Systematics and evolution of the *Triodia basedowii* species complex (Poaceae: Chloridoideae)

Benjamin M. Anderson

MSc BSc BASc

This thesis is presented for the degree of Doctor of Philosophy of the University of Western Australia

School of Plant Biology

2016

Abstract

This thesis investigated evolutionary patterns in a taxonomically challenging species complex of Australian perennial arid zone grasses (*Triodia* R.Br.). With ongoing large-scale disturbance of arid landscapes, especially in the diverse Pilbara region of Western Australia, there is an urgent need to delimit units of biodiversity to assist conservation and restoration efforts. Species delimitation and phylogenetic inference in the *Triodia basedowii* E.Pritz. species complex were undertaken using traditional morphological and genetic methods, and were greatly enhanced by the application of a next-generation sequencing approach: genotyping by sequencing (GBS). In clarifying the systematics of the complex, this thesis contributes to addressing the need for accurate delimitation of biologically significant entities in the Australian arid zone.

To provide an initial assessment of taxon boundaries in the complex, evidence from nuclear (ETS and ITS) and chloroplast (*rps16-trnK* spacer) regions and morphology were evaluated in an integrative taxonomic framework (Chapter 2). There was evidence for multiple new taxa in the complex as well as incongruence between data sets, potentially indicating hybridization. Delimiting taxa revealed high lineage diversity and endemism in the Pilbara, consistent with previous hypotheses that the region acted as a refugium for arid zone species. Taxa showed strong geographic structure in the Pilbara, congruent with recent work on co-occurring animals.

To further test the distinction of taxa and phylogenetic relationships among them, GBS was used to characterise genomic differences between samples (Chapter 3). As the use of reduced representation genomic data sets in species delimitation is relatively new, methodological challenges and solutions were discussed, with broad applicability for similar systems of closely-related species. In addition to traditional concatenation analyses, summary species tree analyses were used with GBS loci for the first time. GBS data and analyses improved the resolution of taxa and relationships among them compared to traditional approaches.

To explore the timing of divergences in the complex in relation to climate and landscape changes in Australia, full chloroplast sequences were used to construct a

fossil-calibrated phylogeny (Chapter 4). Divergences in the *T. basedowii* complex began 1.3–2.8 Ma, prior to a major shift in glacial cycles and the beginning of the formation of sandy dunefields *c*. 1 Ma. The timing of diversification indicates that some lineages in the complex persisted through major climate and landscape changes in the arid zone, possibly assisted by refugial areas such as the Pilbara. To explore earlier indications of a range expansion in *T. basedowii*, previously-generated GBS data (Chapter 3) were used in a program that detects a signal for, and locates the origin of, an expansion (Chapter 4). Across data sets, a consistent signal of westward expansion from central or eastern Australia was detected, suggesting *T. basedowii* maintained populations in central Australia through landscape and climate changes and later expanded onto newer sandy dunefield habitats.

To formalise species delimitations (Chapters 2 & 3), a taxonomic revision was undertaken, describing eight new species and revising four existing species, including two close relatives of the complex (Chapter 5).

Table of Contents

Acknowledgements

First of all, thanks to my supervisors, who provided guidance and encouragement to me while I was undertaking this PhD. Matthew Barrett, thanks for the engaging discussions, good arguments, insights into working with *Triodia* both in the field and in the lab, great field companionship, and a lot of the driving force to push into next-gen methods. Pauline Grierson, thanks for introducing me to the Pilbara for the special place that it is and for exuding an enthusiasm for the *Triodia* work. Siegy Krauss, thanks for quick comments on written work and for the impetus to seek the funding that made a lot of the fieldwork possible. Kevin Thiele, thanks for freely sharing your experience, for teaching me about open air camping, for bringing the key points into focus, for lively debate and for your example of challenging the norm.

One of the highlights of my time in Australia has been the opportunity to work in some amazing places, and I am grateful to those who were part of field trips. Thanks to Russell and Matthew Barrett for allowing me to accompany them to the Kimberley (and patiently answering my "what is that plant?" questions). Thanks to Rob Davis and Tim Hammer for allowing me to accompany them on a trip to central WA. Thanks to Stephen van Leeuwen, Neil Gibson, Kate Brown, Margaret Langley, and Nicholas Casson for an amazing trip to the Little Sandy Desert with camping experiences and damper I won't soon forget. Thanks to Peter Golos for allowing me to accompany him to collect *Triodia* at Nifty Copper Mine. Finally, thanks to Peter Jobson for hospitably hosting me in Alice Springs and taking the time to escort me around central Australia (with an obligatory stop at Uluru) in search of *T. basedowii* samples.

This research has been possible because of facilities to conduct morphological and genetic investigations, including the Western Australian Herbarium and the Biodiversity and Conservation Centre at Kings Park and Botanic Garden. Thanks to the herbarium staff and research associates who helped me use the herbarium and encouraged me in my work, especially Ryonen Butcher, Karina Knight, Kelly Shepherd and Juliet Wege. At Kings Park, thanks go to Annette Johnson for

administrative support and encouragement, to C. Ellery Mayence for assistance in collecting a population of *T. plurinervata*, and to Carole Elliott for good conversations and encouragement.

Some of the analysis of next-gen data were initially trialled at the Pawsey Supercomputing Centre, supported by the governments of Australia and Western Australia. Thanks to the staff for guidance in using their systems and introduction to bash scripts. Thanks to Kym Ottewell, Tara Hopley and Rachel Binks for allowing me to be a part of their project exploring assembly and analysis of GBS data at the Pawsey Centre, and for helpful discussions. Thanks to Deren Eaton for assistance with PyRAD, and to Aaron Chuah for help with scripting and for providing example scripts.

One of the important sources of encouragement in the PhD journey has been meeting other researchers and students at conferences and workshops who share my interest in systematics and phylogeny. Thanks to the Australasian Systematic Botany Society (ASBS) for making their conferences affordable for students to attend, and to the Australian Biological Resources Study for providing funding to attend a workshop and a conference. Thanks as well to the Friends of Kings Park and Botanic Gardens for funds to attend a workshop in Sydney. Thanks to the attendees at ASBS conferences for encouraging student participation and including me in your society. Thanks to fellow students Kaylene Bransgrove, Marina Cheng, Charles Foster, Todd McLay, and Lalita Simpson for comradery. Thanks particularly to Todd McLay for hosting me in his native New Zealand and showing me the marvels of Palmerston North. Thanks to Peter Weston for guidance and encouragement on a visit to Sydney and in subsequent ASBS conferences.

Back in Perth, thanks go to fellow students Emma Dalziel, Wolfgang Lewandrowski, Myles Menz, Alison Ritchie and others for encouragement and inclusion in Australian culture. Thanks to the gang that made Sunday poker nights possible, namely David Leach, Dave Symons, Matthew and Christy Barrett, Arielle Fontaine, Carole Elliott, Sean Tomlinson, and Tony Scalzo, among others.

A major component of my sanity and well-being, influential in getting this thesis done, was a stable home base. Sue and Simon Ford provided a temporary place to stay when I arrived, and I am grateful for their help in the big move. I was particularly fortunate to have been able to stay in the same house during my four years in Perth. Thanks go to Todd Erickson and Ryan Phillips for welcoming me into their share house and for being awesome flatmates. While I am still not sold on this whole cricket thing, I can appreciate it better now (maybe).

I was able to undertake this PhD with support from an International Postgraduate Research Scholarship, an Australian Postgraduate Award and a UWA Top-up scholarship at The University of Western Australia. Funds from the School of Plant Biology also provided me with a computer. This research was partly supported by an Australian Research Council Linkage Project 120100350 (CIs Grierson, Krauss et al. in collaboration with Rio Tinto Iron Ore, Chevron, Botanic Gardens and Parks Authority and the Department of Parks and Wildlife). Collecting in eastern Western Australia and central Australia was partly supported by an ANZ Trustees Foundation - Holsworth Wildlife Research Endowment. Some of the costs of next-gen sequencing were covered by the ASBS through a Hansjörg Eichler Scientific Research Grant.

Finally, thanks are due to my family back in Canada, who supported me from afar and welcomed me back on my two visits. Thanks to my grandparents Don and Ann Williams, who sent cards and money and kept me up to date with the goings on in Vancouver. Thanks to my brother Dan for the Skype sessions and online gaming, to my sisters Lydia and Joanna for their encouragement, and to my brother Gabe for sharing his music. Thanks to my mother Kim, who buoyed my spirits with care packages full of home-made baking and North American candy.

Candidate's declaration

I declare that this thesis has not been previously accepted for a degree, that it was substantially completed during the course of my enrolment, and that all sources of information have been acknowledged in references or in the statement of candidate's contribution. Published work (Chapter 2) is included with the permission of my co-authors.

Benjamin M. Anderson

Statement of candidate contribution

This thesis contains both single-author and multi-author works. B.M. Anderson wrote Chapters 1 and 6, while Chapters 2–5 are co-authored papers intended for publication, with Chapter 2 already published. The relevant chapters and contributions of all authors are outlined below.

Chapter 2. Untangling a species complex of arid zone grasses (*Triodia*) reveals patterns congruent with co-occurring animals

Authors: Benjamin M. Anderson, Matthew D. Barrett, Siegfried L. Krauss, Kevin R. Thiele

BMA contributed to study design, collected and analysed the data, and wrote the manuscript. MDB contributed to study design, and assisted with data collection and interpreting results. SLK contributed to study design. KRT contributed to study design and assisted with interpreting results. All authors commented on the manuscript.

Chapter 3. Insights into using genotyping by sequencing in a species complex: a case study using Australian hummock grasses (*Triodia*)

Authors: Benjamin M. Anderson, Kevin R. Thiele, Siegfried L. Krauss, Matthew D. Barrett

BMA contributed to study design, helped collect the genetic material, wrote the scripts and analysed the data, and wrote the manuscript. KRT helped collect genetic material and interpret results. SLK commented on study design. MDB contributed to study design and helped collect genetic material and interpret results. All authors commented on the manuscript.

Chapter 4. Diversification timing for Australian arid zone grasses (*Triodia*) and evidence for recent range expansion across sandy deserts

Authors: Benjamin M. Anderson, Kevin R. Thiele, Pauline F. Grierson, Siegfried L. Krauss, Paul G. Nevill, Ian D. Small, Xiao Zhong, Matthew D. Barrett

BMA contributed to study design, helped collect genetic material, wrote scripts, analysed the data, and wrote the manuscript. KRT helped collect genetic material and assisted with interpreting results. PFG and SLK helped interpret results. PGN helped collect genetic material and provided sequencing of chloroplast genomes. IDS and XZ assembled chloroplast genomes. MDB contributed to study design, helped collect genetic material, provided sequencing and assembly of some chloroplast genomes, and helped interpret results. All authors commented on or contributed to the manuscript.

Chapter 5. A revision of the *Triodia basedowii* species complex and close relatives (Poaceae: Chloridoideae)

Authors: Benjamin M. Anderson, Kevin R. Thiele, Matthew D. Barrett

BMA contributed to study design, made morphological measurements, synthesised the data, and wrote the manuscript. KRT helped refine taxonomic concepts and format. MDB contributed to study design, examined a morphological character and helped refine taxonomic concepts. All authors commented on the manuscript.

Works arising from this thesis

Publication

Anderson BM, Barrett MD, Krauss SL, Thiele K (2016) Untangling a species complex of arid zone grasses (*Triodia*) reveals patterns congruent with co-occurring animals. *Molecular Phylogenetics and Evolution*, **101**, 142–162.

Conference presentations

Anderson BM, Barrett MD, Grierson PG, Krauss SL, Thiele K (2013) Taxon delimitation in the *Triodia basedowii* E.Pritz. species complex. Oral presentation given at the Australasian Systematic Botany Society Conference: Systematics without Borders, in Sydney, NSW, 1–4 December 2013.

Anderson BM, Barrett MD, Grierson PG, Krauss SL, Thiele K (2014) Complex evolutionary patterns in the diverse Pilbara region of WA: an example from spinifex grass (*Triodia* spp.). Oral presentation given at the Kwongan Workshop in Perth, WA, 22 July 2014.

Anderson BM, Barrett MD, Grierson PG, Krauss SL, Thiele K (2014) Evolutionary diversity in the *Triodia basedowii* E.Pritz. species complex. Oral presentation given at the Ecological Society of Australia 2014 Annual Conference in Alice Springs, NT, 2 October 2014.

Anderson BM, Barrett MD, Grierson PG, Krauss SL, Thiele K (2014) Exploring hybridization in the *Triodia basedowii* E.Pritz. species complex using next-generation sequencing. Oral presentation given at the Australasian Systematic Botany Society Conference: Next Generation Systematics, in Palmerston North, New Zealand, 24 November 2014.

Anderson BM, Barrett MD, Grierson PG, Krauss SL, Thiele K (2015) Using genotyping by sequencing to resolve evolutionary relationships in a species complex of Australian arid zone grasses (*Triodia*). Oral presentation given at the Australasian

Systematic Botany Society Conference: Building our Botanical Capital, in Canberra, ACT, 1 December 2015.

Chapter 1. General introduction

The *Triodia basedowii* E.Pritz. species complex comprises a group of taxonomically difficult, closely-related species that occur across the Australian arid zone, including within the Pilbara region of Western Australia, an area of biological richness and high endemism (Pepper *et al.* 2013b). The Pilbara is also rich in mineral resources, providing 93% (665 Mt) of the state's iron ore production in 2014–15 (Department of Mines and Petroleum 2015). Ongoing resource extraction in the Pilbara and other arid regions is resulting in large disturbances to landscapes. There is thus an urgent need to accurately delimit units of biodiversity in these regions to assist in pre-mining biological surveys and impact assessments, increase effectiveness of post-mining restoration efforts and seed sourcing, and inform conservation planning. To contribute to addressing this need, this thesis aims to clarify the systematics of the *T. basedowii* complex. In addition, this thesis aims to explore the use of a next-generation sequencing method in species delimitation and to elucidate the evolutionary history of the complex, providing insights into the evolution of the Australian arid zone biota.

This introductory chapter provides a general overview of approaches to delimiting species, current knowledge of the genus *Triodia* and the *T. basedowii* complex, and the evolutionary context for Australian arid zone plants.

Species delimitation

The definition of a species has been and continues to be controversial. This can be partly traced to Darwin's idea that species were not just a taxonomic rank but also real evolutionary entities (Brooks & McLennan 1999; de Queiroz 2011). This definition naturally leads to the question of how to detect and delimit these entities. Early approaches were reliant on experts to subjectively define an appropriate degree of differentiation (de Queiroz 2011), but modern approaches attempt to use more objective criteria. Examples of criteria for recognising species in modern approaches include phenetic clustering (Sneath & Sokal 1973), reproductive barriers (Mayr

1942), mate recognition and fertilization (Paterson 1985), ecological niche (Van Valen 1976), fixed characters (Cracraft 1983; Nixon & Wheeler 1990; Davis & Nixon 1992), monophyly (Donoghue 1985; Mishler 1985), and exclusive allele coalescence (Avise & Ball 1990). These varied criteria for detecting species lead to conflict, especially when they are used as definitions of species (necessary properties), since they may be acquired at different stages of lineage divergence (see de Queiroz 2007).

De Queiroz (1998, 2007, 2011) attempted to address these conflicts by suggesting that rather than defining what a species is, the various criteria applied in previous species concepts are better used as lines of evidence that speciation has occurred, and that species are defined as separately evolving segments of metapopulation lineages. Various criteria (e.g. reproductive isolation) are acquired during the course of a species existence and do not actually define what a species is (de Queiroz 1998, 2007, 2011). While this argument did much to unify disparate views of species, in some respects it simply reframed the question of defining species to one of defining lineages. In that sense, species delimitation retains a largely subjective component that requires a researcher to evaluate evidence for lineage divergence (i.e. in an integrative framework), and make a call as to whether to recognise a given lineage or lineages as species. In practice, species delimitation often relies on clear morphological differences between taxa, but in taxonomically challenging cases such as the *T. basedowii* species complex researchers are required to turn to more powerful and costly sources of evidence.

Integrative taxonomy

An important outcome of de Queiroz' unification of disparate species concepts was to shift focus to the examination of multiple lines of evidence rather than a single criterion (de Queiroz 1998). A variety of approaches highlighted the limitations of different data sources and the need to avoid reliance on a single method (Sites & Marshall 2004). In order to assess biological diversity from multiple disciplines and to avoid a taxonomy based solely on newly-emerging DNA barcoding, an "integrative taxonomy" was advocated by Dayrat (2005) and Will *et al*. (2005).

Within integrative taxonomy, lines of evidence are valued for their ability to detect lineage divergence rather than their source (e.g. morphological, molecular or ecological) (Padial *et al.* 2010). Approaches to integrating data from multiple sources include congruence, where species are delimited only when all examined data sources support that delimitation, and cumulation, where each data source alone can be enough to delimit a species (Padial *et al.* 2010). Discordance among multiple lines of evidence can be resolved with recourse to evolutionary explanations, allowing species to be delimited even when one line of evidence (e.g. monophyly of a molecular marker) may not agree with others (Schlick-Steiner *et al.* 2010). Most integrative taxonomic approaches look for congruence between character sets rather than including all characters in a single analysis (Yeates *et al.* 2011). The need to rely on multiple lines of evidence extends to molecular markers as well, where processes such as incomplete lineage sorting and hybridization have the potential to produce conflicting phylogenetic signals between markers (Maddison 1997; Naciri & Linder 2015). The more conservative approach to delimitation, searching for congruence and hesitating to delimit species that are incongruent between methods, is considered preferable to splitting at the highest resolution available from the data (Miralles & Vences 2013; Carstens *et al.* 2013). This conservatism is an important consideration for taxonomists with access to increasingly powerful tools for characterising genomic diversity.

Next-generation sequencing and taxonomy

Next-generation sequencing (NGS) holds tremendous promise for addressing biosystematic questions. A number of approaches are currently available for generating NGS data to tackle phylogenetic and phylogeographic problems (reviewed in McCormack *et al.* 2013). These include sequencing PCR products of specific regions of the genome, using a restriction enzyme to generate fragments of the genome, and using probes to capture conserved sequences (McCormack & Faircloth 2013; McCormack *et al.* 2013). While genomic divergence has traditionally been inferred from e.g. morphological differentiation, NGS approaches allow taxonomists to measure genomic divergence more directly.

Integrative taxonomic studies (e.g. Leaché *et al.* 2009; Burbrink *et al.* 2011; Barrett & Freudenstein 2011; Pepper *et al.* 2013a) have applied species delimitation methods to relatively long loci generated in Sanger sequencing (Sanger *et al.* 1977). NGS methods that generate longer loci (e.g. amplicon sequencing or sequence capture) have also been employed in species delimitation (e.g. O'Neill *et al.* 2013; Giarla & Esselstyn 2015; Potter *et al.* 2016) and phylogenetics (e.g. Pyron *et al.* 2014; Leaché *et al.* 2015), but these methods are challenging to use in plant species, most of which have a history of polyploidy (Soltis *et al.* 2009) and therefore often multiple copies of genomic targets. Studies on non-model plants lacking prior genomic information to design probes make the use of sequence capture approaches even more difficult. Reduced representation data sets created with restriction enzymes offer a cheaper and faster alternative to sequence capture approaches and hold promise for organisms without prior genomic information (reviewed in Davey *et al.* 2011), such as the *T. basedowii* complex.

To date there has been relatively little application of NGS reduced representation data to delimiting species, partly due to the reduced information per shorter locus (species tree methods rely on generating phylogenetic trees for each locus) and the focus on extracting single SNPs (McCormack & Faircloth 2013). Phylogenetic analyses of reduced representation data sets (e.g. Eaton & Ree 2013; Nadeau *et al.* 2013; Wagner *et al.* 2013; Escudero *et al.* 2014; Pante *et al.* 2014; Leaché *et al.* 2015; Qi *et al.* 2015) often concatenate recovered loci rather than account for gene tree incongruence (Edwards *et al.* 2016), but other studies (e.g. Gohli *et al.* 2015; Nicotra *et al.* 2016) have also employed single SNPs per locus in a coalescent approach (Bryant *et al.* 2012), including a study focusing on species delimitation (Leaché *et al.* 2014). The potential of these data sets in species delimitation studies is still being explored. An older approach also using restriction enzymes but based on fragment length rather than sequence, amplified fragment length polymorphism (AFLP; Vos *et al*. 1995), has been useful for differentiating closely-related species and varieties (reviewed in Mueller & Wolfenbarger 1999; e.g. Hodkinson *et al*. 2002). This thesis examines the effectiveness of a reduced representation data set generated with genotyping by sequencing (Elshire *et al.* 2011) for delimiting species in the *T. basedowii* complex.

Triodia

> "The Australian species are all endemic, and include the troublesome prickly desert grasses, sent by early collectors with the name of 'Porcupine Grass,' but now more generally and disagreeably known to explorers by that of 'Spinifex,' totally disconnected however from the botanical genus *Spinifex*."
>
> — Bentham (1878)

The genus *Triodia* R.Br. comprises iconic Australian perennial grasses, commonly known as "spinifex," which are found predominantly in arid and semi-arid regions of every state and territory except Tasmania (Lazarides 1997). The genus forms a dominant component of the vegetation in hummock grasslands, which cover more than 18% of the Australian continent (Department of the Environment 2006). As noted above by Bentham (1878), *Triodia* is notorious for pungent leaf blade tips, which can make traversing spinifex country and collecting plants particularly painful. *Triodia* often forms dense hummocks due to repeated branching and stoloniferous expansion, sometimes forming large rings as older material dies in the centre (Burbidge 1945; Jacobs 1992; Lazarides 1997).

Biology and ecology

Triodia utilises C_4 photosynthesis and has a modified bundle-sheath leaf anatomy (Jacobs 1971, 1992; Carolin *et al.* 1973; McWilliam & Mison 1974; Hattersley & Watson 1976; Prendergast *et al.* 1987). The modified leaf anatomy, with incomplete bundle-sheaths and photosynthetic tissue reduced to small patches adjacent to stomatal grooves, distinguishes *Triodia* from the rest of its subfamily (Renvoize 1983; Van den Borre & Watson 1997). *Triodia* has been divided into two informal groups based on the nature of the leaves, with "hard" species having much more sclerified and pungent leaves, and "soft" species being less sclerified and more flexible, and producing abundant resin (Gardner 1952; Burbidge 1953; Mant *et al.* 2000). Hard species have stomatal grooves evenly distributed across the abaxial surface of the leaf blades while soft species have them only in a narrow central strip, leaving two ungrooved regions on the abaxial surface. *Triodia* leaves fold dorsally

such that in dry conditions the blades are terete, with more sclerified leaves typically less able to unfold in wet conditions (Burbidge 1945). The highly sclerified and modified leaf anatomy may be important in maintaining structural integrity during drought stress.

As dominant components of the vegetation in parts of the arid zone, *Triodia* provides food and shelter for animals (Ealey *et al.* 1965; Kitchener *et al.* 1983; Losos 1988; Daly *et al.* 2008). Examples include the indirect support of lizard populations by sustaining detritivores like termites (Morton & James 1988), and thermal shelter for the hare wallaby (Dawson & Bennett 1978). Human use of *Triodia* has included the harvesting and processing of resin from soft species for use as an adhesive material for hafting or caulking (Gamage *et al.* 2012), with relatively little traditionally added to the resin beyond incidental sand or vegetation fragments (Mondal *et al.* 2012).

Triodia typically occurs on low-nutrient soils (Burbidge 1953; Beadle 1981; Griffin 1984), conditions that are thought to promote sclerophylly (Beadle 1954, 1966). While a recent study (Armstrong & Garnett 2011) suggested that *Triodia* does not have strong substrate specificity, other observations suggest substrates play an important role in species distributions (Burbidge 1959; Beard 1969; Beadle 1981). Distributions are also highly influenced by precipitation patterns and the availability of water (Winkworth 1967; Griffin 1990), with local landscape partitioning thought to reflect differing water availability and requirements for different species (Grigg *et al.* 2008).

It is generally accepted that *Triodia* is adapted to cyclic fire (Winkworth 1967), with fire considered important in community dynamics (Griffin 1992; Jacobs 1992; Wright & Clarke 2007; Armstrong & Phillips 2012) and even essential to keep *Triodia* present in many ecosystems (Jacobs 1980; Chapman 1996; Morrison *et al.* 2005). Resinous species, in particular, are highly flammable (Lazarides 1997). Responses to fire vary both between and within species (Burbidge 1943; Rice & Westoby 1999), with some species fire-killed and returning from seed and others resprouting following fire. There is controversy over whether fire stimulates germination in *Triodia*, with evidence both for (Armstrong & Legge 2011) and

against (Bogusiak *et al.* 1990). Seeds exposed to smoke or a smoke chemical showed varied germination responses between and within species (Wells & Dixon 1999; Erickson *et al.* 2016).

Relatively little is known about breeding systems in *Triodia*. Limited evidence suggests that some species are outcrossers (Wells *et al.* 2003) or self-sterile (Burbidge 1945). Flowering is thought to be dependent on the timing and amount of rainfall (Jacobs 1992), with seed production often quite poor and perhaps dependent on sustained favourable conditions (Burbidge 1953; Jacobs 1992). Seeds often accumulate under hummocks (Westoby *et al.* 1988), with ant harvesting (Jacobs 1992; Chapman 1996; Dickman *et al.* 1999) possibly effecting some dispersal.

There has been very limited investigation of ploidy levels in *Triodia*, with two published tentative counts (2n=30, 2n=60) suggesting the existence of polyploidy (Jacobs 1971). Polyploidy is a common evolutionary process in angiosperms (Soltis *et al*. 2009) and could have important implications for interpreting reproductive barriers and genomic divergence in *Triodia*.

Evolution and systematics

Triodia is part of the grass subfamily Chloridoideae (Clayton & Renvoize 1986; Watson & Dallwitz 1992; Peterson *et al.* 2010), the ancestors of which are thought to have diversified into drier habitats in Africa concurrent with the evolution of C_4 photosynthesis (Bouchenak-Khelladi *et al.* 2010). The closest relatives of *Triodia*, variably recovered as *Aeluropus* or *Orinus* (Peterson *et al.* 2010) or *Cleistogenes* (Grass Phylogeny Working Group II 2012), all occur in dry or sandy environments. *Aeluropus* occurs from the Mediterranean to China in salty soils of seashores and deserts, *Orinus* occurs in Tibet and China on dunes at high altitudes, while *Cleistogenes* is found from Europe to Japan in dry and open environments (Clayton & Renvoize 1986; Watson & Dallwitz 1992). As biome shifts are uncommon in colonisations (Crisp *et al.* 2009), it is likely that ancestors of *Triodia* arrived, by long-distance dispersal, already adapted to the dry environments developing in Australia. Recent molecular divergence dating suggests that *Triodia* likely arrived in

Australia *c*. 24–14 million years ago (Ma) and radiated during Miocene aridification (Toon *et al.* 2015).

Phylogenetic studies indicate that most soft species form a clade within *Triodia* while hard species form a paraphyletic grade (Mant *et al.* 2000; Toon *et al.* 2015). One clade of hard species is recovered as sister to the rest of the genus and contains species that are largely restricted to southern Australia, while subsequent divergences in the rest of the genus are found progressively northward (Toon *et al.* 2015). This might suggest that ancestors of *Triodia* established in southern Australia and diversified northward through the developing arid zone and into the tropical savannahs (where much of the diversity in the derived clade comprising soft species is found).

Triodia was erected by Robert Brown (1810) and initially comprised six species, four of which were retained by Bentham (1878), who added two more. Bentham (1883) and Hackel (1887) later widened the genus to include European, African and North American grasses, but Stapf applied the name solely to Australian material (see Hubbard 1937). Burbidge (1946a; b; c) supported Stapf's application of the name to only Australian species. The type species (*Triodia pungens* R.Br.) was selected by Hitchcock (1920).

Revisionary work on the genus includes early treatments by Burbidge (1946a, 1953, 1960a) and a more recent one by Lazarides (1997), who synonymised the genus *Plectrachne* Henr. with *Triodia*. While Jacobs (2004) maintained these two genera as distinct, phylogenetic work shows that former *Plectrachne* species are nested within *Triodia* (Mant *et al.* 2000). Recently, Crisp *et al*. (2015) synonymised *Monodia* Jacobs and *Symplectrodia* Lazarides with *Triodia* following phylogenetic work (Toon *et al.* 2015) showing that these too are nested within *Triodia*.

Species taxonomy in *Triodia* is based almost exclusively on morphological characters. There are currently 73 described species (Lazarides 1997; Barrett *et al.* 2005; Armstrong 2008; Barrett & Barrett 2011, 2015; Hurry *et al.* 2012; Crisp *et al.* 2015), with more being discovered based on collections from relatively unexplored areas such as northern Australia and taxonomic studies of species complexes and

broadly applied names. One important but unresolved species complex occurring in the arid zone is the *T. basedowii* complex.

The *Triodia basedowii* complex and close relatives

The *Triodia basedowii* species complex and its relatives comprise an early-diverging clade of hard species, sister to most of the rest of the genus (not including a small clade of southern species) and geographically distributed (Fig. 1.1) between (and partially overlapping both) southern and northern clades (Toon *et al.* 2015). Within this central area, an important shift in rainfall pattern occurs south to north, from predominantly winter to predominantly summer rainfall (Burbidge 1960b). The clade contains species spanning this shift and is thus phylogenetically important for understanding evolution in the genus as it diversified northward.

Species in the clade are distinguished from other hard species by their many-nerved glumes. Two of the more divergent species, *T. rigidissima* (Pilg.) Lazarides and *T. desertorum* (C.E.Hubb) Lazarides, were formerly included in *Plectrachne* and have long awns on their lemmas, while the rest of the species, which are the focus of this thesis, have awnless lemmas. These species comprise two close relatives of the *T. basedowii* complex, *T. concinna* N.T.Burb. and *T. plurinervata* N.T.Burb., both raised by Burbidge (1960a), and the complex itself.

The *T. basedowii* complex (examples in Figs. 1.2 & 1.3) currently comprises two described species, *T. basedowii* E.Pritz (1918) and *T. lanigera* Domin (1912), along with five phrase-named species, some of which were raised by the authors during the course of this study. These phrase-named species are *T.* sp. Shovelanna Hill (S. van Leeuwen 3835), *T.* sp. Little Sandy Desert (S. van Leeuwen 4935), *T.* sp. Peedamulla (A.A. Mitchell PRP 1636), *T.* sp. Warrawagine (A.L. Payne PRP 1859), and *T.* sp. Pannawonica (B.M. Anderson & M.D. Barrett BMA 89). Names of informally recognised taxa with "sp." and a geographical epithet follow the Australian convention for undescribed taxa (see Barker 2005) and are on the Australian Plant Census. Additional entities have been noted by the author during investigations of the complex and are referred to in the text of this thesis with temporary designations (e.g. *T.* "shova").

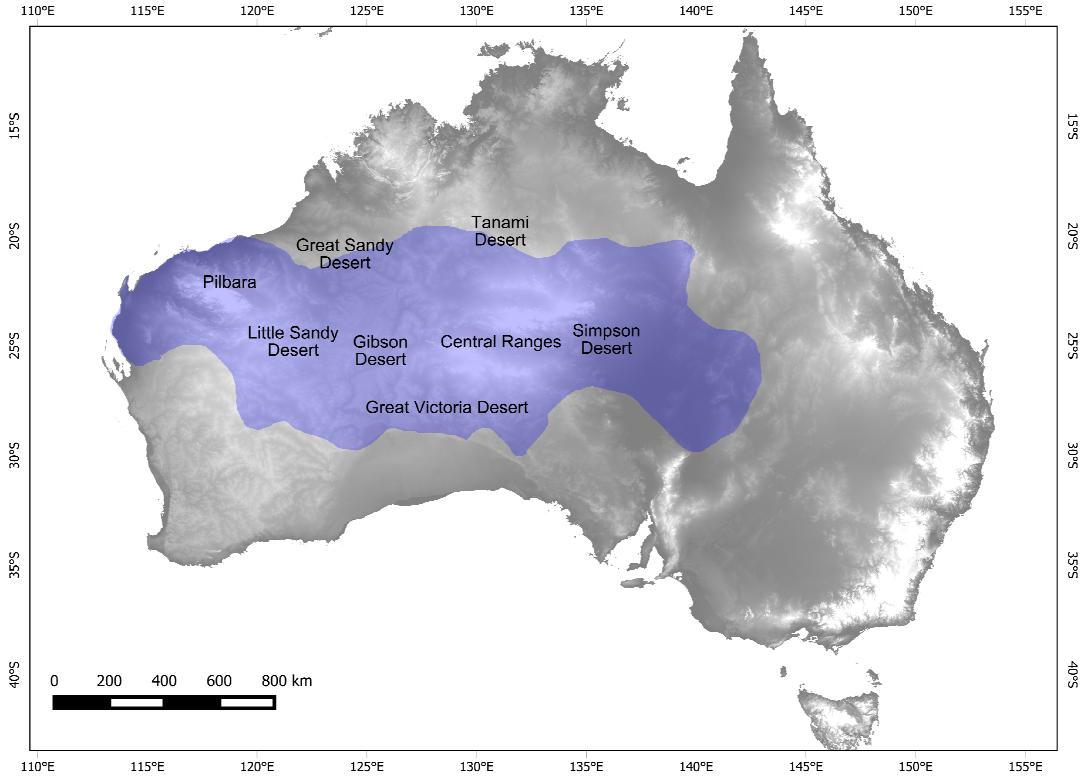

Figure 1.1. Approximate geographic distribution of the *Triodia basedowii* complex in Australia. The distribution (purple), major deserts and two topographically complex areas (Pilbara and Central Ranges) are indicated.

Figure 1.2. Examples of members of the *T. basedowii* complex and their habitats. (A) a robust plant of *Triodia basedowii*; (B) co-occurrence of *T. basedowii* and *T.* sp. Shovelanna Hill on sandy to rocky substrates; (C) dune environment of *T.* sp. Little Sandy Desert; (D) sandy conditions for *T. lanigera*, contrasted with rockier conditions in the south (F); (E) rocky landscape in the Hammersley sub-region of the Pilbara (non-*T. basedowii* complex *Triodia* sp. in the foreground).

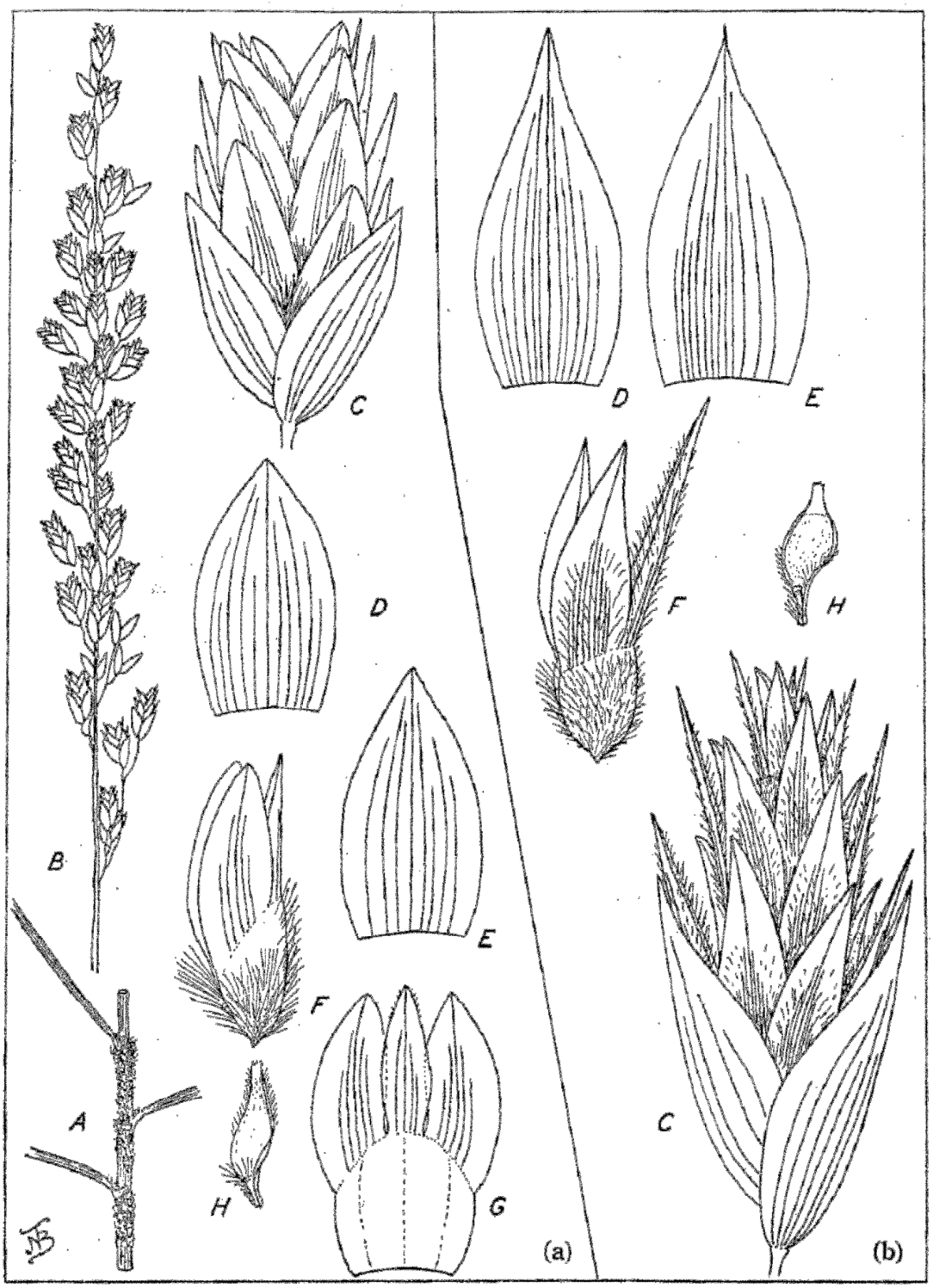

Figure 1.3. Line drawings of (a) *Triodia basedowii* and (b) *T. lanigera*. (A) portion of culm showing orifice, (B) portion of panicle, (C) spikelet, (D) lower glume, (E) upper glume, (F) lateral view of floret, (G) lemma from within, (H) palea. Reproduced from Burbidge (1953) with permission from CSIRO Publishing.

Taxonomic diversity in the *T. basedowii* complex is concentrated in Western Australia, with only one of the currently described species (*T. basedowii*) occurring widely across much of the arid zone. Previous workers (Burbidge 1953; Lazarides *et al.* 2005) have commented on the high degree of similarity and close relationship between *T. basedowii* and *T. lanigera*. Difficulty identifying plants as one or the other of these species, especially in the Pilbara region of Western Australia, and the proliferation of phrase names to account for additional observed morphological diversity, suggest the complex requires taxonomic investigation and the application of molecular methods to disentangle taxa and relationships. Given the complex spans the Pilbara and adjacent regions, noted for significant phylogeographic structure in the biota (Pepper *et al.* 2013b), it is an ideal system for exploring evolutionary patterns in Australian arid zone plants.

Evolution in the Australian arid zone

The current arid and semi-arid zones in Australia (here defined where average annual precipitation is less than *c*. 500 mm/year) cover more than 60% of the continent, but Australia has not always been a predominantly dry continent. Australia was warm and wet and supported widespread rainforests during the Eocene (Martin 2006). The development of the Circum-Antarctic Current near the Eocene-Oligocene transition *c*. 33 Ma (Walker *et al.* 2012) coincided with the onset of cooler conditions, but subsequent warming resulted in a mid-Miocene (*c*. 16 Ma) warm peak with evidence of rainforest plants in the Australian interior (Martin 2006; Macphail 2007). Since the mid-Miocene peak there has been a continued trend of cooling and aridification as Australia drifted northward, until another relatively warm period in the early Pliocene (Martin 2006; Macphail 2007).

From the onset of the Pleistocene (*c*. 2.5 Ma), glacial cycles produced cooler and more arid conditions through globally lower temperatures and lower sea levels (both resulting in reduced precipitation in the Australian interior) (Williams 1984). The stony deserts of central Australia are thought to have begun to form *c*. 2–4 Ma, during this period of global cooling (Fujioka *et al.* 2005). Roughly 0.8–1.2 Ma there was a global shift from *c*. 40 thousand year (ky), low amplitude glacial cycles to

c. 100 ky, higher amplitude cycles (Pisias & Moore 1981; Mudelsee & Stattegger 1997; Clark *et al.* 1999). This shift in glacial cycle amplitude corresponds to the earliest evidence of sandy dunefields in central Australia (Fujioka *et al.* 2009). Subsequent glacial cycles have followed a trend of decreasing precipitation up to the present interglacial, which is drier than the previous one *c*. 120 thousand years ago (Martin 2006). The aridification of Australia and the more recent Pleistocene glacial cycles have likely had a significant impact on the evolution and composition of the flora.

Australian arid zone plants comprise groups descended from mesic ancestors that have been in Australia since before the Oligocene, as well as more recently arrived arid-adapted taxa (Crisp & Cook 2013). Older groups include *Acacia* (Ariati *et al.* 2006), *Callitris* (Sakaguchi *et al.* 2013) and Goodeniaceae (Jabaily *et al.* 2014). More recent long-distance dispersers include *Lepidium* (Mummenhoff *et al.* 2004), chenopods (Salicornioideae, Shepherd *et al.* 2004; Camphorosmoideae, Kadereit & Freitag 2011) and *Triodia* (Toon *et al.* 2015). The timing of diversification in the flora is important for understanding how these plants responded to climate changes, but there are few studies that have investigated this. It has been suggested that diversification in arid zone organisms occurred during the late Miocene and Pliocene from the onset of aridification, and that diversity was maintained during Pleistocene glacial cycles in multiple refugial areas, possibly including the Pilbara (Byrne *et al.* 2008). The Australian Goodeniaceae contain multiple examples of arid zone lineages that have diversified since the late Miocene (Jabaily *et al.* 2014); arid lineages of *Santalum* also diverged roughly 13–19 Ma (Harbaugh & Baldwin 2007). The phylogeographic structure in *Callitris* further suggests the persistence of lineages in refugial areas in central Australia through the Pleistocene (Sakaguchi *et al.* 2013). In contrast, more recent diversification is noted for Australian Camphorosmeae, with a crown age of *c*. 5 Ma (Kadereit & Freitag 2011), and *Lepidium,* with *c*. 20 Australian species thought to have radiated within the last *c*. 1.3 Ma (Mummenhoff *et al.* 2004). More investigations of diversification timing in Australian plants are needed to establish the commonality of older *versus* more recent radiations. Diversification timing in the *T. basedowii* complex is investigated in this thesis to partly address this knowledge gap.

In addition to diversification timing, spatial patterns of diversity are also important in revealing how organisms responded to climate changes (e.g. Hewitt 1999). One of the emerging patterns in studies on Australian arid zone organisms is the high diversity and endemism in the Pilbara region (Pepper *et al.* 2013b). Multiple animal groups, such as lizards (Pepper *et al.* 2008) and frogs (Catullo *et al.* 2011), show striking differentiation between subregions within the Pilbara and between the Pilbara and adjacent sandy deserts. One possible explanation for the high diversity and endemism is that the region has functioned as a refugium during adverse climatic conditions (Burbidge 1959; Byrne *et al.* 2008). Higher diversity in lizards from the Pilbara and other topographically complex regions compared to lower-lying deserts is consistent with this refugial hypothesis (Pepper *et al.* 2011).

The concept of a refugium implies the expansion of populations to adjoining areas following amelioration of adverse conditions. A study on geckos (Fujita *et al.* 2010) showed evidence for multiple invasions of the arid zone from mesic refugia and range expansions in the last *c*. 1 Ma. Similarly, a late Pleistocene expansion was inferred for parthenogenic grasshoppers (Kearney *et al.* 2006). The formation of sandy dunefields starting *c*. 1 Ma (Fujioka *et al.* 2009) could have provided new habitat for species adapted to sandy environments, facilitating their expansion, either from these mesic refugia or from previously restricted sandy areas.

While refugial and range expansion patterns have been explored in studies on Australian arid zone animals, detailed genetic studies of plant taxa inhabiting both putative refugial areas and the sandy deserts are lacking. Both refugial and range expansion patterns are explored in the *T. basedowii* complex in this thesis.

Aims and structure of this thesis

This thesis aims broadly to resolve taxonomic problems in the *T. basedowii* species complex using morphological, genetic and NGS methods, to provide insights into the evolutionary history of the complex with extension to Australian arid zone plants, and to provide a taxonomic revision of the group. The research is presented as a series of papers; as such, some repetition is unavoidable. Each paper has been

prepared with a different journal in mind, but style has been modified where possible for coherence.

Chapter 2 focuses on evaluating evidence for independent lineages in the *T. basedowii* complex using morphological and Sanger sequence data. Distribution patterns of putative taxa are then compared with those of co-occurring animals from previous studies, especially in the Pilbara. The entities identified in this chapter informed sampling for subsequent chapters. This chapter also proposes biogeographical hypotheses to explain patterns of species richness in the complex (addressed again in Chapter 4). This chapter has been published and will be referred to in the remainder of this thesis as (Anderson *et al.* 2016; Chapter 2).

Chapter 3 presents a case study of the use of genotyping by sequencing (GBS) data (NGS reduced representation data) in a species complex. It aims to test the distinction of entities in the complex, especially those identified in (Anderson *et al.* 2016; Chapter 2), and improve phylogenetic resolution. In doing so, it presents a number of methodological insights for assembling and analysing GBS data, with broad applicability to similar systems of closely-related species.

Chapter 4 uses alignments across whole chloroplast genomes from representatives of each grass subfamily (Genbank) and from a denser sampling of Chloridoideae and *Triodia* (newly sequenced) to create a fossil-calibrated chronogram (time-calibrated phylogeny), in order to provide a temporal context for diversification in the *T. basedowii* complex. In addition, GBS data from Chapter 3 is used to test for a signal, and to locate the origin, of a recent range expansion in the widespread species *T. basedowii*. Together, the time-calibrated phylogeny and evidence for expansion allow for a new hypothesis of the historical biogeography of the complex.

In Chapter 5, the taxonomic implications of combined traditional (Anderson *et al.* 2016; Chapter 2) and genomic (Chapter 3) results are used to support a revision of the *T. basedowii* complex and close relatives, including the naming and description of eight new species.

Finally, Chapter 6 provides a discussion of the implications of this study for the taxonomy of the *T. basedowii* complex, the use of GBS in a species complex, and the evolution of Australian arid zone plants.

References

Anderson BM, Barrett MD, Krauss SL, Thiele K (2016) Untangling a species complex of arid zone grasses (*Triodia*) reveals patterns congruent with co-occurring animals. *Molecular Phylogenetics and Evolution*, **101**, 142–162.

Ariati SR, Murphy DJ, Udovicic F, Ladiges PY (2006) Molecular phylogeny of three groups of acacias (*Acacia* subgenus *Phyllodineae*) in arid Australia based on the internal and external transcribed spacer regions of nrDNA. *Systematics and Biodiversity*, **4**, 417–426.

Armstrong G (2008) *Triodia caelestialis* (Triodieae: Chloridoideae: Poaceae), a new species from the central Kimberley, Western Australia. *Journal of the Royal Society of Western Australia*, **91**, 313–317.

Armstrong G, Garnett S (2011) Landscape partitioning among *Triodia* spp. (Poaceae) in the fire prone Kimberley, north-west Australia. *Austral Ecology*, **36**, 849–857.

Armstrong G, Legge S (2011) The post-fire response of an obligate seeding *Triodia* species (Poaceae) in the fire-prone Kimberley, north-west Australia. *International Journal of Wildland Fire*, **20**, 974–981.

Armstrong G, Phillips B (2012) Fire history from life-history: determining the fire regime that a plant community is adapted using life-histories. *PLOS ONE*, **7**, e31544.

Avise J, Ball RJ (1990) Principles of genealogical concordance in species concepts and biological taxonomy (D Futuyma, J Antonovics, Eds,). *Oxford Surveys in Evolutionary Biology*, **7**, 45–67.

Barker B (2005) Standardising informal names in Australian publications. *Australian Systematic Botany Society Newsletter*, **122**, 11–12.

Barrett RL, Barrett MD (2011) Two new species of *Triodia* (Poaceae: Triodieae) from the Kimberley region of Western Australia. *Telopea*, **13**, 57–67.

Barrett RL, Barrett MD (2015) Twenty-seven new species of vascular plants from Western Australia. *Nuytsia*, **26**, 21–87.

Barrett CF, Freudenstein J V (2011) An integrative approach to delimiting species in a rare but widespread mycoheterotrophic orchid. *Molecular Ecology*, **20**, 2771–2786.

Barrett RL, Wells GB, Dixon KW (2005) New taxa and combinations: Subfam. Chloridoideae, Trib. Triodieae, *Triodia*. In: *Flora of Australia* (ed Mallett K), pp. 458–459. CSIRO Publishing, Melbourne.

Beadle NCW (1954) Soil phosphate and the delimitation of plant communities in

eastern Australia. *Ecology*, **35**, 370–375.

Beadle NCW (1966) Soil phosphate and its role in molding segments of the Australian flora and vegetation, with special reference to xeromorphy and sclerophylly. *Ecology*, **47**, 992–1007.

Beadle NCW (1981) *The Vegetation of Australia*. Cambridge University Press, Cambridge.

Beard J (1969) The natural regions of the deserts of Western Australia. *Journal of Ecology*, **57**, 677–711.

Bentham G (1878) Gramineae. In: *Flora Australiensis* , pp. 605–607. L. Reeve & Co., London.

Bentham G (1883) Gramineae. In: *Genera Plantarum* (eds Bentham G, Hooker J), pp. 1175–1176. L. Reeve & Co., London.

Bogusiak A, Rice B, Westoby M (1990) Seedling emergence of hummock grasses in relation to the effects of fire. *Australian Rangeland Journal*, **12**, 25–28.

Bouchenak-Khelladi Y, Verboom GA, Savolainen V, Hodkinson TR (2010) Biogeography of the grasses (Poaceae): a phylogenetic approach to reveal evolutionary history in geographical space and geological time. *Botanical Journal of the Linnean Society*, **162**, 543–557.

Brooks DR, McLennan DA (1999) Species: turning a conundrum into a research program. *Journal of Nematology*, **31**, 117–133.

Brown R (1810) Gramineae. In: *Prodromus floræ Novæ Hollandiæ et Insulæ Van-Diemen* , pp. 182–183. R. Taylor.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A (2012) Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, **29**, 1917–1932.

Burbidge N (1943) Ecological succession observed during regeneration of *Triodia pungens* R.Br. after burning. *Journal of the Royal Society of Western Australia*, **28**, 149–156.

Burbidge N (1945) Morphology and anatomy of the Western Australian species of *Triodia* R.Br. I. General morphology. *Transactions of the Royal Society of South Australia*, **69**, 303–313.

Burbidge N (1946a) A revision of the Western Australian species of *Triodia* R. Br. *Journal of the Royal Society of Western Australia*, **30**, 15–33.

Burbidge N (1946b) Foliar anatomy and the delimitation of the genus *Triodia* R. Br. *Blumea, Supplement*, **3**, 83–89.

Burbidge N (1946c) Morphology and anatomy of the Western Australian species of *Triodia* R.Br. II. Internal anatomy of leaves. *Transactions of the Royal Society of South Australia*, **70**, 221–236.

Burbidge N (1953) The genus *Triodia* R.Br. (Gramineae). *Australian Journal of Botany*, **1**, 121–184.

Burbidge N (1959) *Notes on Plants and Plant Habitats Observed in the Abydos-Woodstock Area, Pilbara District, Western Australia*. CSIRO, Melbourne.

Burbidge N (1960a) Further notes on *Triodia* R.Br. (Gramineae) with description of five new species and one variety. *Australian Journal of Botany*, **8**, 381–395.

Burbidge N (1960b) The phytogeography of the Australian region. *Australian Journal of Botany*, **8**, 75–211.

Burbrink FT, Yao H, Ingrasci M *et al.* (2011) Speciation at the Mogollon Rim in the Arizona mountain kingsnake (*Lampropeltis pyromelana*). *Molecular Phylogenetics and Evolution*, **60**, 445–54.

Byrne M, Yeates DK, Joseph L *et al.* (2008) Birth of a biome: insights into the assembly and maintenance of the Australian arid zone biota. *Molecular Ecology*, **17**, 4398–4417.

Carolin R, Jacobs S, Vesk M (1973) The structure of the cells of the mesophyll and parenchymatous bundle sheath of the Gramineae. *Botanical Journal of the Linnean Society*, **66**, 259–275.

Carstens BC, Pelletier TA, Reid NM, Satler JD (2013) How to fail at species delimitation. *Molecular Ecology*, **22**, 4369–4383.

Catullo R, Doughty P, Roberts J, Keogh J (2011) Multi-locus phylogeny and taxonomic revision of *Uperoleia* toadlets (Anura: Myobatrachidae) from the western arid zone of Australia, with a description of a new species. *Zootaxa*, **43**, 1–43.

Chapman G (1996) *The Biology of Grasses*. CAB International, Wallingford, UK.

Clark P, Alley R, Pollard D (1999) Northern Hemisphere ice-sheet influences on global climate change. *Science*, **286**, 1104–1111.

Clayton W, Renvoize S (1986) *Genera Graminum: Grasses of the World*. H.M.S.O., London.

Cracraft J (1983) Species concepts and speciation analysis. *Current Ornithology*, **1**, 159–187.

Crisp MD, Arroyo MTK, Cook LG *et al.* (2009) Phylogenetic biome conservatism on a global scale. *Nature*, **458**, 754–756.

Crisp M, Cook L (2013) How was the Australian flora assembled over the last 65 million years? A molecular phylogenetic perspective. *Annual Review of Ecology, Evolution, and Systematics*, **44**, 303–324.

Crisp MD, Mant J, Toon A, Cook LG (2015) Australian spinifex grasses: new names in *Triodia* for *Monodia* and *Symplectrodia*. *Phytotaxa*, **230**, 293–296.

Daly BG, Dickman CR, Crowther MS (2008) Causes of habitat divergence in two species of agamid lizards in arid central Australia. *Ecology*, **89**, 65–76.

Davey JW, Hohenlohe PA, Etter PD *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews. Genetics*, **12**, 499–510.

Davis J, Nixon K (1992) Populations, genetic variation, and the delimitation of

phylogenetic species. *Systematic Biology*, **41**, 421–435.

Dawson W, Bennett A (1978) Energy metabolism and thermoregulation of the spectacled hare wallaby (*Lagorchestes conspicillatus*). *Physiological Zoology*, **51**, 114–130.

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

de Queiroz K (1998) The general lineage concept of species, species criteria, and the process of speciation. In: *Endless Forms: Species and Speciation* (eds Howard D, Berlocher S), pp. 57–75. Oxford University Press, Oxford.

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

de Queiroz K (2011) Branches in the lines of descent: Charles Darwin and the evolution of the species concept. *Biological Journal of the Linnean Society*, **103**, 19–35.

Department of Mines and Petroleum (2015) *Western Australian Mineral and Petroleum Statistics Digest 2014–15*. Government of Western Australia, East Perth.

Department of the Environment (2006) *MVG 20 - Hummock Grasslands*. Australian Government, Canberra, ACT.

Dickman CR, Letnic M, Mahon PS (1999) Population dynamics of two species of dragon lizards in arid Australia: the effects of rainfall. *Oecologia*, **119**, 357–366.

Domin K (1912) Additions to the flora of western and north-western Australia. *Journal of the Linnean Society. Botany*, **41**, 245–283.

Donoghue M (1985) A critique of the biological species concept and recommendations for a phylogenetic alternative. *Bryologist*, **88**, 172–181.

Ealey E, Bentley P, Main A (1965) Studies on water metabolism of the hill kangaroo, *Macropus robustus* (Gould), in northwest Australia. *Ecology*, **46**, 473–479.

Eaton DAR, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). *Systematic Biology*, **62**, 689–706.

Edwards SV, Xi Z, Janke A *et al.* (2016) Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. *Molecular Phylogenetics and Evolution*, **94**, 447–462.

Elshire RJ, Glaubitz JC, Sun Q *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLOS ONE*, **6**, e19379.

Erickson TE, Shackelford N, Dixon KW, Turner SR, Merritt DJ (2016) Overcoming physiological dormancy in seeds of *Triodia* (Poaceae) to improve restoration in the arid zone. *Restoration Ecology*.

Escudero M, Eaton DAR, Hahn M, Hipp AL (2014) Genotyping-by-sequencing as a

tool to infer phylogeny and ancestral hybridization: A case study in *Carex* (Cyperaceae). *Molecular Phylogenetics and Evolution*, **79**, 359–367.

Fujioka T, Chappell J, Fifield LK, Rhodes EJ (2009) Australian desert dune fields initiated with Pliocene-Pleistocene global climatic shift. *Geology*, **37**, 51–54.

Fujioka T, Chappell J, Honda M *et al.* (2005) Global cooling initiated stony deserts in central Australia 2–4 Ma, dated by cosmogenic 21Ne-10Be. *Geology*, **33**, 993–996.

Fujita MK, McGuire JA, Donnellan SC, Moritz C (2010) Diversification and persistence at the arid-monsoonal interface: Australia-wide biogeography of the Bynoe's gecko (*Heteronotia binoei*; Gekkonidae). *Evolution*, **64**, 2293–2314.

Gamage H, Mondal S, Wallis L *et al.* (2012) Indigenous and modern biomaterials derived from *Triodia* ('spinifex') grasslands in Australia. *Australian Journal of Botany*, **60**, 114–127.

Gardner C (1952) *Flora of Western Australia*. William H. Wyatt, Government Printer, Perth.

Giarla TC, Esselstyn JA (2015) The challenges of resolving a rapid, recent radiation: empirical and simulated phylogenomics of Philippine shrews. *Systematic Biology*, **64**, 727–740.

Gohli J, Leder EH, Garcia-Del-Rey E *et al.* (2015) The evolutionary history of Afrocanarian blue tits inferred from genomewide SNPs. *Molecular Ecology*, **24**, 180–191.

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

Griffin G (1984) Hummock grasslands. In: *Management of Australia's Rangelands* (eds Harrington G, Wilson A, Young M), pp. 271–284. CSIRO, East Melbourne.

Griffin GF (1990) Characteristics of three spinifex alliances in Central Australia. *Journal of Vegetation Science*, **1**, 435–444.

Griffin G (1992) Will it burn — should it burn? Management of the spinifex grasslands of inland Australia. In: *Desertified Grasslands: Their Biology and Management* (ed Chapman G), pp. 63–76. Academic Press, London.

Grigg A, Veneklaas E, Lambers H (2008) Water relations and mineral nutrition of *Triodia* grasses on desert dunes and interdunes. *Australian Journal of Botany*, **56**, 408–421.

Hackel E (1887) Gramineae. In: *Die Natürlichen Pflanzenfamilien* (eds Engler A, Prantl K), p. 68. W. Engelmann, Leipzig.

Harbaugh DT, Baldwin BG (2007) Phylogeny and biogeography of the sandalwoods (*Santalum*, Santalaceae): repeated dispersals throughout the Pacific. *American Journal of Botany*, **94**, 1028–1040.

Hattersley PW, Watson L (1976) C_4 Grasses: an anatomical criterion for

distinguishing between NADP-malic enzyme species and PCK or NAD-malic enzyme species. *Australian Journal of Botany*, **24**, 297–308.

Hewitt G (1999) Post-glacial re- colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87–112.

Hitchcock A (1920) *The Genera of Grasses of the United States: with Special Reference to the Economic Species*. United States Department of Agriculture, Washington, D.C.

Hodkinson TR, Chase MW, Renvoize SA (2002) Characterization of a genetic resource collection for *Miscanthus* (Saccharinae, Andropogoneae, Poaceae) using AFLP and ISSR PCR. *Annals of Botany*, **89**, 627–636.

Hubbard C (1937) *Triodia pungens* R. Br. *Hooker's Icones Plantarum*, **34**, t. 3336.

Hurry CR, Walsh NG, Murphy DJ (2012) A taxonomic review of *Triodia bunicola* and *T. scariosa* (Poaceae: Chloridoideae), based on morphological and molecular data. *Australian Systematic Botany*, **25**, 304–312.

Jabaily RS, Shepherd KA, Gardner AG *et al.* (2014) Historical biogeography of the predominantly Australian plant family Goodeniaceae. *Journal of Biogeography*, **41**, 2057–2067.

Jacobs S (1971) Systematic position of the genera *Triodia* R.Br. and *Plectrachne* Henr. (Gramineae). *Proceedings of the Linnean Society of New South Wales*, **96**, 175–185.

Jacobs S (1980) Pyric cycles in *Triodia* mallee. In: *Aeolian Landscapes in the Semi-Arid Zone of South Eastern Australia* (eds Storrier R, Stannard M), pp. 145–146. Australian Society of Soil Science, Inc., Riverina Branch.

Jacobs S (1992) Spinifex (*Triodia*, *Plectrachne*, *Symplectrodia* and *Monodia*: Poaceae) in Australia. In: *Desertified Grasslands: Their Biology and Management* (ed Chapman G), pp. 47–62. Academic Press, London.

Jacobs S (2004) The tribe Triodieae (Chloridoideae: Gramineae). *Telopea*, **10**, 701–703.

Kadereit G, Freitag H (2011) Molecular phylogeny of Camphorosmeae (Camphorosmoideae, Chenopodiaceae): implications for biogeography, evolution of C_4-photosynthesis and taxonomy. *Taxon*, **60**, 51–78.

Kearney M, Blacket MJ, Strasburg JL, Moritz C (2006) Waves of parthenogenesis in the desert: evidence for the parallel loss of sex in a grasshopper and a gecko from Australia. *Molecular Ecology*, **15**, 1743–1748.

Kitchener DJ, Stoddart J, Henry J (1983) A taxonomic appraisal of the genus *Ningaui* Archer (Marsupialia: Dasyuridae), including description of a new species. *Australian Journal of Zoology*, **31**, 361–379.

Lazarides M (1997) A revision of *Triodia* including *Plectrachne* (Poaceae, Eragrostideae, Triodiinae). *Australian Systematic Botany*, **10**, 381–489.

Lazarides M, Weiller C, McCusker A (2005) *Triodia* (K Mallett, Ed,). *Flora of Australia*, **44B**, 203–256.

Leaché AD, Chavez AS, Jones LN *et al.* (2015) Phylogenomics of phrynosomatid lizards: conflicting signals from sequence capture versus restriction site associated DNA sequencing. *Genome Biology and Evolution*, **7**, 706–719.

Leaché AD, Fujita MK, Minin VN, Bouckaert RR (2014) Species delimitation using genome-wide SNP Data. *Systematic Biology*, **63**, 534–542.

Leaché AD, Koo MS, Spencer CL *et al.* (2009) Quantifying ecological, morphological, and genetic variation to delimit species in the coast horned lizard species complex (*Phrynosoma*). *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 12418–12423.

Losos J (1988) Thermoregulatory correlates of escape behavior by a desert lizard, *Ctenophorus isolepis*. *Journal of Herpetology*, **22**, 353–356.

Macphail M (2007) *Australian Palaeoclimates: Cretaceous to Tertiary: a Review of Palaeobotanical and Related Evidence to the Year 2000*. Bentley, Western Australia.

Maddison W (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523–536.

Mant J, Bayer R, Crisp M, Trueman J (2000) A phylogeny of Triodieae (Poaceae: Chloridoideae) based on the ITS region of nrDNA: testing conflict between anatomical and inflorescence characters. In: *Grasses: Systematics and Evolution* (eds Jacobs S, Everett J), pp. 213–217. CSIRO, Melbourne.

Martin H (2006) Cenozoic climatic change and the development of the arid vegetation in Australia. *Journal of Arid Environments*, **66**, 533–563.

Mayr E (1942) *Systematics and the Origin of Species from the Viewpoint of a Zoologist*. Columbia University Press, New York.

McCormack J, Faircloth B (2013) Next-generation phylogenetics takes root. *Molecular Ecology*, **22**, 19–21.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**, 526–538.

McWilliam J, Mison K (1974) Significance of the C_4 pathway in *Triodia irritans* (spinifex), a grass adapted to arid environments. *Australian Journal of Plant Physiology*, **1**, 171–175.

Miralles A, Vences M (2013) New metrics for comparison of taxonomies reveal striking discrepancies among species delimitation methods in *Madascincus* lizards. *PLOS ONE*, **8**, e68242.

Mishler B (1985) The morphological, developmental, and phylogenetic basis of species concepts in bryophytes. *Bryologist*, **88**, 207–214.

Mondal S, Memmott P, Wallis L, Martin D (2012) Physico-thermal properties of spinifex resin bio-polymer. *Materials Chemistry and Physics*, **133**, 692–699.

Morrison B, Lamb D, Hundloe T (2005) Assessing the likelihood of mine site revegetation success: A Queensland case study. *Australasian Journal of Environmental Management*, **12**, 165–182.

Morton S, James S (1988) The diversity and abundance of lizards in arid Australia: a new hypothesis. *American Naturalist*, **132**, 237–256.

Mudelsee M, Stattegger K (1997) Exploring the structure of the mid-Pleistocene revolution with advanced methods of time-series analysis. *Geologische Rundschau*, **86**, 499–511.

Mueller UG, Wolfenbarger LL (1999) AFLP genotyping and fingerprinting. *Trends in Ecology and Evolution*, **14**, 389–394.

Mummenhoff K, Linder P, Friesen N *et al.* (2004) Molecular evidence for bicontinental hybridogenous genomic constitution in *Lepidium sensu stricto* (Brassicaceae) species from Australia and New Zealand. *American Journal of Botany*, **91**, 254–261.

Naciri Y, Linder HP (2015) Species delimitation and relationships: The dance of the seven veils. *Taxon*, **64**, 3–16.

Nadeau NJ, Martin SH, Kozak KM *et al.* (2013) Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Molecular Ecology*, **22**, 814–826.

Nicotra AB, Chong C, Bragg JG *et al.* (2016) Population and phylogenomic decomposition via genotyping-by-sequencing in Australian *Pelargonium*. *Molecular Ecology*, **25**, 2000–2014.

Nixon K, Wheeler Q (1990) An amplification of the phylogenetic species concept. *Cladistics*, **6**, 211–223.

O'Neill EM, Schwartz R, Bullock CT *et al.* (2013) Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (*Ambystoma tigrinum*) species complex. *Molecular Ecology*, **22**, 111–129.

Padial JM, Aurélien M, De la Riva I, Vences M (2010) The integrative future of taxonomy. *Frontiers in Zoology*, **7**, 16.

Pante E, Abdelkrim J, Viricel A *et al.* (2014) Use of RAD sequencing for delimiting species. *Heredity*, **114**, 450–459.

Paterson H (1985) The recognition concept of species. In: *Species and Speciation* (ed Vrba E), pp. 21–29. Transvaal Museum, Pretoria.

Pepper M, Doughty P, Arculus R, Keogh JS (2008) Landforms predict phylogenetic structure on one of the world's most ancient surfaces. *BMC Evolutionary Biology*, **8**, 152.

Pepper M, Doughty P, Fujita MK, Moritz C, Keogh JS (2013a) Speciation on the rocks: Integrated systematics of the *Heteronotia spelea* species complex (Gekkota; Reptilia) from western and central Australia (D Fontaneto, Ed,). *PLOS ONE*, **8**, e78110.

Pepper M, Doughty P, Keogh JS (2013b) Geodiversity and endemism in the iconic Australian Pilbara region: a review of landscape evolution and biotic response in an ancient refugium. *Journal of Biogeography*, **40**, 1225–1239.

Pepper M, Ho SYW, Fujita MK, Keogh JS (2011) The genetic legacy of aridification: climate cycling fostered lizard diversification in Australian montane refugia and left low-lying deserts genetically depauperate. *Molecular Phylogenetics and Evolution*, **61**, 750–759.

Peterson PM, Romaschenko K, Johnson G (2010) A classification of the Chloridoideae (Poaceae) based on multi-gene phylogenetic trees. *Molecular Phylogenetics and Evolution*, **55**, 580–598.

Pisias N, Moore TJ (1981) The evolution of Pleistocene climate: a time series approach. *Earth and Planetary Science Letters*, **52**, 450–458.

Potter S, Bragg JG, Peter BM, Bi K, Moritz C (2016) Phylogenomics at the tips: inferring lineages and their demographic history in a tropical lizard, *Carlia amax*. *Molecular Ecology*, **25**, 1367–1380.

Prendergast H, Hattersley P, Stone N (1987) New structural/biochemical associations in leaf blades of C_4 grasses (Poaceae). *Australian Journal of Plant Physiology*, **14**, 403–420.

Pritzel E (1918) Species novae ex Australia centrali. *Repertorium novarum specierum regni vegetabilis*, **15**, 356–361.

Pyron RA, Hendry CR, Chou VM *et al.* (2014) Effectiveness of phylogenomic data and coalescent species-tree methods for resolving difficult nodes in the phylogeny of advanced snakes (Serpentes: Caenophidia). *Molecular Phylogenetics and Evolution*, **81**, 221–231.

Qi ZC, Yu Y, Liu X *et al.* (2015) Phylogenomics of polyploid *Fothergilla* (Hamamelidaceae) by RAD-tag based GBS—insights into species origin and effects of software pipelines. *Journal of Systematics and Evolution*, **53**, 432–447.

Renvoize S (1983) A survey of leaf-blade anatomy in grasses IV. Eragrostideae. *Kew Bulletin*, **38**, 469–478.

Rice B, Westoby M (1999) Regeneration after fire in *Triodia* R. Br. *Australian Journal of Ecology*, **24**, 563–572.

Sakaguchi S, Bowman DMJS, Prior LD *et al.* (2013) Climate, not Aboriginal landscape burning, controlled the historical demography and distribution of fire-sensitive conifer populations across Australia. *Proceedings of the Royal Society B. Biological Sciences*, **280**, 20132182.

Sanger F, Nicklen S, Coulson A (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, **74**, 5463–5467.

Schlick-Steiner BC, Steiner FM, Seifert B *et al.* (2010) Integrative taxonomy: a multisource approach to exploring biodiversity. *Annual Review of Entomology*, **55**, 421–438.

Shepherd K, Waycott M, Calladine A (2004) Radiation of the Australian Salicornioideae (Chenopodiaceae)—based on evidence from nuclear and chloroplast DNA sequences. *American Journal of Botany*, **91**, 1387–1397.

Sites JW, Marshall JC (2004) Operational criteria for delimiting species. *Annual Review of Ecology, Evolution, and Systematics*, **35**, 199–227.

Sneath P, Sokal R (1973) *Numerical Taxonomy: the Principles and Practices of Numerical Classification*. W. H. Freeman, San Francisco.

Soltis DE, Albert VA, Leebens-Mack J *et al.* (2009) Polyploidy and angiosperm diversification. *American Journal of Botany*, **96**, 336–348.

Toon A, Crisp MD, Gamage H *et al.* (2015) Key innovation or adaptive change? A test of leaf traits using Triodiinae in Australia. *Scientific Reports*, **5**, 12398.

Van den Borre A, Watson L (1997) On the classification of the Chloridoideae (Poaceae). *Australian Systematic Botany*, **10**, 491–531.

Van Valen L (1976) Ecological species, multispecies, and oaks. *Taxon*, **25**, 233–239.

Vos P, Hogers R, Bleeker M *et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.

Wagner CE, Keller I, Wittwer S *et al.* (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, **22**, 787–798.

Walker J, Geissman J, Bowring S, Babcock L (2012) *Geological Time Scale v. 4.0.* Geological Society of America.

Watson L, Dallwitz MJ (1992) The grass genera of the world: descriptions, illustrations, identification, and information retrieval; including synonyms, morphology, anatomy, physiology, phytochemistry, cytology, classification, pathogens, world and local distribution, and references. http://delta-intkey.com/grass/.

Wells G, Dixon K (1999) *Post-mining Restoration of Spinifex Species in the Arid-tropical Zone*. East Perth, WA.

Wells G, Zawko G, Rossetto M, Dixon K (2003) A molecular approach to provenance delineation for the restoration of hummock grasslands (*Triodia* spp.) in arid-tropical Australia. *Ecological Management & Restoration*, **4**, S60–S68.

Westoby M, Rice B, Griffin G, Friedel M (1988) The soil seed bank of *Triodia basedowii* in relation to time since fire. *Australian Journal of Ecology*, **13**, 161–169.

Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology*, **54**, 844–851.

Williams MAJ (1984) Cenozoic evolution of arid Australia. In: *Arid Australia: Proceedings of a Symposium on the Origins, Biota and Ecology of Australia's Arid Regions* (eds Cogger HG, Cameron EE), pp. 59–78. Australian Museum, Sydney.

Winkworth R (1967) The composition of several arid spinifex grasslands of central Australia in relation to rainfall, soil water relations, and nutrients. *Australian Journal of Botany*, **15**, 107–130.

Wright BBR, Clarke PJP (2007) Fire regime (recency, interval and season) changes the composition of spinifex (*Triodia* spp.)-dominated desert dunes. *Australian Journal of Botany*, **55**, 709–724.

Yeates DK, Seago A, Nelson L *et al.* (2011) Integrative taxonomy, or iterative taxonomy? *Systematic Entomology*, **36**, 209–217.

Chapter 2.

Untangling a species complex of arid zone grasses (*Triodia*) reveals patterns congruent with co-occurring animals

Abstract

The vast Australian arid zone formed over the last 15 million years, and gradual aridification as well as more extreme Pliocene and Pleistocene climate shifts have impacted the evolution of its biota. Understanding the evolutionary history of groups of organisms or regional biotas such as the Australian arid biota requires clear delimitation of the units of biodiversity (taxa). Here we integrate evidence from nuclear (ETS and ITS) and chloroplast (*rps16-trnK* spacer) regions and morphology to clarify taxonomic boundaries in a species complex of Australian hummock grasses (*Triodia*) to better understand the evolution of Australian arid zone plants and to evaluate congruence in distribution patterns with co-occurring organisms. We find evidence for multiple new taxa in the *T. basedowii* species complex, but also incongruence between data sets and indications of hybridization that complicate delimitation. We find that the *T. basedowii* complex has high lineage diversity and endemism in the biologically important Pilbara region of Western Australia, consistent with the region acting as a refugium. Taxa show strong geographic structure in the Pilbara, congruent with recent work on co-occurring animals and suggesting common evolutionary drivers across the biota. Our findings confirm recognition of the Pilbara as an important centre of biodiversity in the Australian arid zone, and provide a basis for future taxonomic revision of the *T. basedowii* complex and more detailed study of its evolutionary history and that of arid Australia.

Introduction

Australia is dominated by an arid zone that extends over more than 60% of the continent. This arid zone formed over the past *c*. 15 Ma, with significant impacts from more recent Pleistocene (2.6 Ma to 10 ka; Walker *et al.* 2012) climate instability (reviewed in Byrne *et al.* 2008). In Australia, Pleistocene glacial cycles were marked by cooler and drier conditions rather than the ice sheets that characterized the northern hemisphere (Williams 1984). The formation of the arid zone, including the more intense Pleistocene aridification, is likely to have had a marked effect on the evolution of organisms that currently inhabit this region.

An important part of the development of the Australian arid zone was the formation of sandy dune fields, which began to form *c.* 1 Ma (Fujioka *et al.* 2009), coinciding with a global shift to lower frequency, higher amplitude glacial cycles roughly 0.8–1.2 Ma (Pisias & Moore 1981; Mudelsee & Stattegger 1997; Clark *et al.* 1999). These newly formed dune fields may have eliminated existing habitat for some organisms and created new habitat for others to expand into (e.g. geckos; Fujita *et al.* 2010; Pepper *et al.* 2011). It is not yet clear how plants were impacted by the formation of these dunes and associated aridification. The remarkable stability of the dune fields since their formation could be partly explained by persistent vegetation and/or soil particle binding (Hesse 2011), but dust (Hesse *et al.* 2004) and pollen (Martin 2006) records indicate some dune activity and more open vegetation during glacial maxima, suggesting some plants were unable to persist.

It has been hypothesised that mesic refugia in the Australian arid zone could have allowed local persistence of organisms through climate changes and served as a safe retreat for currently widespread taxa during drier periods associated with glacial maxima (e.g. Beadle 1981a; Byrne 2008). The semi-arid Pilbara region of Western Australia (WA), with its complex topography and proximity to the coast, has been hypothesized to have functioned as one such arid zone refugium (Burbidge 1959; Byrne *et al.* 2008). Consistent with the refugium hypothesis, recent work on geckos (Pepper *et al.* 2011) indicates that in topographically complex areas such as the Pilbara, extant genetic diversity is greater and coalescent times older than in

surrounding sandy deserts. Also consistent with this hypothesis, the Pilbara has been recognized for its rich biodiversity and strong phylogeographic patterns that set it apart from surrounding regions (reviewed in Pepper *et al.* 2013b), and as a centre of endemism (e.g. Halse *et al.* 2014).

The Pilbara biogeographic region (Department of the Environment 2012) is divided into four subregions with distinctive substrates: sandy to clayey coastal plains (Roebourne), granite and greenstone plains and basaltic ranges (Chichester), sandy and clayey alluvial plains (Fortescue), and mountainous sedimentary and volcanic rock with banded ironstone formations and skeletal soils (Hamersley) (McKenzie *et al.* 2009) (see Fig. 2.1). Surveys of Pilbara biodiversity and more detailed studies have revealed strong geographic structure in frogs (Catullo *et al.* 2011), dragon lizards (Doughty *et al.* 2012), pebble-mimic lizards (Shoo *et al.* 2008) and geckos (Pepper *et al.* 2008, 2013a; Doughty *et al.* 2010), as well as in beetle (Guthrie *et al.* 2010) and spider (Durrant *et al.* 2010) species assemblages. The main phylogeographic patterns seen in the fauna are north-south differentiation between the Chichester and Hamersley subregions (separated by the Fortescue River valley) and between the Pilbara itself and the sandplains to the west (Pepper *et al.* 2013b). Detailed genetic studies of Pilbara plants are currently lacking but are critical for insight into the evolution of the region and the Australian arid zone, especially in comparison with recent work on animals.

Perennial hummock grasses of the genus *Triodia* R.Br. (Poaceae: Chloridoideae) are major and structurally dominant components of the vegetation across much of arid and semi-arid Australia including the Pilbara, covering more than 18% of the continent (Department of the Environment and Water Resources 2007) where mean annual precipitation is typically <350 mm and soils are infertile (Beadle 1981b; Griffin 1990). *Triodia* species are commonly known as "spinifex," and are notorious for their tough, needle-like rolled leaf blades. They are ecologically important and provide food and habitat for a range of animals including termites, which in turn support one of the most diverse lizard assemblages on Earth (Pianka 1981; Morton & James 1988; Colli *et al.* 2006). *Triodia* currently comprises 73 described species

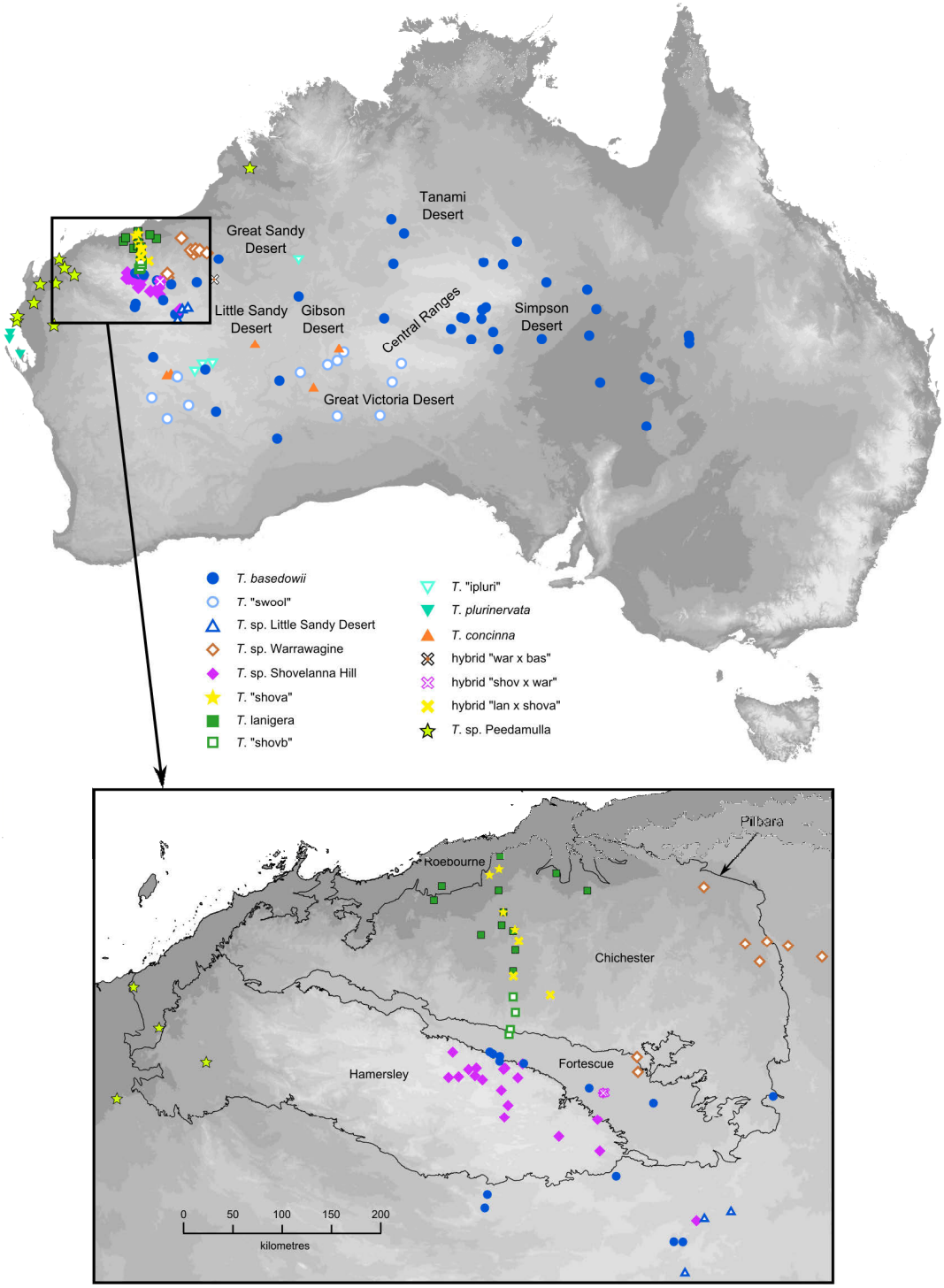

Figure 2.1. Geographic distribution of samples from the *T. basedowii* species complex used to sequence the ITS region. Symbols denote putative taxa and individuals with ITS copies from two distinct taxa. The Pilbara region is indicated in the enlarged area with the IBRA subregions labelled, and the central deserts and central ranges are also labelled.

endemic to Australia (Lazarides 1997; Barrett *et al.* 2005; Armstrong 2008; Barrett & Barrett 2011, 2015; Hurry *et al.* 2012; Crisp *et al.* 2015). Based on leaf anatomy and the production of resin, Burbidge (1953) recognized two morphological groups of spinifex: "hard" spinifexes with highly sclerified and pungent leaves, and typically without resin; and "soft" spinifexes with less sclerified and less pungent leaves, and typically with abundant resin. Soft spinifexes (with one possible exception) form a clade nested within a paraphyletic assemblage of hard species (Mant *et al.* 2000; Toon *et al.* 2015). A recent dated phylogeny of *Triodia* (Toon *et al.* 2015) suggests that its ancestors probably arrived in Australia *c*. 24–14 Ma and radiated extensively during Miocene aridification. *Triodia* is thus a key genus for understanding the evolution of Australian arid zone plants.

Triodia still poses significant taxonomic challenges, however, especially in poorly explored areas and for broadly applied names that likely encompass multiple species. This taxonomic impediment affects our understanding of evolutionary history in the group and restricts our ability to make broader comparisons, which have the potential to reveal evolutionary drivers affecting entire biomes (Bermingham & Moritz 1998; Carstens *et al.* 2005; Arbogast & Kenagy 2008).

One group of *Triodia* species that exemplifies this taxonomic challenge is the *Triodia basedowii* E.Pritz. species complex. *Triodia basedowii* and related species *T. lanigera* Domin, *T. plurinervata* N.T.Burb., *T. concinna* N.T.Burb., *T. rigidissima* (Pilg.) Lazarides and *T. desertorum* (C.E.Hubb.) Lazarides form a clade (see Toon *et al.* 2015) distinguished from other hard spinifexes by their many-nerved glumes. Within the clade, *T. rigidissima* and *T. desertorum* are morphologically distinct in having long awns on their lemmas (*vs*. no awns) and were originally placed in a separate genus (*Plectrachne*) that is now synonymous with *Triodia* (Lazarides 1997). Previous morphological treatments (Lazarides 1997; Lazarides *et al.* 2005) placed *T. plurinervata* and *T. concinna* in a separate informal species group from *T. basedowii* and *T. lanigera*, the former two taxa having shortly lobed (*vs*. deeply lobed) lemmas. The *T. basedowii* species complex comprises *T. basedowii* and *T. lanigera* along with four informally named taxa (Western Australian Herbarium 1998): *T*. sp. Shovelanna Hill (S. van Leeuwen 3835), *T*. sp. Little Sandy Desert (S. van Leeuwen

4935), *T.* sp. Peedamulla (A.A. Mitchell PRP1636) and *T.* sp. Warrawagine (A.L. Payne PRP 1859). *T. lanigera* is commonly differentiated from *T. basedowii* based on the relative length and shape of its lemma lobes (typically having a proportionately longer midlobe). *T.* sp. Shovelanna Hill is distinct from *T. basedowii* in its smaller habit with shorter leaves and in its often unbranched inflorescence. *T.* sp. Warrawagine is morphologically similar to *T.* sp. Shovelanna Hill, but has a branched inflorescence with a greater number of spikelets. *T.* sp. Peedamulla differs from *T. lanigera* in its glabrous (*vs*. woolly) leaf sheaths and orifices, and in its shorter glumes and less hairy lemma midlobes. *T.* sp. Little Sandy Desert differs from *T. basedowii* in its glabrous (*vs*. woolly) leaf orifices and in the shape of its spikelets and textures of its lemmas (more induration in the lobes). Members of the complex occur throughout much of central and western Australia, including the Pilbara region.

Due to overlapping morphological variation and variability within currently recognized species, taxa in the *T. basedowii* complex are challenging to differentiate from each other. Wide morphological variation across specimens recognized as *T. basedowii* led Lazarides et al. (2005) to suggest that taxonomic resolution within *T. basedowii* would require assessment of genetic variation. Given the challenging nature of this species complex and the limited resolution from morphological evidence, additional sources of evidence are needed to delimit taxa ("integrative taxonomy"; Dayrat 2005; Padial *et al.* 2010; Schlick-Steiner *et al.* 2010). Accurately resolving units of biodiversity (taxonomy) in this complex will allow broader comparisons of distribution patterns across the biota.

This study seeks to clarify taxonomic boundaries in the *T. basedowii* complex using integrated morphological and molecular evidence, and to compare geographic patterns of taxon distribution and genetic partitioning in this plant group with those found in previous studies of co-occurring animals. We assess whether our results are consistent with the hypothesis of the Pilbara as a refugium and what they suggest about the impacts of geologically recent climate changes on Australian arid zone plants.

Methods

Taxon sampling and a priori *identification*

Putative taxa were identified *a priori* primarily based on morphological differences observed during a survey of herbarium material from across the distribution of the complex and on field trips to the Pilbara and central WA. We refer to six of these putative taxa using informal names: *T.* "shova" is similar to *T. lanigera* but with relatively unbranched inflorescences, glabrous leaf sheaths and short leaves; *T.* "shovb" is similar to *T. lanigera* but with short and long-leaved plants and a glabrous lemma midlobe; *T.* "swool" is similar to *T. basedowii* but has densely woolly leaf sheaths (*vs*. usually sparsely woolly to glabrous); *T.* "wcoast" is similar to *T.* sp. Peedamulla but with smaller spikelets; *T.* "broad" is similar to *T. basedowii* but has broadened spikelets with indurated lemma lobes; and *T.* "ipluri" is similar to *T. plurinervata* but with larger spikelets and a disjunct inland distribution. A total of 153 collections from the *T. basedowii* complex plus 19 collections from outgroups were utilized across data sets (Table S2.1), including both herbarium and field material.

Sampling of both field and herbarium material was an iterative and exploratory process. Discovery of new variations in the field led to broader sampling and influenced the refinement of herbarium piles. Due to resource limitations, most populations are represented by a single genetic sample. Herbarium sampling and measurement of specimens was undertaken attempting to capture the range of variation within putative taxa across their geographic distribution, subject to available material and the restrictions of time and effort. As a result, not all available specimens were measured (especially in the case of the widespread *T. basedowii*).

Molecular sampling

Leaf and/or inflorescence fragments were removed for DNA extraction from 90 herbarium specimens and fresh leaf material collected into silica gel was obtained from 66 field collections (Table S2.2). Field specimens of four other *Triodia* species (*T. danthonioides* (F.Muell.) Lazarides, *T. pungens* R.Br. (a soft spinifex), *T.*

wiseana C.A.Gardner, and *T. rigidissima* (Pilg.) Lazarides) were also included. Specimens of *Aeluropus lagopoides* (L.) Trin. ex Thwaites (two specimens), *Orinus kokonorica* (K.S. Hao) Tzvelev and *Orinus thoroldii* (Stapf ex Hemsl.) Bor were used as outgroups for *Triodia* based on Peterson *et al.* (2010).

DNA extraction, amplification and sequencing

Approximately 500 mg of leaf and/or inflorescence material was ground in liquid nitrogen, and genomic DNA extracted using a CTAB extraction protocol (Doyle & Dickson 1987) modified with Carlson extraction buffer (Csaikl *et al.* 1998), a 15 minute 65°C heated stage, an equal volume of isopropanol, a 20–30 minute -20°C precipitation step, a straight 70% ethanol wash, and omission of RNase at the penultimate step. We targeted the nuclear internal transcribed spacers (ITS1 and ITS2) including the 5.8S subunit of ribosomal DNA (rDNA), and a portion of the external transcribed spacer (ETS) of the rDNA repeats. This region was easily amplified from both field and herbarium material, had appreciable variation within and between entities, and in most cases showed evidence of only a single ribotype, which made Sanger sequencing (Sanger *et al.* 1977) possible without cloning.

While the ITS region of rDNA has been a mainstay of recent phylogenetic studies on plant groups, there are concerns over determining orthology for copies sequenced within and between species (Álvarez & Wendel 2003; Nieto Feliner & Rosselló 2007). Low copy nuclear markers, a recommended alternative to ITS for nuclear phylogenetic inference, may face similar challenges in determining orthology and identifying pseudogenes (Nieto Feliner & Rosselló 2007). We attempted to use the xanthine dehydrogenase gene (Morton 2011) as a low copy nuclear region, but did not include it in our analyses due to difficulty amplifying it from herbarium material and the presence of many heterozygous positions in some samples, indicative of multiple alleles or paralogs.

We also targeted the chloroplast *rps16-trnK* spacer to compare with the nuclear marker. This region was recently used in constructing a phylogeny of the subfamily Chloridoideae to which *Triodia* belongs (Peterson *et al.* 2010).

Target regions were amplified using polymerase chain reaction (PCR) on either an Applied Biosystems® Veriti®, GeneAmp® PCR System 9700, or BIOER: GenePro thermal cycler with primers and conditions as shown in Table 2.1. Because many of the samples came from herbarium specimens, PCR reactions were nested to recover products from low-concentration DNA. In some cases, failure to amplify entire regions (such as ETS and the *rps16-trnK* spacer) using the full-length primer pair led us to design primers that split the regions into two smaller parts that were then recoverable from many of the herbarium samples. Products were verified and roughly quantified on an agarose gel before being cleaned with 1 µL ExoSAP-IT® (Affymetrix USB) per 10 µL PCR product.

Sequencing reactions were done primarily using BigDye® Terminator v. 1.1/3.1 chemistry (Applied Biosystems) with the same primers as were used for the second round of the nested PCR amplification, using the following PCR protocol: 96°C for 1 min, 25×(96°C for 10 sec, 50°C for 5 sec, 60°C for 4 min). Some preliminary sequencing was outsourced to Macrogen (Korea), but most was carried out on an in-lab Applied Biosystems® 3500 Genetic Analyzer, in both the forward and reverse directions. Chromatograms of contigs were first scored automatically then corrected manually and combined into sequences in Geneious® v. 6.1.7 (http://www.geneious.com/; Kearse *et al.* 2012). In samples with heterozygous positions, two or more copies could sometimes be recovered manually (due to a length difference); if not, the positions were scored using ambiguity codes.

Phylogenetic analyses

We analysed four datasets: (1) 181 ITS sequences from 164 samples (multiple copies were recovered from some samples), (2) 73 ETS sequences from 62 samples, (3) 75 concatenated ETS+ITS sequences from 62 samples, and (4) 134 *rps16-trnK* spacer sequences from 134 samples. Specimen representation across these datasets was unequal (Table S2.2) due to unsuccessful sequencing from some specimens. The ITS dataset included two divergent copies from some individuals, designated putative hybrids, but these were excluded from the concatenated ETS+ITS dataset. Sequences were aligned automatically in Geneious® v. 6.1.7 (http://www.geneious.com/;

Table 2.1. Primers used for PCR and sequencing, and PCR conditions for each of the target regions. Primers are listed as the nested three for the full region followed by those designed to amplify subsections of the region. PCR conditions for the *rps16-trnK* spacer differed between full length and partial region amplification.

Region	Primers	Direction	Sequence (5'–3')	References	PCR conditions
ITS (ITS1, 5.8S rDNA, ITS2)	ITS5A	Forward	CCTTATCATTTAGAGGAAGGAG	(Wurdack in Stanford *et al.* 2000)	95℃ for 3 min, 37×(95℃ for 30 sec, 48℃ for 30 sec, 72℃ for 1.5 min), 72℃ for 10 min
	ITSL	Forward	TCGTAACAAGGTTTCCGTAGGTG	(Hsiao *et al.* 1994)	
	ITS4	Reverse	TCCTCCGCTTATTGATATGC	(White *et al.* 1990)	
ETS	ETSall F1	Forward	CTGCCTGGGGCTCAGTGGGR	this study	95℃ for 3 min, 37 or 40×(95℃ for 30 sec, 48℃ for 30 sec, 72℃ for 1.5 min), 72℃ for 10 min
	ETSall F2n	Forward	TGCCTGGGGCTCAGTGGGRT	this study	
	18S-R	Reverse	AGACAAGCATATGACTACTGGCAGG	(Starr *et al.* 2003)	
	ETSall Int-R	Reverse	ACGCCCGAGCATCCCATCGT	this study	
	ETSshort F1	Forward	CGTGTCACCGGCGAGAAGCAG	this study	
	ETSshort R	Reverse	CTTTGAGACAAGCATATGACTACTGG	this study	
rps16-trnK spacer	rpS16-900F	Forward	TATCGAATCGTTGCAATTGATG	(Peterson *et al.* 2010)	80℃ for 5 min, 35×(95℃ for 1 min, 50℃ for 1 min, 65℃ for 4 min)
	trnK(-rps16)out_n	Forward	CTCTACCATTGAGTTAGCAACCCAG	this study	
	3914PR	Reverse	CATTGAGTTAGCAACCCAGATA	(Peterson *et al.* 2010)	
	rps16-tKA F1	Forward	TCGAATCGTTGCAATTGATG	this study	95℃ for 2 min, 40×(95℃ for 30 sec, 51℃ for 1 min, 72℃ for 1.5 min), 72℃ for 7 min
	rps16-tKA R1	Reverse	ACCTAATAGGGAGTGACGAT	this study	
	rps16-tKA R2n	Reverse	AGGGAGTGACGATAAAGAGA	this study	
	rps16-tKB F1	Forward	TGGTTTTCTTACCATTGTATTTCT	this study	
	rps16-tKB R1	Reverse	TACCATTGAGTTAGCAACCC	this study	
	rps16-tKB R2n	Reverse	CATTGAGTTAGCAACCCAGA	this study	

Kearse *et al.* 2012) then refined manually. Gaps were coded using simple indel coding (Simmons & Ochoterena 2000) as implemented in the program SeqState (Müller 2005). Gaps were included as restriction characters in Bayesian analyses and as binary characters in maximum likelihood and maximum parsimony analyses.

We used three methods of phylogenetic inference for the nuclear data: maximum parsimony (MP) in PAUP* v. 4.0b10 (Swofford 2002), maximum likelihood (ML) in RAxML v. 7.8.7 (Stamatakis 2006), and Bayesian inference (BI) in MrBayes v. 3.2.1 (Ronquist *et al.* 2012). Tests for partitioning schemes for ML and BI were run in PartitionFinder v. 1.1.0 (Lanfear *et al.* 2012) using the Bayesian Information Criterion. Optimum partitioning for the ITS dataset was determined as ITS1+ITS2 under SYM+G and 5.8S rDNA under K2P, while for the concatenated ETS+ITS dataset it was ITS1+ITS2+ETS under GTR+G and 5.8S rDNA under JC. The selected models are similar between the two datasets, and differences may be due to the exclusion of the more divergent outgroups (*Aeluropus* and *Orinus*) from the ETS+ITS dataset. The ETS+ITS tree was rooted on *T. danthonioides*, the most divergent shared outgroup in the ITS tree (albeit with low support).

PAUP* was run using 1000 random addition replicates, keeping 10 trees per replicate and using tree bisection-reconnection to swap, followed by filtering for most parsimonious trees at the end of the search. To obtain support values we ran 500 bootstrap heuristic searches using 50 random addition replicates, keeping two trees per replicate. RAxML was run using 50 searches for the maximum likelihood tree. Since RAxML only implements GTR-based models, partitions were run under the more general GTR+G. A third partition was added for the binary indel characters. To obtain support values, we ran 500 bootstrap replicates. MrBayes was run using three concurrent runs of 4,000,000 generations. Partitions were run under the models found using PartitionFinder, and an additional partition was created for the restriction indel characters and run under the 'coding=variable' option. Each run consisted of four chains and the default temperature setting, attempting three swaps every five generations. Samples were taken from the MCMC chain every 500 generations. When summarizing parameters and trees, 25% of the samples were discarded as burn-in. We used Tracer v. 1.6 (Rambaut *et al.* 2014) to evaluate the

three runs, which all converged on the same log likelihood score, appeared to have reached stationarity, and had effective sample sizes greater than 499 for every parameter. Trees were visualized and edited using FigTree v. 1.4.0 and v. 1.4.2 (Rambaut 2014) and Inkscape™ (The Inkscape Team 2012).

We used statistical parsimony (Templeton *et al.* 1992) as implemented in the program TCS (Clement *et al.* 2000) to analyse the chloroplast *rps16-trnK* spacer sequences. This approach may be better than MP for addressing intraspecific relationships, where there is often low variation in molecular markers (Clement *et al.* 2000). The gap-coded *rps16-trnK* spacer sequence alignment created with SeqState was manually edited (changing the gap codes "1" to "A" and "0" to "C") then reformatted for TCS using ALTER (Glez-Peña *et al.* 2010). When running TCS, gaps were treated as missing data (since they had already been coded). A variable portion of the *rps16-trnK* spacer consisting of a cytosine repeat (4–9 bp) was removed from the analysis after a preliminary analysis (results not shown) in PAUP* suggested it was homoplasious (i.e. lower bootstrap support values for clades in runs with compared to without the variable region). The *Aeluropus* sequences could not be connected to the network with the default 0.95 parsimony probability connection limit, so were excluded. Networks were edited graphically in Inkscape™ (The Inkscape Team 2012), with haplotype circles adjusted to make area proportional to sample size.

Morphology

Morphological measurements were made from 82 herbarium specimens and 18 field specimens selected to cover the geographic and morphological range of the *T. basedowii* complex (Table S2.2). Where possible, specimens used for morphological measurements were also targeted for sequencing, but sequences were not always recoverable from older specimens with degraded DNA.

We measured or assessed 28 quantitative and qualitative morphological characters, including ratios (Table 2.2). Characters were chosen based on whether they had been used previously to distinguish taxa in the complex (Burbidge 1946, e.g. 1953; Lazarides 1997) and on whether they appeared to differ between taxa based on

Table 2.2. Morphological characters measured for use in morphometric analyses. Grass morphological terminology: leaves are composed of a sheath clasping the stem (culm) and a blade extending away from the stem; the orifice is the area at the junction of the sheath and blade; the ligule is (in *Triodia*) a row of hairs on the upper surface of the orifice (against the stem); inflorescences are composed of spikelets, which are made of two bracts (glumes) and multiple florets, each composed of two bracts (lemma and palea) enclosing the ovary and stamens; the florets in a spikelet are connect by a rhachilla, which (in *Triodia*) breaks apart and a piece remains attached to each floret. (table continued on next page)

Character	Mean	SE	SD	Min–max
Quantitative				
Maximum leaf blade length (mm)	131.70	5.81	58.09	45–385
Ligule length (mm)	0.65	0.03	0.32	0.25–2
Maximum inflorescence length (mm)	89.47	3.04	30.38	30–195
Maximum pedicel length (mm)	8.54	0.32	3.24	1–19
Quantitative average				
Number of inflorescence branches	4.79	0.20	2.75	0–10
Number of spikelets per inflorescence	17.71	0.65	8.98	4–54
Spikelet length (mm)	11.28	0.15	2.65	5.6–26.6
Number of lower glume nerves	9.89	0.14	2.41	5–17
Lower glume length (mm)	6.82	0.10	1.72	3.0–15.75
Lower glume width (mm)	3.37	0.04	0.67	1.5–5.5
Number of florets per spikelet	7.69	0.09	1.61	4–13
Number of lemma midlobe nerves	3.97	0.07	1.15	1–7
Number of lemma lateral lobe nerves	5.50	0.07	1.24	2–10
Lemma body width (mm)	3.25	0.03	0.58	2–5.25
Lemma midlobe length (mm)	4.43	0.08	1.40	2–11
Lemma lateral lobe length (mm)	3.79	0.05	0.81	2–6.5
Palea length (mm)	3.73	0.04	0.66	2.3–6
Palea rhachilla length (mm)	0.87	0.01	0.23	0.3–1.75
Quantitative ratio (average)				
Spikelet exertion (length beyond highest glume apex : total length)	0.36	0.01	0.12	0–0.64
Lemma midlobe relative length (midlobe length : body length)	1.69	0.03	0.54	0.9–5

Character	Mean	SE	SD	Min–max
Palea shape (length : width)	3.31	0.04	0.68	1.9–5.5
Qualitative multistate				
Leaf sheath indumentum	0=glab, 1=hairy margins, 2=villous/tangled woolly exterior			
Orifice indumentum	0=glab/puberulent, 1=sparse hairs 2=villous/tangled woolly			
Blade indumentum	0=glab, 1=glab with rare hairs extend, 2=hairs extend			
Glume marginal hairs	0=glab, 1=hairs present			
Lemma midlobe indumentum	0=glab, 1=sparse hairs, 2=hairy beyond lower half			
Lemma lateral lobe indumentum	0=glab or sparsely hairy, 1=hairy beyond lower half			
Lemma lobe fusion	0=no fusion, 1=lobes fused			

preliminary assessment. Measurements were made by hand with a ruler (recorded to the nearest 0.25 mm), with the aid of a dissecting microscope when appropriate.

Spikelets were rehydrated in boiling water and detergent for *c*. 5 to 10 minutes prior to measurement. Quantitative averages were based on a variable number of observations per individual depending on available material. Measurements of floret characters were made on the lowest floret in each of three spikelets. Multiple measurements per individual were averaged for inclusion in morphometric analysis, and ratios were calculated prior to averaging.

Constrained and unconstrained approaches (see Anderson & Willis 2003) were used to visualize morphological patterns using PRIMER v. 6 (Clarke & Gorley 2006) with the add-on PERMANOVA+ (Anderson *et al.* 2008). Similarity matrices were constructed using Gower's similarity coefficient (Gower 1971), as the data contained both quantitative and multistate characters. Unconstrained analyses used non-metric multidimensional scaling (MDS; see Pimentel 1981; Clarke 1993) and hierarchical group average clustering (see Clarke 1993). MDS was conducted on the full sample set, using 1000 restarts to identify the global stress minimum. Clustering was conducted on the full sample set and on a subset lacking the more divergent samples from the first clustering analysis and MDS. Constrained analyses used canonical analysis of principal components (CAP; Anderson & Willis 2003) using the *a priori* identifications. The CAP analyses were conducted on groups identified in the clustering analysis to test whether entities in those groups could be separated. The number of principal coordinates was chosen automatically in PRIMER to minimize the "leave-one-out" classification error and its residual sum of squares. The trace statistic for the null hypothesis of no group differences was calculated with 999 permutations.

Taxon delimitation

We qualitatively compared our three data sets (nuclear DNA, chloroplast DNA and morphology) for evidence of lineage divergence. We evaluated our molecular data using the criterion of exclusivity (genealogical species concept; Baum & Shaw 1995). Entities that were monophyletic in ETS+ITS were considered divergent, as

were entities with unique chloroplast haplotypes. In cases where chloroplast haplotypes were shared across putative taxa but haplotypes within a taxon were similar, the evidence was interpreted as congruent with the recognition of that taxon. Relying on monophyly of a single molecular marker to delimit species has been criticized (e.g. Edwards 2009) because of the stochastic nature of genetic inheritance (Knowles & Carstens 2007) and expected gene tree incongruence with species trees because of such processes as incomplete lineage sorting and introgression (Maddison 1997), underscoring the importance of integrating additional evidence. Morphological evidence was evaluated using the criteria of diagnostic combinations of character states (phylogenetic species concepts; Cracraft 1983; Nixon & Wheeler 1990) and phenetic discontinuity (morphological species concept; e.g. Sokal & Sneath 1963). Where lines of evidence were incongruent, taxa could still be recognized if the incongruence could be explained with recourse to evolutionary theory (see Schlick-Steiner *et al.* 2010).

Nucleotide diversity and summary statistics

We quantified nucleotide diversity with the program DnaSP v. 5.10.01 (Librado & Rozas 2009) for each marker across taxa in the *T. basedowii* complex that we recognized based on integrated evidence. We also calculated Tajima's D (Tajima 1989), Fu's Fs (Fu 1997) and Ramos-Onsins and Rozas' R2 (Ramos-Onsins & Rozas 2002) for taxa with sample sizes of at least 10 sequences. While Fu's Fs is more powerful than Tajima's D and Ramos-Onsins and Rozas' R2 at detecting population expansion, it is sensitive to recombination and may be significant even under constant population size (Ramírez-Soriano *et al.* 2008). Both Tajima's D and Ramos-Onsins and Rozas' R2 are preferred when there is recombination, and R2 is better for smaller sample sizes (Ramos-Onsins & Rozas 2002; Ramírez-Soriano *et al.* 2008). Since recombination is likely in ETS and ITS, we did not use Fu's Fs to test for neutrality in those sequences. Significance for Fu's Fs and Ramos-Onsins and Rozas' R2 was calculated with 10,000 coalescent simulations under the constant population size model. To reconstruct haplotypes for the ITS and ETS sequences, we used the built-in PHASE (Stephens *et al.* 2001; Stephens & Scheet 2005) module in DnaSP

with default settings. As DnaSP ignores positions with missing data in the dataset, we trimmed and excluded some sequences to reduce missing data.

Results

Molecular phylogenies

The ITS alignment was 662 bp and provided 128 parsimony-informative characters (Table S2.3). Trees produced from the MP, ML and BI analyses are largely congruent, at least for strongly supported clades. The MP analysis resulted in 1497 most parsimonious trees, one of which is shown (Fig. 2.2) with indications of support from all three analyses. There is strong support (>89/>89/1.00) for two *T. concinna* groups, *T. plurinervata*, *T.* "ipluri", *T.* "shova" and *T.* sp. Peedamulla. For the most part, relationships between clades are poorly supported. ITS copies characteristic of two different clades were recovered from some individuals (e.g. the putative *T.* sp. Shovelanna Hill × *T.* sp. Warrawagine hybrids, marked in Fig. 2.2). Select ITS positions variable between putatively hybridizing entities in the *T. basedowii* complex (mapped in Table 2.3) show that putative hybrids have ITS copies from each proposed parent, with both fixed and heterozygous positions. The ITS primer combination did not appear to amplify highly divergent sequences indicative of pseudogenes.

Analysis of ETS alone (not shown) resolved the same clades as in the ITS tree but with greater support for relationships between clades. As ETS is part of the same rDNA repeats as ITS and showed a congruent signal, we concatenated the two regions. The concatenated ETS+ITS alignment was 1753 bp and provided 215 parsimony-informative characters (Table S2.3). Trees produced from the MP, ML and BI analyses are largely congruent, except that the placement of the eastern *T. concinna* specimen sister to the *T. basedowii* complex is not supported by BI (placed on a polytomy with the western *T. concinna* specimens and the *T. basedowii* complex instead). The MP analysis resulted in 5130 most parsimonious trees, one of which is shown (Fig. 2.3) with support values from all analyses. Within a strongly supported (100/100/1.00) *T. basedowii* complex there is strong support (>97/100/1.00) for clades corresponding to most of the informally named taxa (*T.* sp.

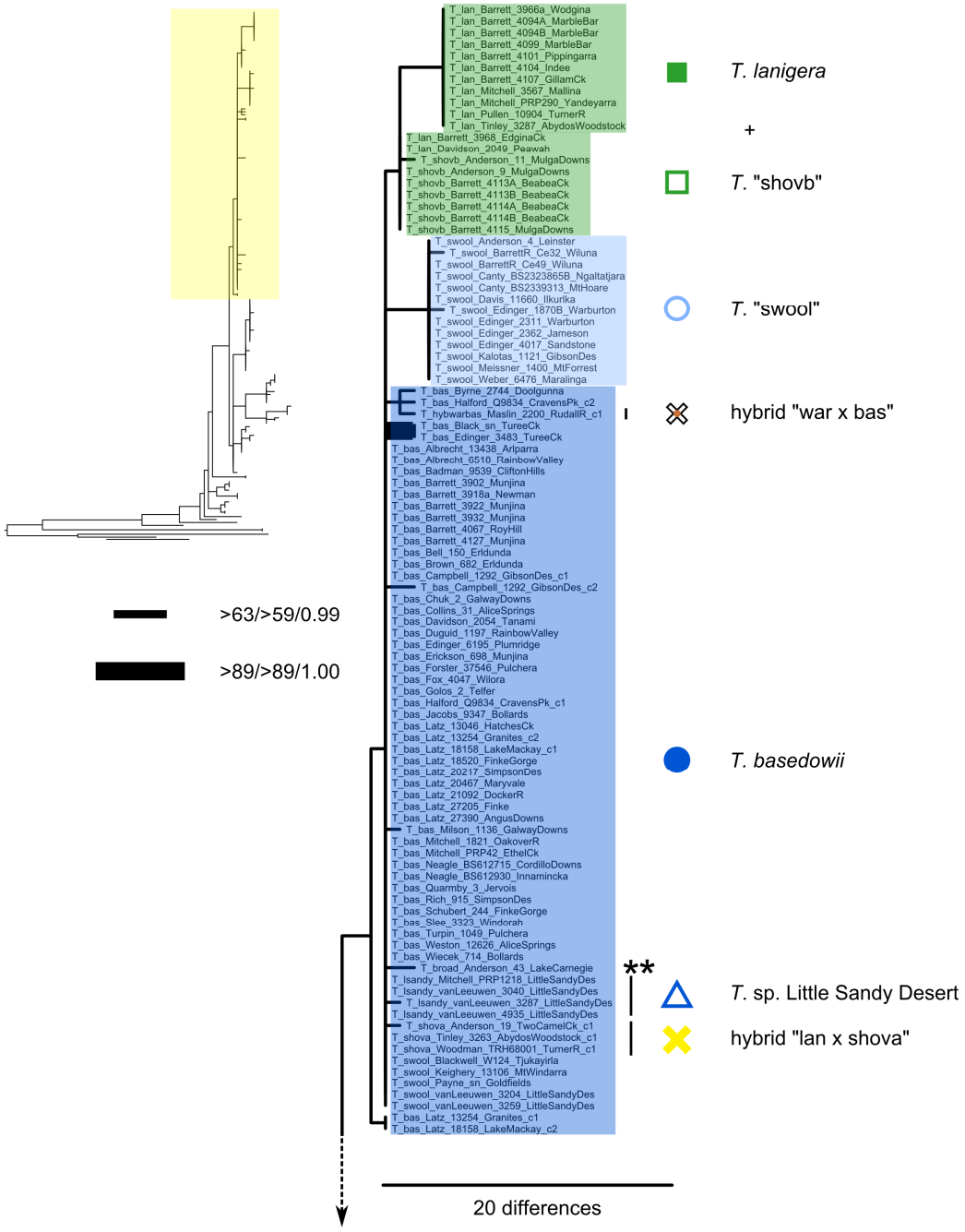

Figure 2.2. One of 1497 most parsimonious MP trees showing inferred relationships in the *T. basedowii* species complex based on ITS. (continued on next page)

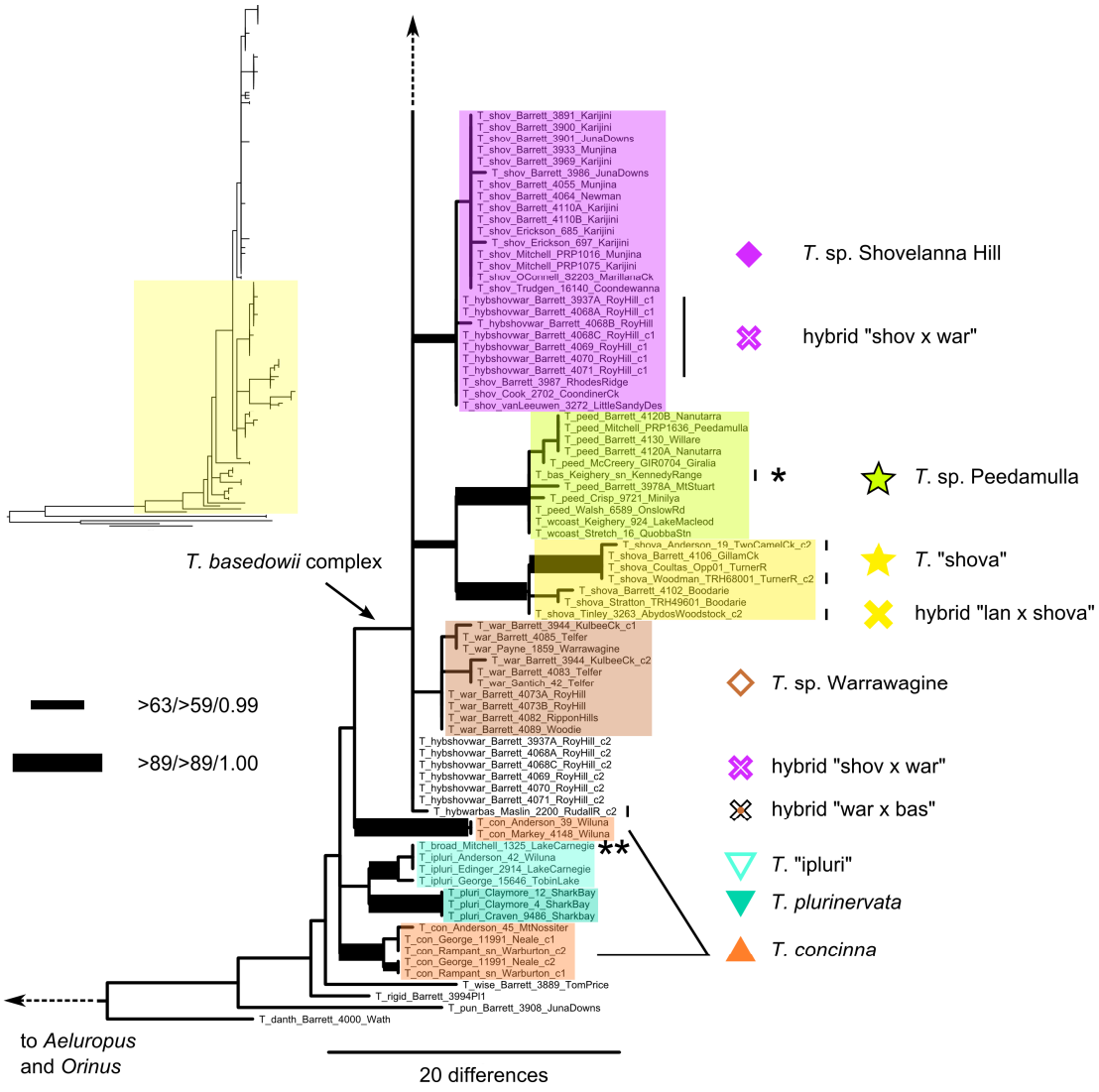

Figure 2.2. (cont.) Branches are drawn proportional to the number of character state differences. Clades within the ingroup having more than 89% bootstrap support (500 bootstrap replicates in PAUP* and RAxML, in that order) and posterior probabilities of 1 from MrBayes are drawn with thickened lines, while those with some but lower support (posterior probabilities of 0.99) are drawn with lines half as thick. ITS copies from putative hybrids are indicated on both locations in the tree. * An individual with a *T.* sp. Peedamulla ITS sequence but a *T. basedowii rps16-trnK* spacer sequence. ** Two individuals morphologically identified as *T.* "broad" with disparate placement in the ITS tree.

Table 2.3. ITS sequence variation in putatively hybridizing taxa in the *T. basedowii* species complex. The upper set of sequences show *T. lanigera*, *T.* "shova" and both major sequence types found in putative hybrids between them. The lower set of sequences show the same for *T.* sp. Shovelanna Hill and *T.* sp. Warrawagine. The solid border shows a region where length variation allows the reconstruction of two sequence types in hybrids using Sanger sequencing of both strands. Dashed borders around positions provide examples of putative copy homogenization and partial homogenization at positions of fixed differences, from left to right. Fixed differences are shown with bold letters. Ambiguity codes are: Y=C/T, W=A/T, R=G/A. P1 and P2 are putative parental sequences for each set of hybrids.

		ITS alignment position																			
	Sample	88	92	126	135	217	218	232	435	445	453	467	468	469	472	473	485	535	569	577	580
	T. basedowii (typical)	C	A	G	C	A	C	G	C	C	C	-	-	-	T	G	A	G	C	A	C
P1	*T. lanigera* (typical)	T	**A**	**G**	**T**	G	C	G	C	C	C	-	-	-	**T**	**G**	A	**G**	T	A	C
	T_lan_Davidson_2049_Peawah	Y	**A**	**G**	**T**	R	Y	G	C	C	C	-	-	-	**T**	**G**	A	**G**	Y	A	C
	T_shova_Anderson_19_TwoCamelCk_c1	C	W	R	Y	A	Y	G	C	C	C	-	-	-	T	G	A	G	C	A	C
	T_shova_Tinley_3263_AbydosWoodstock_c1	C	W	R	Y	A	C	G	Y	C	C	-	-	-	T	G	A	R	C	A	C
hybrids	T_shova_Woodman_TRH68001_TurnerR_c1	C	W	R	Y	R	Y	G	C	Y	C	-	-	-	T	G	A	G	Y	A	C
	T_shova_Anderson_19_TwoCamelCk_c2	C	W	R	Y	A	Y	G	T	T	T	-	A	C	G	C	A	A	C	A	C
	T_shova_Tinley_3263_AbydosWoodstock_c2	C	W	R	Y	A	C	G	Y	C	C	-	G	C	G	C	A	R	C	A	C
	T_shova_Woodman_TRH68001_TurnerR_c2	C	W	R	Y	R	Y	G	T	Y	Y	-	A	C	G	C	A	R	Y	A	C
	T_shova_Coultas_Opp01_TurnerR	C	**T**	**A**	**C**	A	Y	G	T	Y	Y	-	R	C	**G**	**C**	A	**A**	C	A	C
P2	T_shova_Barrett_4106_GillamCk	C	**T**	**A**	**C**	A	T	G	T	T	T	-	A	C	**G**	**C**	A	**A**	C	A	C
	T_shova_Stratton_TRH49601_Boodarie	C	**T**	**A**	**C**	A	C	G	T	C	C	-	G	C	**G**	**C**	A	**A**	C	A	C
P1	*T.* sp. Shovelanna Hill (typical)	C	T	G	C	A	**C**	**A**	C	C	C	T	G	-	-	G	**A**	G	C	**T**	C
	T_shov_Cook_2702_Newman	C	T	G	C	A	**C**	**A**	C	C	C	T	G	-	-	G	**A**	G	C	**T**	**Y**
	T_hybshovwar_Barrett_4068A_RoyHill_c1	C	T	G	C	A	Y	R	C	C	C	T	G	-	-	G	A	G	C	W	Y
hybrids	T_hybshovwar_Barrett_3937A_RoyHill_c1	C	T	G	C	A	Y	R	C	C	C	T	G	-	-	G	A	G	C	W	Y
	T_hybshovwar_Barrett_4068A_RoyHill_c2	C	T	G	C	A	Y	R	C	C	C	T	G	C	T	G	R	G	C	W	Y
	T_hybshovwar_Barrett_3937A_RoyHill_c2	C	T	G	C	A	Y	R	C	C	C	T	G	C	T	G	R	G	C	A	Y
P2	*T.* sp. Warrawagine (typical)	C	T	G	C	A	**T**	**G**	C	C	C	T	G	C	T	G	**G**	G	C	**A**	C

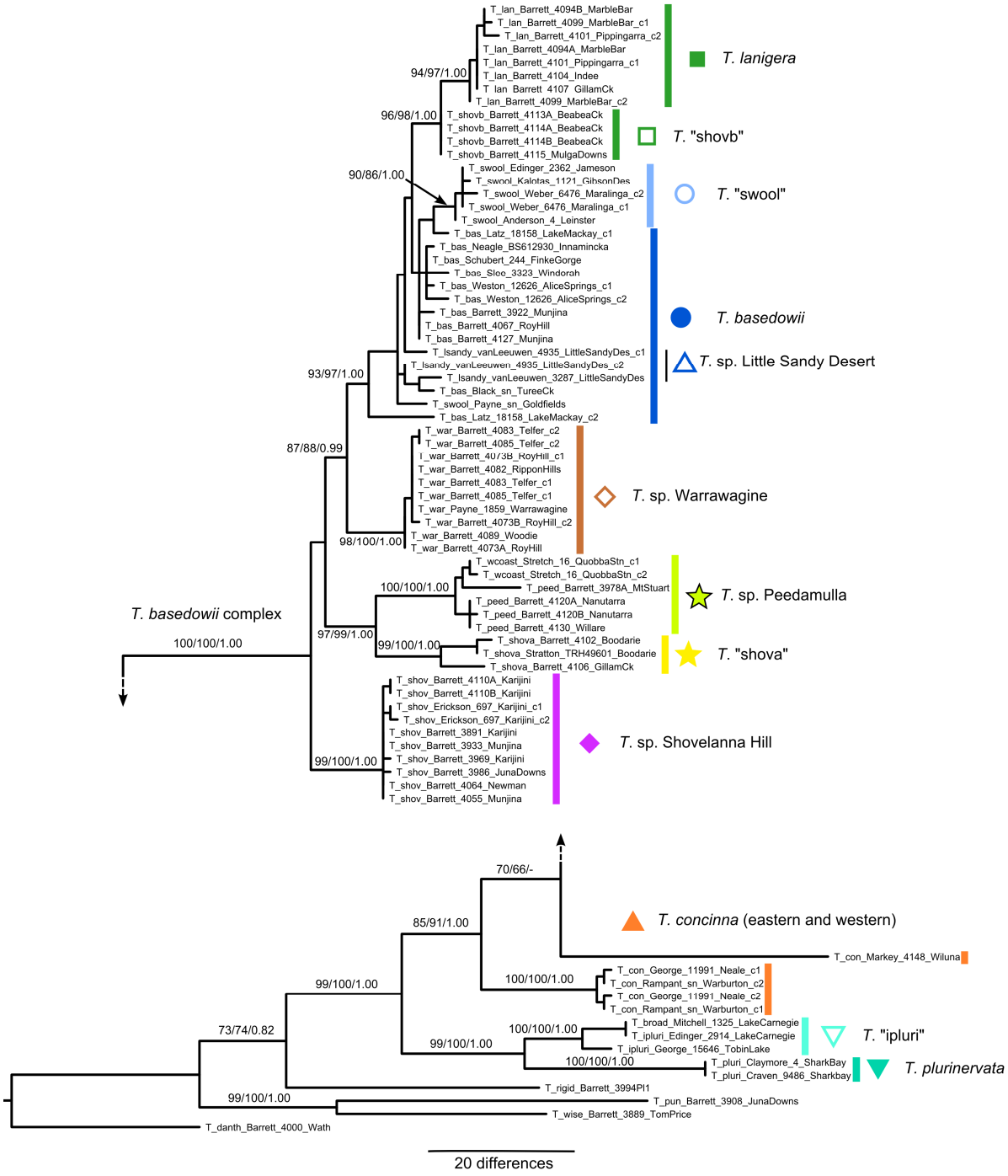

Figure 2.3. One of 5130 most parsimonious MP trees showing inferred relationships in the *T. basedowii* species complex based on concatenated ETS+ITS sequences. Branches are drawn proportional to the number of character state differences. Selected clade support values shown are based on 500 bootstrap replicates in PAUP*, 500 bootstrap replicates in RAxML, and posterior probabilities from MrBayes, in that order. No putative hybrids are included.

Shovelanna Hill, *T.* sp. Warrawagine, and *T.* sp. Peedamulla). In contrast, *T.* sp. Little Sandy Desert is part of the same clade as the widespread *T. basedowii*. Specimens identified as *T. lanigera* are monophyletic and sister to *T.* "shovb" on a polytomy with *T. basedowii* and *T.* "swool". The currently unrecognized lineage *Triodia* "shova" is strongly supported (97/99/1.00) as sister to *T.* sp. Peedamulla. There is also strong support (99/100/1.00) for the grouping of *T. plurinervata* and *T. concinna* with the *T. basedowii* complex. Specimens identified as *T. plurinervata* form a clade of coastal specimens and inland specimens (designated *T.* "ipluri"). *T. concinna* is broken into eastern and western groups (the eastern "group" consisting of a single sample), but there is no support for their monophyly.

The *rps16-trnK* spacer alignment was 793 bp (trimmed to 788 bp to remove the cytosine microsatellite) and provided 36 parsimony-informative characters (Table S2.3). The haplotype network (Fig. 2.4) reveals two main haplotype groups in the *T. basedowii* complex (designated A and B) with some intermediate haplotypes (designated "M" and "Peed"). Outgroups are ambiguously connected to the network but are not the focus of this study. There is both congruence and incongruence between the haplotype network and the ETS+ITS tree. Of the taxa forming clades in the ETS+ITS tree, some (e.g. *T.* sp. Peedamulla) are represented by unique chloroplast haplotypes, others (e.g. *T. lanigera*) are consistently of one or two closely related but non-unique chloroplast haplotypes, and one (*T.* sp. Shovelanna Hill) has multiple divergent chloroplast haplotypes. In addition, there are individuals of *T.* sp. Warrawagine, *T. basedowii* and *T.* "swool" which have divergent haplotypes compared to most individuals sequenced for those taxa. Sister taxa of the *T. basedowii* complex in the ETS+ITS tree (*T. plurinervata*, *T.* "ipluri" and *T. concinna*) each have unique chloroplast haplotypes. Haplotype diversity in the *T. basedowii* complex is geographically structured (Fig. 2.5), with more haplotypes found in the Pilbara (14) compared to the Australian deserts (7).

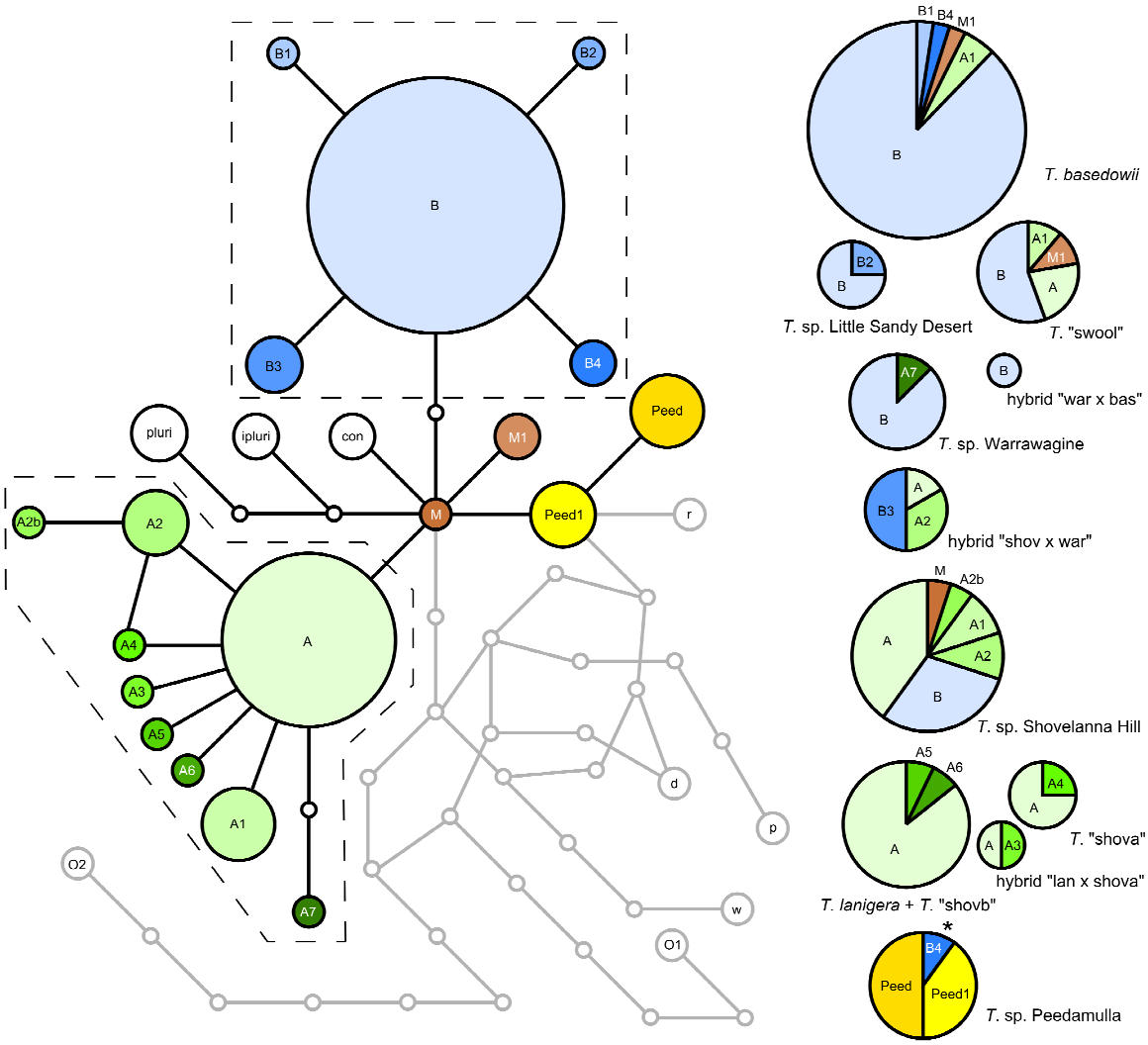

Figure 2.4. Chloroplast haplotype distribution in the *T. basedowii* species complex based on sequences of the *rps16-trnK* spacer. **Left:** Haplotype network, showing haplotype diversity and relationships; each line between circles represents one difference between haplotypes, and empty circles represent missing or unsampled haplotypes. The haplotypes "pluri," "ipluri" and "con" were found in *T. plurinervata*, *T.* "ipluri" and *T. concinna*, respectively. Outgroups (shown in grey) are the two *Orinus* spp. (O1 and O2), *Triodia rigidissima* (r), *T. wiseana* (w), *T. danthonioides* (d) and *T. pungens* (p). The two main haplotype groups are designated with dashed lines for reference. Haplotypes "M" and "M1" are intermediate between the two main groups. **Right:** Correspondence between chloroplast haplotypes in the *T. basedowii* complex and putative taxa. Circles are drawn with area proportional to number of individuals sampled. * An individual with a *T.* sp. Peedamulla ITS sequence but a *T. basedowii* chloroplast haplotype (see Fig. 2.2).

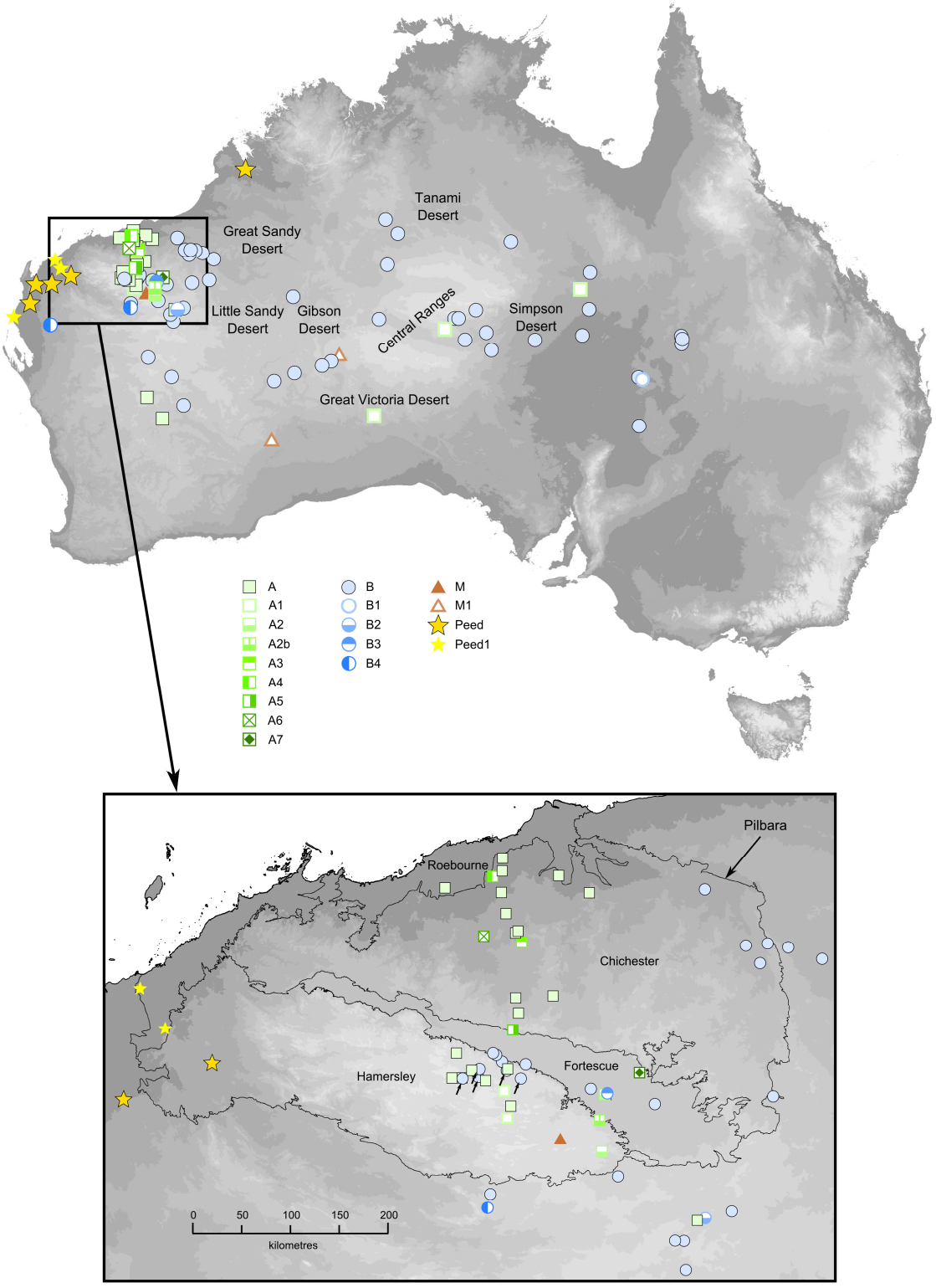

Figure 2.5. Geographic distribution of chloroplast *rps16-trnK* spacer haplotypes in the *T. basedowii* species complex. Colours and haplotypes are as in Fig. 2.4, and geographic labels are as in Fig. 2.1. *T. basedowii* haplotypes occurring in the Hamersley subregion are marked with arrows.

Morphology

Hierarchical clustering (Fig. 2.6A) and unconstrained MDS analysis (Fig. 2.6B; 2D stress of 0.15) of the full morphological dataset showed that *T. concinna*, *T. plurinervata, T.* "ipluri", and three samples of *T. lanigera* were morphologically divergent from the rest of the samples. A CAP analysis of *T. concinna*, *T. plurinervata* and *T.* "ipluri" (Fig. 2.6C) samples showed clear differentiation between those taxa. With these divergent samples removed, a second clustering analysis (Fig. 2.7A) revealed two main clusters, one comprising all samples of *T. basedowii* and *T.* "swool" plus two samples of *T. lanigera*, the second comprising all remaining samples. The *T. lanigera* specimens in the first cluster are clearly distinguishable in the CAP analysis (Fig. 2.7D), but the *T. basedowii* and *T.* "swool" samples are poorly differentiated; only a portion of the *T.* "swool" samples have a *T.* "swool" ITS sequence (circled in Fig. 2.7A). CAP analysis of important groupings in the second cluster (Fig. 2.7B,C) indicate that *a priori* entities can be differentiated within them. Some specimens are incongruously placed: *T.* "shovb" samples do not consistently group together and neither do *T.* "wcoast" samples (asterisks in Figs. 2.6 & 2.7), and the Keighery s.n. sample (highlighted in Figs. 2.6 & 2.7) groups with *T. basedowii* in the first clustering but outside it in the second.

Taxon delimitation

Congruent molecular and morphological evidence support recognition of the following taxa: *T. plurinervata*, *T. concinna*, *T. basedowii*, *T. lanigera*, *T.* sp. Peedamulla, and *T.* "shova" (see Table 2.4). In addition, there is support for the recognition of *T.* sp. Shovelanna Hill, *T.* sp. Warrawagine, *T.* sp. Little Sandy Desert and *T.* "ipluri", but incongruence (*T.* sp. Little Sandy Desert), limited sampling (*T.* "ipluri"), and hybridization/morphological similarities (*T.* sp. Shovelanna Hill and *T.* sp. Warrawagine) mean these taxa require further investigation. We tentatively recognize them for the purposes of this study. In contrast, there is insufficient evidence to support recognition of the following taxa: *T.* "wcoast" (in *T.* sp. Peedamulla), *T.* "swool" (in *T. basedowii*), and *T.* "shovb" (in *T. lanigera*). Possibly

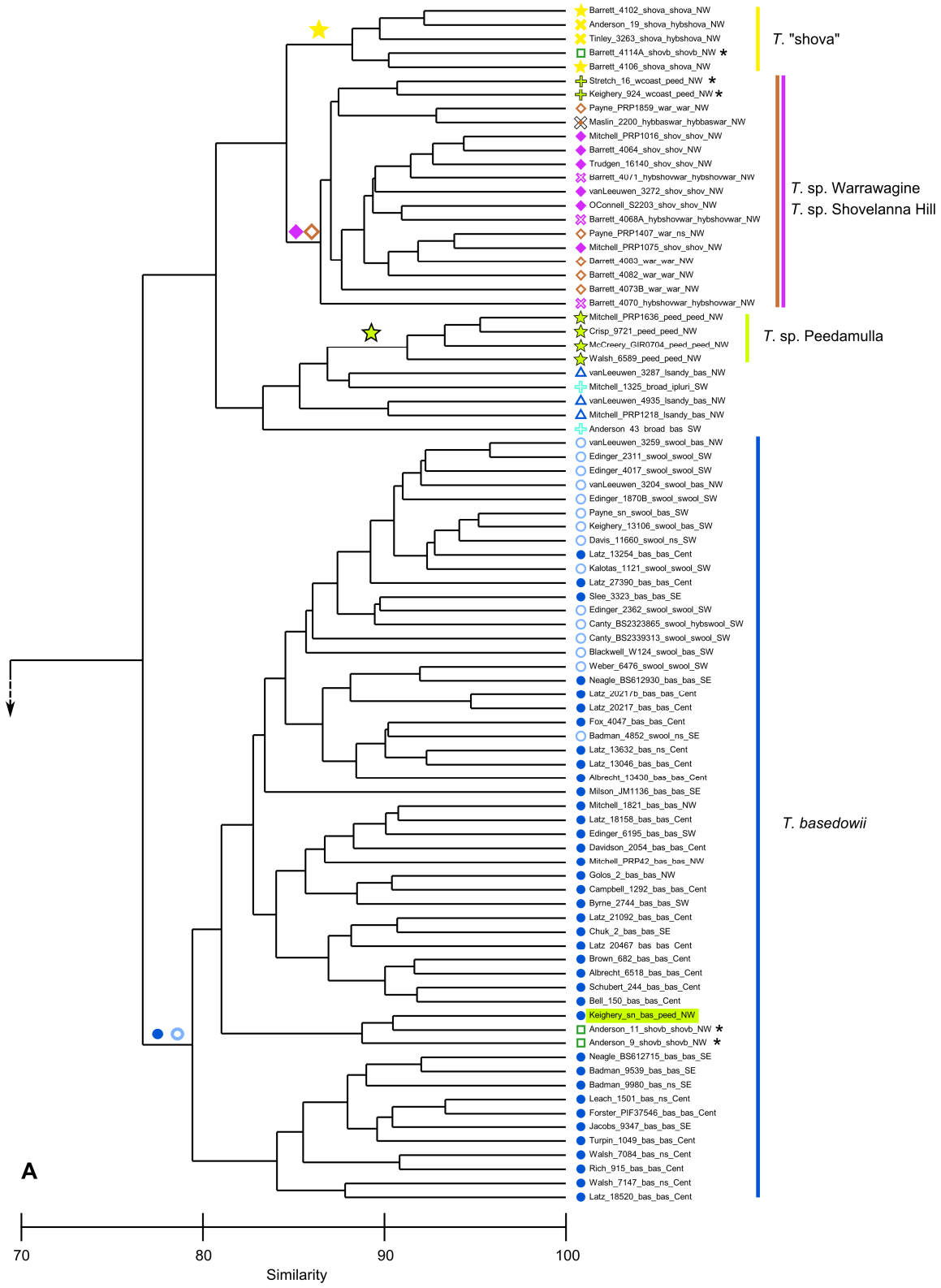

Figure 2.6. Morphometric analyses of specimens from the *T. basedowii* species complex.

(continued on next page)

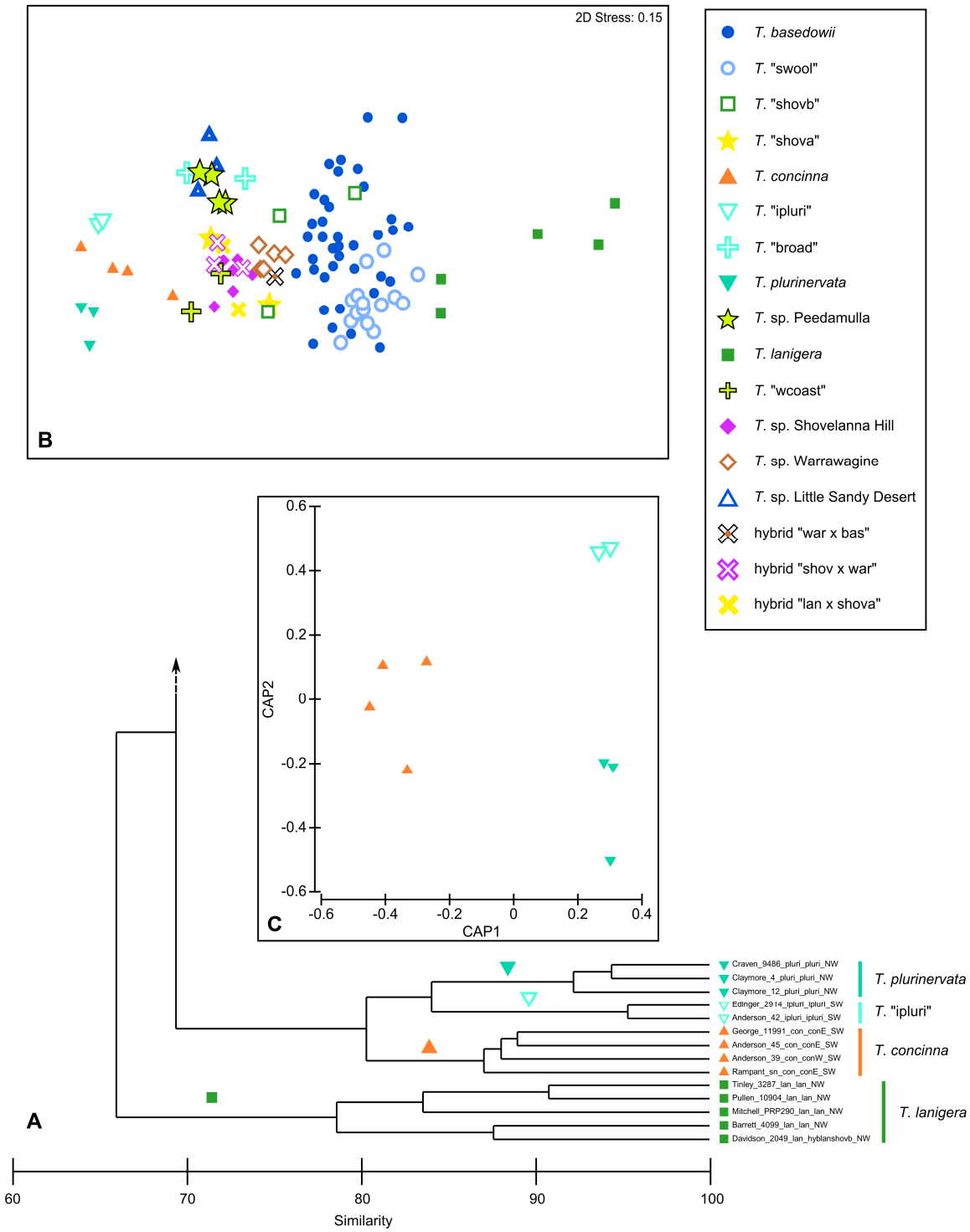

Figure 2.6. (cont.) Symbols are based on putative taxa. (**A**) group average clustering analysis of all specimens; (**B**) 2D MDS plot for all specimens measured; (**C**) 2D ordination of CAP analysis for differences between *T. concinna*, *T. plurinervata* and *T.* "ipluri".

* Incongruently placed specimens. An individual with a *T.* sp. Peedamulla ITS sequence and a *T. basedowii* chloroplast haplotype is highlighted.

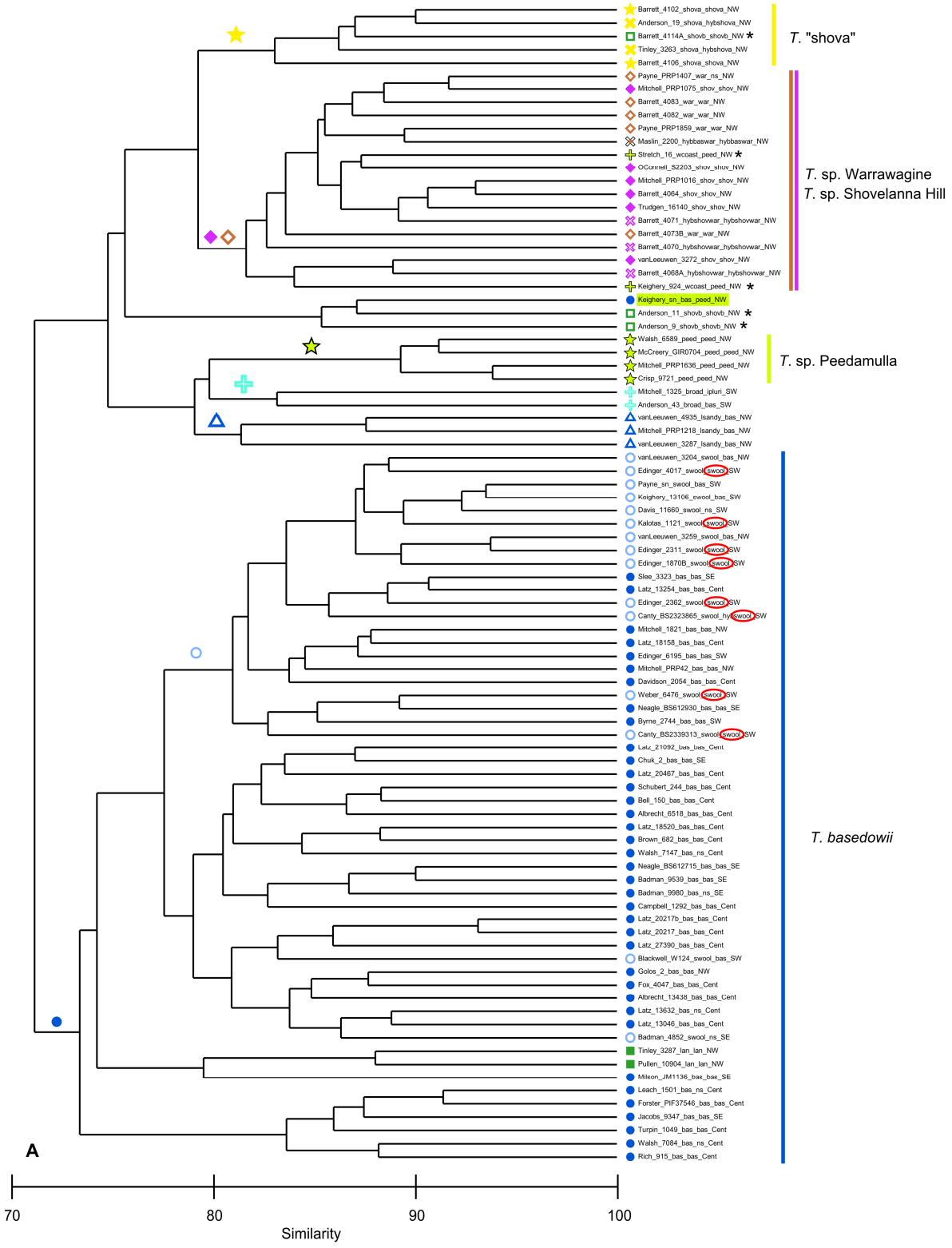

Figure 2.7. Morphometric analyses of specimens from the *T. basedowii* species complex (continued). (continued on next page)

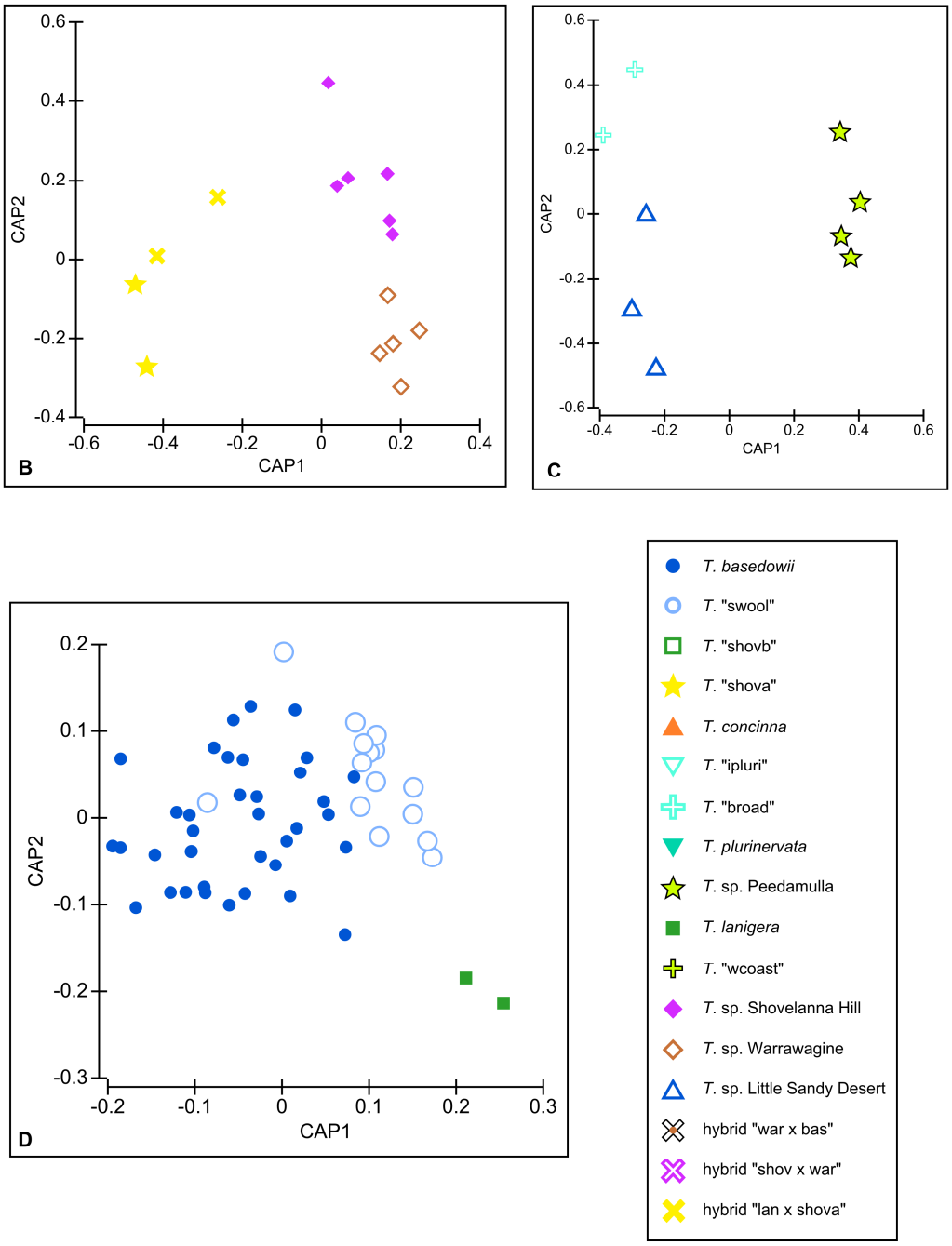

Figure 2.7. (cont.) (**A**) group average clustering analysis of specimens after removing *T. concinna*, *T. plurinervata*, *T.* "ipluri" and three divergent *T. lanigera* specimens; (**B**) 2D ordination of CAP analysis for differences between *T.* sp. Shovelanna Hill, *T.* sp. Warrawagine and *T.* "shova", excluding hybrids and anomalous specimens; (**C**) 2D ordination of CAP analysis for differences between *T.* sp. Peedamulla, *T.* sp. Little Sandy Desert and *T.* "broad"; (**D**) 2D ordination of CAP analysis for differences between *T. basedowii*, *T.* "swool" and *T. lanigera* (two less divergent specimens). * Incongruently placed specimens. Specimens with *T.* "swool" ITS sequences are circled. An individual with a *T.* sp. Peedamulla ITS sequence and a *T. basedowii* chloroplast haplotype is highlighted.

Table 2.4. Summary of three sources of evidence (nuclear DNA, chloroplast DNA and morphology) for taxon distinction in the *T. basedowii* species complex. Red lettering indicates conflict or inadequate evidence. (table continued on next two pages)

Putative taxon	Nuclear evidence (ITS+ETS)	Chloroplast evidence (*rps16-trnK* spacer)	Morphological evidence (distinctive characters)	Morphological evidence (phenetic clustering)	Summary	Conclusion
T. plurinervata	monophyletic (few samples)	unique (few samples)	fused lemma lobes plus a distinctive inflorescence (spicate); shared with the larger *T.* "ipluri"	distinct (few samples)	congruent evidence suggests this is a distinct taxon	support current recognition
T. "ipluri"	monophyletic (few samples)	unique (few samples)	fused lemma lobes plus a distinctive inflorescence (spicate); shared with the smaller *T. plurinervata*	distinct (few samples)	morphologically similar to *T. plurinervata* but larger and geographically disjunct; divergence in all three lines of evidence suggests distinction	new taxon pending further sequencing
T. concinna	paraphyletic eastern and western populations (low support and sampling)	unique (few samples; shared by both populations)	fused lemma lobes, branched inflorescence and visible palea wings	distinct (few samples; no difference between populations)	no distinctive characters between eastern and western populations, and low support for monophyly; additional sampling is needed to support splitting	support current recognition
T. basedowii	a broad polytomy (low support)	predominantly one haplotype (a few exceptions); shared with other taxa	no consistent diagnostic characters, except perhaps leaf blade hairiness, which it shares with *T.* "swool", *T. lanigera* and *T.* "shovb", differing from the latter two in relative lemma midlobe length	largely distinct and variable, but clusters with *T.* "swool"	widespread species with population variation; chloroplast exceptions may be explained by hybridization or incomplete lineage sorting; morphological intergrading with *T.* "swool" suggest the two may be one taxon	support current recognition
T. lanigera	monophyletic (some specimens are heterozygous at fixed differences with *T.* "shovb")	consistent similar haplotypes; shared with other taxa	woolly leaf sheaths, branched inflorescences and a long lemma midlobe relative to body	largely distinct; two specimens closer to *T. basedowii* but differentiable	evidence suggests distinction; potential hybridization with *T.* "shova" and uncertain relationship with *T.* "shovb" specimens	support current recognition

Putative taxon	Nuclear evidence (ITS+ETS)	Chloroplast evidence (*rps16-trnK* spacer)	Morphological evidence (distinctive characters)	Morphological evidence (phenetic clustering)	Summary	Conclusion
T. sp. Shovelanna Hill	monophyletic; some putative hybrids with *T.* sp. Warrawagine	multiple divergent haplotypes	smaller than *T. basedowii* with less hairy leaf blades and a relatively unbranched (vs. branched) inflorescence; relatively wider palea and less branched inflorescence than *T.* sp. Warrawagine	distinct (but close to *T.* sp. Warrawagine)	chloroplast incongruence possibly explained by introgression or incomplete lineage sorting; additional sampling/evidence needed to determine distinction from *T.* sp. Warrawagine	new taxon pending further sequencing
T. sp. Warrawagine	monophyletic; some putative hybrids with *T.* sp. Shovelanna Hill	mostly one consistent haplotype (one anomaly); shared with other taxa	typically smaller and less hairy leaf blades than those of *T. basedowii*; a more branched inflorescence and relatively narrower palea than *T.* sp. Shovelanna Hill	distinct (but close to *T.* sp. Shovelanna Hill)	chloroplast anomaly possibly explained by introgression or incomplete lineage sorting; additional sampling/evidence needed to determine distinction from *T.* sp. Shovelanna Hill	new taxon pending further sequencing
T. sp. Peedamulla	monophyletic	unique (one anomaly)	glabrous leaves, branched inflorescence and a long lemma midlobe relative to body	distinct (few samples)	evidence suggests distinction; chloroplast anomaly probably indicative of hybridization (specimen morphologically similar to *T. basedowii* or *T.* "shovb")	new taxon
T. sp. Little Sandy Desert	part of the *T. basedowii* polytomy	consistent similar haplotypes shared with *T. basedowii*; one somewhat distinct	differs from *T. basedowii* in its glabrous leaves and distinctive spikelet appearance	distinct (few samples)	morphologically distinct but with little molecular difference from *T. basedowii*; the incongruence could be explained by hybridization	new taxon pending further sequencing
T. "shova"	monophyletic; some putative hybrids with *T. lanigera* or *T.* "shovb"	consistent similar haplotypes; shared with other taxa	a relatively unbranched inflorescence, no hairs on leaf blades or sheaths, and long lemma midlobe relative to body	distinct (few samples, and one anomalous *T.* "shovb" specimen close)	evidence suggests distinction, with possible hybridization; shared chloroplast haplotype possibly explained by introgression or incomplete lineage sorting	new taxon

Putative taxon	Nuclear evidence (ITS+ETS)	Chloroplast evidence (*rps16-trnK* spacer)	Morphological evidence (distinctive characters)	Morphological evidence (phenetic clustering)	Summary	Conclusion
T. "wcoast"	part of the *T.* sp. Peedamulla clade	(not sequenced)	smaller spikelet compared to *T.* sp. Peedamulla	indistinct (few samples), clustering with *T.* sp. Shovelanna Hill and *T.* sp. Warrawagine	not clear whether this entity is a coastal form of *T.* sp. Peedamulla, of hybrid origin or a distinct taxon but with nuclear introgression	do not recognize
T. "broad"	incongruently placed: one sequence is part of *T.* "ipluri" and the other *T. basedowii*	(not sequenced)	distinct from *T.* "ipluri" in its branched inflorescence and longer pedicels; distinct from *T. basedowii* in its broad, indurated lateral lemma lobes	distinct (few samples)	the morphological distinctiveness is incongruent with nuclear evidence	do not recognize
T. "swool"	monophyletic but branching off of the *T. basedowii* polytomy	multiple divergent haplotypes	typically woolly leaf sheaths, but this character is found in the widespread *T. basedowii*	clusters somewhat peripherally to *T. basedowii*, but with specimens having *T. basedowii* sequences	potentially population differentiation in *T. basedowii*; morphological and nuclear incongruence may be explained by hybridization and introgression	do not recognize
T. "shovb"	a polytomy sister to *T. lanigera* (some specimens of which are heterozygous at fixed differences between the two taxa)	consistent similar haplotypes; shared with other taxa	longer lemma midlobe relative to body than *T. basedowii*, and differs from *T. lanigera* in its glabrous lemma lobes and relatively broader palea	no distinct cluster; incongruously placed	morphologically inconsistent, and possibly hybridizing with *T.* "shova"; possibly differentiation within *T. lanigera* or a hybrid	do not recognize

of hybrid origin, *T.* "broad" is also not recognized, due to restricted sampling and incongruence between molecular and morphological data.

Nucleotide diversity and summary statistics

Higher levels of nucleotide diversity (pi) are expected for taxa with greater genetic diversity, and significant deviations from neutrality (e.g. negative values of Tajima's D) can be used to infer processes such as population expansion or selection. Across the *T. basedowii* complex, within-taxon nucleotide diversity varied (Table 2.5) but did not show a consistent pattern of relatively higher diversity within Pilbara taxa. Two Pilbara taxa (*T.* sp. Warrawagine and *T.* "shova") had the highest levels of ITS diversity, but other Pilbara taxa (e.g. *T. lanigera* and *T.* sp. Shovelanna Hill) had lower ITS diversity than the desert taxon *T. basedowii*. The taxa *T. basedowii* and *T.* sp. Shovelanna Hill had significant values for Tajima's D (negative) and Ramos-Onsins and Rozas' R2 (small) for both ITS and ETS. *T. lanigera* similarly had significant values (except for Tajima's D) for ITS. For the *rps16-trnK* spacer, Tajima's D, Fu's Fs and Ramos-Onsins and Rozas' R2 were not significantly different from neutral expectations for any tested taxa.

Discussion

As the first morphological and molecular analysis of the *T. basedowii* species complex, our study provides a basis for new taxonomic hypotheses for the group. This resolution of taxa, and elucidation of their geographic distributions, in turn allows us to evaluate both congruence of genetic patterns in the Pilbara biota and the influence of geologically recent environmental changes on plant diversity in arid Australia.

Evidence for multiple new taxa in the Triodia basedowii *species complex*

Species delimitation

Species delimitation is a challenging endeavour in species complexes, not only because of complicated evolutionary relationships or recent divergence, but also because what a species is remains subjective. We adopt de Queiroz's (1998, 2007)

Table 2.5. Nucleotide diversity and summary statistics for the three molecular markers used. Statistical significance is indicated by asterisks (* P < 0.05, ** P < 0.01). Sample size for the multi-copy markers ITS and ETS also indicates the number of individuals in parentheses, which is not always half the sample size because some individuals had multiple recoverable copies. Tests of neutrality were only conducted when sample size was at least 10.

	ITS					ETS					*rps16-trnK* spacer					
Taxon	**Sample size**	**Nuc. diversity Pi**	**SD**	**Tajima's D**	**R2**	**Sample size**	**Nuc. diversity Pi**	**SD**	**Tajima's D**	**R2**	**Sample size**	**Nuc. diversity Pi**	**SD**	**Tajima's D**	**R2**	**Fu's Fs**
T. basedowii	96 (46)	0.00594	0.00054	**-2.48599****	**0.0208****	34 (14)	0.01026	0.00092	**-1.83747***	**0.0525****	47	0.00151	0.00043	-0.75587	0.0802	-0.284
T. lanigera	38 (19)	0.00465	0.00052	-1.47811	**0.0618***	18 (8)	0.00647	0.00121	**-2.08322***	**0.0647****	12	0.00026	0.00021	-1.14053	0.2764	-0.476
T. sp. Little Sandy Desert	6 (3)	0.00164	0.00045			4 (1)	0.01669	0.00371			3	0.00104	0.00049			
T. sp. Shovelanna Hill	36 (18)	0.00297	0.00068	**-2.32093****	**0.0437****	20 (9)	0.00337	0.00079	**-2.09164***	**0.0663****	19	0.00298	0.00041	0.3578	0.1546	-0.494
T. sp. Warrawagine	20 (9)	0.00847	0.00125	-1.09064	0.0978	16 (6)	0.01362	0.00238	-1.05685	0.0984	6	0.00209	0.00135			
T. sp. Peedamulla	20 (10)	0.00397	0.00047	-0.94584	0.0867	10 (5)	0.00565	0.00122	-0.9701	0.1216	8	0.00084	0.00019			
T. "shova"	8 (4)	0.00855	0.00131			6 (3)	0.01053	0.00217			4	0.00078	0.00041			

general lineage concept, under which species are considered to be separately evolving segments of metapopulation lineages, and criteria considered necessary under other species concepts (e.g. monophyly, reproductive isolation) are used as lines of evidence for determining lineage divergence. While combining lines of evidence is the focus of integrative taxonomy (e.g. Schlick-Steiner *et al.* 2010), most current practice relies on congruence between sources of evidence rather than quantitative integration in one analysis (see Yeates *et al.* 2011). We interpret congruence between molecular and morphological results as support for lineage divergence.

Based on the criteria we used to evaluate lineage divergence (monophyly/distinct haplotypes and diagnostic/discontinuous morphology), there is strong evidence for two new taxa (*T.* sp. Peedamulla and *T.* "shova") and moderate evidence for three more (*T.* sp. Shovelanna Hill, *T.* sp. Warrawagine, and *T.* sp. Little Sandy Desert) in the *T. basedowii* species complex (Table 2.4). In some cases, lines of evidence are incongruent. For example, *T.* sp. Little Sandy Desert is strongly differentiated morphologically from *T. basedowii* but not differentiated in nuclear and chloroplast markers. In this case, we hypothesize that these are distinct species because they are ecologically differentiated (where they co-occur, *T.* sp. Little Sandy Desert grows on dune slopes while *T. basedowii* grows in dune swales) and we can posit evolutionary explanations for the shared nuclear sequence such as hybridization and introgression or recent divergence without accumulation of rDNA differences. Under these hypotheses, examining additional molecular markers should reveal differentiation between these two species. Based on our results, we recognize the following seven taxa in the *T. basedowii* species complex: *T. basedowii* (including *T.* "swool"), *T. lanigera* (including *T.* "shovb"), *T.* sp. Little Sandy Desert, *T.* sp. Shovelanna Hill, *T.* sp. Warrawagine, *T.* sp. Peedamulla (including *T.* "wcoast"), and *T.* "shova" (see Table 2.4).

Hybridization

Observations of co-occurrence of some species (e.g. *T.* sp. Shovelanna Hill and *T. basedowii*) without obvious morphological intermediates or shared nuclear markers

supports their distinction, but our molecular results also provide evidence of hybridization in some cases, complicating delimitation. This evidence includes incongruence between nuclear and chloroplast markers suggestive of chloroplast capture (Rieseberg & Soltis 1991) and the presence of ITS copies from two distinct lineages within single individuals. Some ITS copies (Table 2.3) retain nucleotides diagnostic for their respective source parent taxon while being heterozygous at other positions, perhaps indicating incomplete concerted evolution following hybridization (e.g. Nieto Feliner *et al.* 2004).

Examples of putative hybrids include a population occurring over an isolated area of low, outcropping ironstone formation in the Fortescue River valley between *T.* sp. Shovelanna Hill and *T.* sp. Warrawagine. The hybrid individuals possess ITS copies from both, and *rps16-trnK* spacer haplotypes similar to one or the other of the parent taxa (though chloroplast haplotypes are variable in *T.* sp. Shovelanna Hill). The presence of both chloroplast haplotype groups in this population suggests that both parents dispersed by seed to the location, rather than one introgressing by pollen flow from its adjacent populations. Possible hybridization and introgression is also found in the central Chichester, where a number of individuals of *T.* "shova" possess ITS copies characteristic of *T. lanigera* (Table 2.3) and could represent hybrids between them. Our chloroplast data are unhelpful in exploring hybridity in this case because the taxa share chloroplast haplotypes. The hybrids cluster morphologically with *T.* "shova" samples (though note one anomalous *T.* "shovb" = *T. lanigera* sample also clustering with *T.* "shova" in Figs. 2.6 & 2.7). Another example of putative hybridization is an individual at the southern extent of the range of *T.* sp. Peedamulla that has the same chloroplast haplotype as is found in the closest *T. basedowii* population, yet retains a *T.* sp. Peedamulla ITS sequence (asterisk in Figs. 2.2 & 2.4). Morphologically, the individual (highlighted in Figs. 2.6 & 2.7) does not group with *T.* sp. Peedamulla but is more closely associated with *T. basedowii* in the first clustering and *T.* sp. Peedamulla in the second.

In contrast to the above cases of possible hybridization, a mixed population of *T.* sp. Shovelanna Hill and *T. basedowii* without ITS admixture occurs at the edge of the Hamersley and Fortescue subregions. A distinct difference in floret maturity between

the two taxa in the population observed during sampling could indicate a flowering time difference contributing to isolation. The presence of chloroplast haplotypes typical of *T. basedowii* extending into the Hamersley subregion in populations of *T.* sp. Shovelanna Hill (arrows in Fig. 2.5), however, could indicate past chloroplast capture despite evidence the species are not currently mixing.

While the above examples of hybridization complicate delimitation, they also provide insight into evolutionary processes shaping the complex and the lack of complete reproductive isolation characteristic of closely-related species. In general, hybrids are rare and are not always evident where taxa co-occur. Importantly, hybridization on its own doesn't invalidate species distinction, as barriers to gene flow may be present between two species despite introgression in contact zones (e.g. Rieseberg *et al.* 1995). Consequently, we do not consider that the presence of hybrids invalidates taxon recognition. Their rarity and limited distribution suggest they are not established hybrid species, and we do not recognize them as taxa.

Geographic distributions of taxa and congruent patterns across the biota

Taxa in the *Triodia basedowii* complex exhibit strong geographic structure (Fig. 2.8), both across Australia and within the Pilbara. *T. basedowii* is widespread across the Australian sandy deserts, while the remaining taxa are restricted to Western Australia, predominantly the Pilbara region. Within the Pilbara, distributions of taxa are coincident with the boundaries of biogeographic subregions and track different underlying geologies. In particular, there is a north-south replacement series across the Fortescue River valley (Fig. 2.8), with *T.* "shova" and *T. lanigera* in the central Chichester on old granitic plains and basaltic ranges; *T. basedowii* in the Fortescue River valley on patchy, young, sandy soils; and *T.* sp. Shovelanna Hill in the topographically rugged Hamersley on skeletal soils over banded ironstone formations. Additionally, the two taxa in the central Chichester are replaced by *T.* sp. Warrawagine in the eastern Chichester and along the northern edge of the Fortescue River valley, potentially tracking the extension of ironstone formations from the south. Similar north-south differentiation across the Fortescue River valley has been

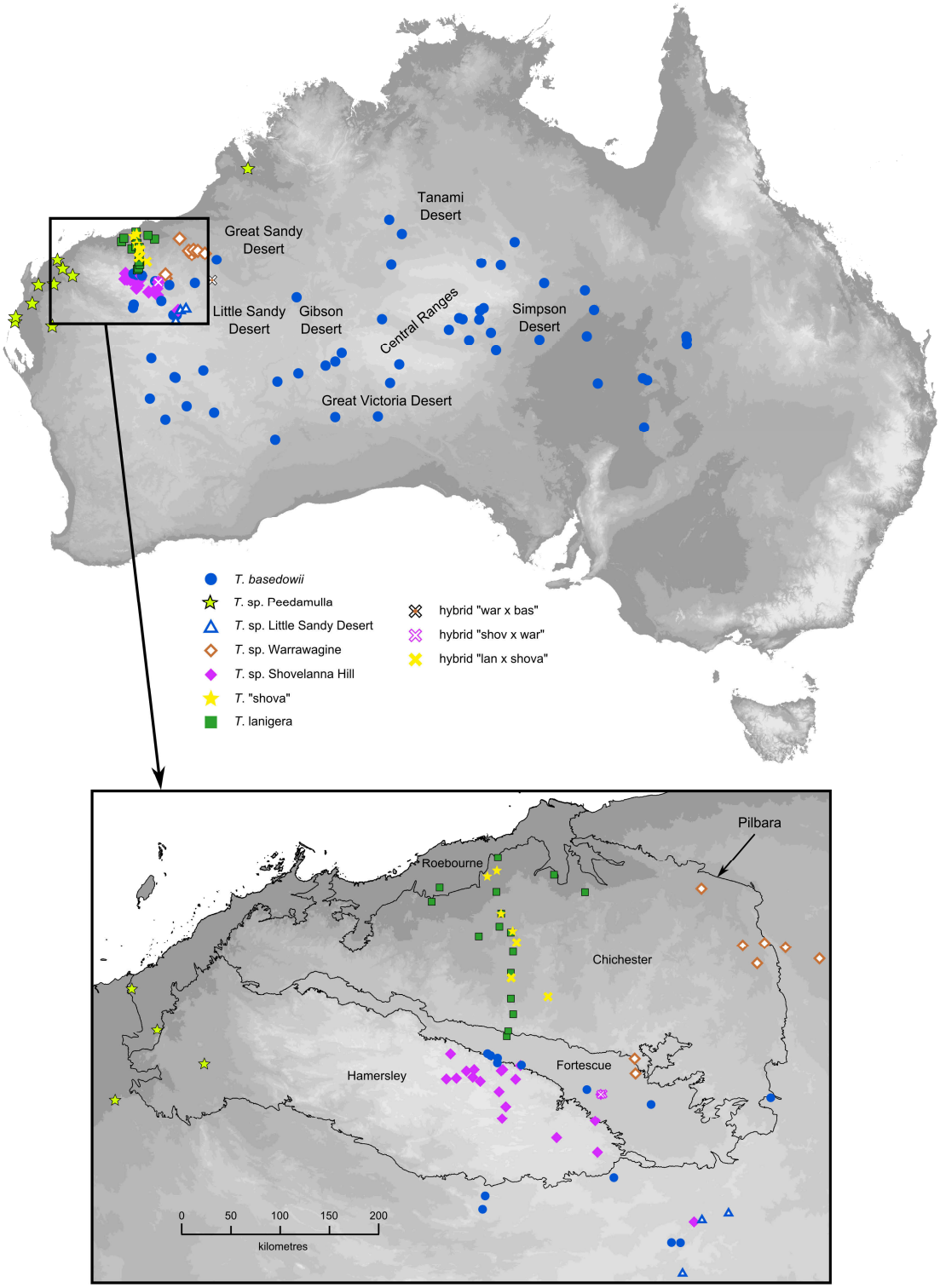

Figure 2.8. Geographic distributions of taxa from the *T. basedowii* species complex recognized in this study. Putative hybrids are included. The Pilbara region is indicated in the enlarged area with the IBRA subregions labelled, and the central deserts and central ranges are also labelled.

observed for lizards (Shoo *et al.* 2008; Doughty *et al.* 2012; Pepper *et al.* 2013a) and beetle (Guthrie *et al.* 2010) and spider (Durrant *et al.* 2010) species assemblages, while some east-west differentiation in the Chichester subregion was found in a species delimitation study on geckos (Pepper *et al.* 2013a). These geographic patterns of differentiation shared across distantly-related organisms suggest common underlying causes.

The strong correspondence between distributions and underlying geology (e.g. *T.* sp. Shovelanna Hill is restricted to rockier terrain in the Hamersley subregion, while *T. basedowii* is restricted in the Pilbara to sandy soils in the Fortescue River valley) suggests substrate could be driving these patterns. Specialization for different substrates may have prevented localised diverging populations from expanding and mixing with each other, leading to divergent lineages with different substrate requirements. Earlier work by Burbidge (1959) highlighted the importance of substrate for predicting the distribution of *Triodia* species, and underlying geology and terrain features have been implicated in driving adaptive radiations in geckos (Pepper *et al.* 2008). Doughty et al. (2011) observed that lizard species assemblages in the Pilbara largely corresponded to surface type (rocky, clayey, sandy), with endemics typically not associated with sandy substrates found outside the Pilbara. Substrate differences might also explain differentiation between the Pilbara and the sandplains to the west, where the genetically and morphologically distinct *T.* sp. Peedamulla occurs, a pattern paralleled in dragon lizards (Doughty *et al.* 2012), frogs (Catullo *et al.* 2011), and geckos (Pepper *et al.* 2008).

Hypotheses for drivers of diversity patterns in the Australian arid zone

The formation of sandy deserts as new habitat

The climatic changes associated with the global shift in glacial cycle frequency and amplitude *c*. 0.8–1.2 Ma (Clark *et al.* 1999) are probably causally linked to the concurrent formation of sandy dune fields in Australia *c*. 1 Ma (Fujioka *et al.* 2009), and would have likely had a significant impact on plants and animals in the developing Australian arid zone. Our results show that species diversity in the *T. basedowii* complex is substantially lower in the sandy deserts than in rockier areas of

Western Australia. This pattern is consistent with the sandy desert habitat becoming accessible only recently. Significant deviations from neutrality for nuclear markers in *T. basedowii* (see Table 2.5) may reflect a recent range expansion in this taxon, corresponding to a spread across the newly formed dune fields. Recent range expansions in taxa inhabiting the sandy deserts have also been inferred for lizards (Fujita *et al.* 2010; Pepper *et al.* 2011). If *T. basedowii* recently expanded its range across the deserts, we would also expect a reduction in its genetic diversity (see Hewitt 1996), but levels of nucleotide diversity within *T. basedowii* are not markedly lower than for other taxa in the complex (see Table 2.5). In addition, the chloroplast marker shows no deviation from neutrality, and the observed deviations from neutrality in the nuclear markers may be explained by a selective sweep of the rDNA region rather than demographic expansion. *T. basedowii* is found in the Pilbara only along the Fortescue River valley (Fig. 2.8) on pockets of sandy soils similar to substrates across its wider range outside the Pilbara. Whether the presence of *T. basedowii* in the Pilbara reflects an ancestral source for, or incursion from, the outside sandy deserts is unclear. The dynamics, direction and timing of possible population expansion for *T. basedowii* require more investigation.

The Pilbara as a refugium or cradle of diversity?

Regions with high topographic heterogeneity such as the Pilbara may be expected to have acted as refugia for organisms impacted by climate changes, both on a long timescale corresponding to intensifying aridification through the Pliocene and Pleistocene and on a short timescale corresponding to heightened aridity during glacial cycles. The recent dated phylogeny for *Triodia* (Toon *et al.* 2015) has limited sampling within the *T. basedowii* complex, but suggests that the complex diversified within the last *c.* 5 Ma, consistent with it being subjected to these climate changes. Refugia are often identified by high levels of endemism (e.g. Fujita *et al.* 2010), and our results indicate that the Pilbara is such a region, with four of the seven taxa in the *T. basedowii* complex almost entirely restricted to the region. Thus on a long timescale (e.g. > 1 Ma, affecting species diversity), our results are consistent with the Pilbara acting as a refugium. To fully test this hypothesis, we would need to know whether the ancestors of the complex were widespread prior to the formation of the

deserts and then only persisted in the Pilbara. Given that sister taxa to the complex (*T. plurinervata*, *T.* "ipluri" and *T. concinna*) are broadly distributed in WA, and given the ecological plasticity apparent in the complex, we can speculate that the ancestors of the complex were likely widespread. As species diversity in the complex would then have been maintained in the Pilbara but not in the outside deserts, the Pilbara would have functioned as a refugium at this longer timescale.

Refugia, however, are also characterised as source areas for widespread taxa, which persist in the region during adverse climate conditions and expand from the region as climate ameliorates (e.g. Byrne 2008). Only one taxon in the complex (*T. basedowii*) is widespread outside the Pilbara, but our sampling was not sufficient to detect whether Pilbara populations of *T. basedowii* had greater genetic diversity than sandy desert populations, as would be expected if *T. basedowii* had expanded from the Pilbara. *T.* sp. Shovelanna Hill and *T. lanigera* also have significant deviations from neutrality in the nuclear markers, possibly indicating local population expansions within the Pilbara as suggested for Pilbara lizards (Pepper *et al.* 2011), but other Pilbara taxa (e.g. *T.* sp. Warrawagine) do not show the same signal or are poorly sampled (e.g. *T.* "shova"). If the Pilbara functioned as a shorter timescale refugium for species of the *T. basedowii* complex, some taxa do not seem to have been able to expand following alleviation of adverse climatic conditions. So while our results of higher species diversity and endemism in the Pilbara are consistent with a longer timescale refugial hypothesis, our results are inconclusive about whether the Pilbara functioned as a refugium for individual taxa through more recent glacial cycles.

An alternative explanation for higher diversity in the Pilbara is that the region has provided a focus for species diversification, whereas areas outside the Pilbara may have either not promoted divergence in the group or only been more recently colonized. More climatically stable or topographically complex regions are often more species rich (Dynesius & Jansson 2000). The higher diversity in the Pilbara could be a combination of increased speciation (e.g. from landscape heterogeneity) and reduced extinction (e.g. from climatic buffering) (see Jetz *et al.* 2004). Our results do not distinguish between alternate hypotheses of the Pilbara as a refugium or cradle of diversity. A more resolved temporal and phylogenetic framework for

diversification in the complex and a better understanding of the ecological and physiological tolerances of the species could be helpful in distinguishing these two hypotheses. For instance, if some Pilbara taxa are more closely related to taxa outside the Pilbara (e.g. *T.* sp. Little Sandy Desert) than to other Pilbara taxa, this would contradict the cradle of diversity hypothesis and favour the refugium. Our ITS+ETS tree suggests that, e.g., *T. lanigera* is more closely related to *T.* sp. Little Sandy Desert than to other Pilbara taxa such as *T.* "shova", but this relationship is incongruent with chloroplast data. Additional loci may clarify relationships between taxa in the complex and enable more rigorous testing of this expectation. It is possible, however, that both processes are contributing to the genetic partitioning patterns, and that attempting to distinguish between them may be oversimplifying a complex evolutionary history.

In contrast to the Pilbara, ranges in central Australia (Central Ranges in Fig. 2.8) contain only *T. basedowii*, like their surrounding deserts. Pepper et al. (2011) found a similar pattern in geckos, with the Central Ranges having only relatively young lineages compared with other range systems such as the Pilbara. Earlier hypotheses that the Central Ranges acted as a refugium (e.g. Beadle 1981a) have been questioned by Ingham et al. (2013) and Crisp et al. (2001), the latter suggesting that glacial maxima were associated with extinction of narrow-range endemics there. Our results provide no indication that the Central Ranges functioned as refugia for the *T. basedowii* complex, but more extensive sampling of *T. basedowii* populations there may allow comparisons of diversity with populations across the sandy deserts.

Conclusions

This study provides one of the first detailed analyses of genetic patterns in a group of arid-adapted plants in Australia. Our results suggest that multiple new taxa can be distinguished in the complex and these warrant taxonomic recognition, although further investigation is needed to clarify some boundaries. The complexity of relationships between taxa is highlighted by evidence of hybridization where geographic distributions abut, though co-occurrence of taxa in other locations without apparent mixing suggests some reproductive barriers. Genetic partitioning in

the complex is strongly geographically structured and exhibits similar patterns to those found in some co-occurring animals, suggesting common historical drivers. Our results support recognition of the Pilbara as a centre of arid zone biodiversity, and while they are consistent with the hypothesis that the Pilbara acted as a refugium through increasing Pliocene and Pleistocene aridity, they are inconclusive as to whether it acted as a refugium for taxa during Pleistocene glacial cycles. Our results also reveal a diversity gradient in the Australian arid zone, reflecting the impact of geologically recent climate changes, but similar careful taxonomic dissection of groups across the biota are needed to explore how common this pattern is.

Acknowledgements

We thank Rob Davis, Stephen van Leeuwen and Neil Gibson for allowing B.M.A. to accompany them on collecting trips to visit sites in Western Australia. Pauline Grierson is thanked for introducing B.M.A. to the Pilbara and for supporting the logistics of fieldwork. The Western Australian Herbarium (PERTH) is thanked for the use of its facilities and collection, and the following herbaria are thanked for loans/samples of material: AD, BRI, CANB, DNA, E, MEL, NSW, NT (abbreviations follow Thiers 2015). We thank two anonymous reviewers for helpful comments on previous versions of this manuscript. This research is partly supported by an Australian Research Council Linkage Project 120100350 (CIs Grierson, Krauss *et al*. in collaboration with Rio Tinto Iron Ore, Chevron, Botanic Gardens and Parks Authority and the Department of Parks and Wildlife). Collecting in eastern Western Australia was supported through an ANZ Trustees Foundation - Holsworth Wildlife Research Endowment. B.M.A. is supported by an International Postgraduate Research Scholarship, an Australian Postgraduate Award and a UWA Top-up scholarship at The University of Western Australia.

Supplementary Information

Supplementary information can be found in Appendix A.

References

Álvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*, **29**, 417–434.

Anderson M, Gorley R, Clarke K (2008) *PERMANOVA+ for PRIMER: guide to software and statistical methods*. PRIMER-E Ltd., Plymouth.

Anderson M, Willis T (2003) Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology*, **84**, 511–525.

Arbogast BS, Kenagy GJ (2008) Comparative phylogeography as an integrative approach to historical biogeography. *Journal of Biogeography*, **28**, 819–825.

Armstrong G (2008) *Triodia caelestialis* (Triodieae: Chloridoideae: Poaceae), a new species from the central Kimberley, Western Australia. *Journal of the Royal Society of Western Australia*, **91**, 313–317.

Barrett RL, Barrett MD (2011) Two new species of *Triodia* (Poaceae: Triodieae) from the Kimberley region of Western Australia. *Telopea*, **13**, 57–67.

Barrett RL, Barrett MD (2015) Twenty-seven new species of vascular plants from Western Australia. *Nuytsia*, **26**, 21–87.

Barrett RL, Wells GB, Dixon KW (2005) New taxa and combinations: Subfam. Chloridoideae, Trib. Triodieae, *Triodia*. In: *Flora of Australia* (ed Mallett K), pp. 458–459. CSIRO Publishing, Melbourne.

Baum D, Shaw K (1995) Genealogical perspectives on the species problem. *Experimental and Molecular Approaches to Plant Biosystematics*, **53**, 289–303.

Beadle NCW (1981a) Origins of the Australian angiosperm flora. In: *Ecological Biogeography of Australia* (ed Keast A), pp. 407–426. Dr W Junk bv Publishers, The Hague.

Beadle NCW (1981b) *The Vegetation of Australia*. Cambridge University Press, Cambridge.

Bermingham E, Moritz C (1998) Comparative phylogeography: concepts and applications. *Molecular Ecology*, **7**, 367–369.

Burbidge N (1946) A revision of the Western Australian species of *Triodia* R. Br. *Journal of the Royal Society of Western Australia*, **30**, 15–33.

Burbidge N (1953) The genus *Triodia* R.Br. (Gramineae). *Australian Journal of Botany*, **1**, 121–184.

Burbidge N (1959) *Notes on Plants and Plant Habitats Observed in the Abydos-Woodstock Area, Pilbara District, Western Australia*. CSIRO, Melbourne.

Byrne M (2008) Evidence for multiple refugia at different time scales during Pleistocene climatic oscillations in southern Australia inferred from phylogeography. *Quaternary Science Reviews*, **27**, 2576–2585.

Byrne M, Yeates DK, Joseph L *et al.* (2008) Birth of a biome: insights into the assembly and maintenance of the Australian arid zone biota. *Molecular*

Ecology, **17**, 4398–4417.

Carstens B, Brunsfeld S, Demboski J, Good J, Sullivan J (2005) Investigating the evolutionary history of the pacific northwest mesic forest ecosystem: hypothesis testing within a comparative phylogeographic framework. *Evolution*, **59**, 1639–1652.

Catullo R, Doughty P, Roberts J, Keogh J (2011) Multi-locus phylogeny and taxonomic revision of *Uperoleia* toadlets (Anura: Myobatrachidae) from the western arid zone of Australia, with a description of a new species. *Zootaxa*, **43**, 1–43.

Clark P, Alley R, Pollard D (1999) Northern Hemisphere ice-sheet influences on global climate change. *Science*, **286**, 1104–1111.

Clarke K (1993) Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology*, **18**, 117–143.

Clarke K, Gorley R (2006) *PRIMER v6: user manual/tutorial*. PRIMER-E Ltd., Plymouth.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Colli GR, Constantino R, Costa GC (2006) Lizards and termites revisited. *Austral Ecology*, **31**, 417–424.

Cracraft J (1983) Species concepts and speciation analysis. *Current Ornithology*, **1**, 159–187.

Crisp M, Laffan S, Linder H, Monro A (2001) Endemism in the Australian flora. *Journal of Biogeography*, **28**, 183–198.

Crisp MD, Mant J, Toon A, Cook LG (2015) Australian spinifex grasses: new names in *Triodia* for *Monodia* and *Symplectrodia*. *Phytotaxa*, **230**, 293–296.

Csaikl U, Bastian H, Brettschneider R *et al.* (1998) Comparative analysis of different DNA extraction protocols: a fast, universal maxi-preparation of high quality plant DNA for genetic evaluation and phylogenetic studies. *Plant Molecular Biology Reporter*, **16**, 69–86.

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

de Queiroz K (1998) The general lineage concept of species, species criteria, and the process of speciation. In: *Endless Forms: Species and Speciation* (eds Howard D, Berlocher S), pp. 57–75. Oxford University Press, Oxford.

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

Department of the Environment (2012) *Interim Biogeographic Regionalisation for Australia v. 7 (IBRA).* Australian Government, Canberra, ACT.

Department of the Environment and Water Resources (2007) *Australia's Native Vegetation: A summary of Australia's Major Vegetation Groups, 2007*. Australian Government, Canberra, ACT.

Doughty P, Kealley L, Melville J (2012) Taxonomic assessment of *Diporiphora* (Reptilia: Agamidae) dragon lizards from the western arid zone of Australia. *Zootaxa*, **3518**, 1–24.

Doughty P, Pepper M, Keogh J (2010) Morphological and molecular assessment of the *Diplodactylus savagei* species complex in the Pilbara region, Western Australia, with a description of a new species. *Zootaxa*, **2393**, 33–45.

Doughty P, Rolfe JK, Burbidge AH, Pearson DJ, Kendrick PG (2011) Herpetological assemblages of the Pilbara biogeographic region, Western Australia: ecological associations, biogeographic patterns and conservation. *Records of the Western Australian Museum*, **Supplement**, 315–341.

Doyle JJ, Dickson EE (1987) Preservation of plant samples for DNA restriction endonuclease analysis. *Taxon*, **36**, 715–722.

Durrant B, Harvey M, Framenau V, Ott R, Waldock J (2010) Patterns in the composition of ground-dwelling spider communities in the Pilbara bioregion, Western Australia. *Records of the Western Australian Museum*, **Supplement**, 185–204.

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

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

Fu Y (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Fujioka T, Chappell J, Fifield LK, Rhodes EJ (2009) Australian desert dune fields initiated with Pliocene-Pleistocene global climatic shift. *Geology*, **37**, 51–54.

Fujita MK, McGuire JA, Donnellan SC, Moritz C (2010) Diversification and persistence at the arid-monsoonal interface: Australia-wide biogeography of the Bynoe's gecko (*Heteronotia binoei*; Gekkonidae). *Evolution*, **64**, 2293–2314.

Glez-Peña D, Gómez-Blanco D, Reboiro-Jato M, Fdez-Riverola F, Posada D (2010) ALTER: program-oriented conversion of DNA and protein alignments. *Nucleic Acids Research*, **38**, W14–W18.

Gower J (1971) A general coefficient of similarity and some of its properties. *Biometrics*, **27**, 857–871.

Griffin GF (1990) Characteristics of three spinifex alliances in Central Australia. *Journal of Vegetation Science*, **1**, 435–444.

Guthrie N, Weir T, Will K (2010) Localised and regional patterns in ground-dwelling beetle assemblages in a semi-tropical arid zone environment. *Records of the Western Australian Museum*, **Supplement**, 169–184.

Halse SA, Scanlon MD, Cocking JS *et al.* (2014) Pilbara stygofauna: deep groundwater of an arid landscape contains globally significant radiation of biodiversity. *Records of the Western Australian Museum*, **Supplement**, 443–

483.

Hesse P (2011) Sticky dunes in a wet desert: Formation, stabilisation and modification of the Australian desert dunefields. *Geomorphology*, **134**, 309–325.

Hesse PP, Magee JW, van der Kaars S (2004) Late Quaternary climates of the Australian arid zone: a review. *Quaternary International*, **118-119**, 87–102.

Hewitt G (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hsiao C, Chatterton N, Asay K, Jensen K (1994) Phylogenetic relationships of 10 grass species: an assessment of phylogenetic utility of the internal transcribed spacer region in nuclear ribosomal DNA in monocots. *Genome*, **37**, 112–120.

Hurry CR, Walsh NG, Murphy DJ (2012) A taxonomic review of *Triodia bunicola* and *T. scariosa* (Poaceae: Chloridoideae), based on morphological and molecular data. *Australian Systematic Botany*, **25**, 304–312.

Ingham J, Forster P, Crisp M, Cook L (2013) Ancient relicts or recent dispersal: how long have cycads been in central Australia? *Diversity and Distributions*, **19**, 307–316.

The Inkscape Team (2012) Inkscape v. 0.48.4. http://www.inkscape.org/.

Jetz W, Rahbek C, Colwell RK (2004) The coincidence of rarity and richness and the potential signature of history in centres of endemism. *Ecology Letters*, **7**, 1180–1191.

Kearse M, Moir R, Wilson A *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.

Knowles LL, Carstens BC (2007) Delimiting species without monophyletic gene trees. *Systematic Biology*, **56**, 887–895.

Lanfear R, Calcott B, Ho SYW, Guindon S (2012) Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**, 1695–1701.

Lazarides M (1997) A revision of *Triodia* including *Plectrachne* (Poaceae, Eragrostideae, Triodiinae). *Australian Systematic Botany*, **10**, 381–489.

Lazarides M, Weiller C, McCusker A (2005) *Triodia* (K Mallett, Ed,). *Flora of Australia*, **44B**, 203–256.

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Maddison W (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523–536.

Mant J, Bayer R, Crisp M, Trueman J (2000) A phylogeny of Triodieae (Poaceae: Chloridoideae) based on the ITS region of nrDNA: testing conflict between anatomical and inflorescence characters. In: *Grasses: Systematics and Evolution* (eds Jacobs S, Everett J), pp. 213–217. CSIRO, Melbourne.

Martin H (2006) Cenozoic climatic change and the development of the arid

vegetation in Australia. *Journal of Arid Environments*, **66**, 533–563.

McKenzie N, van Leeuwen S, Pinder A (2009) Introduction to the Pilbara Biodiversity Survey, 2002–2007. *Records of the Western Australian Museum*, **Supplement**, 3–89.

Morton CM (2011) Newly sequenced nuclear gene (Xdh) for inferring angiosperm phylogeny. *Annals of the Missouri Botanical Garden*, **98**, 63–89.

Morton S, James S (1988) The diversity and abundance of lizards in arid Australia: a new hypothesis. *American Naturalist*, **132**, 237–256.

Mudelsee M, Stattegger K (1997) Exploring the structure of the mid-Pleistocene revolution with advanced methods of time-series analysis. *Geologische Rundschau*, **86**, 499–511.

Müller K (2005) SeqState. *Applied Bioinformatics*, **4**, 65–69.

Nieto Feliner G, Gutiérrez Larena B, Fuertes Aguilar J (2004) Fine-scale geographical structure, intra-individual polymorphism and recombination in nuclear ribosomal internal transcribed spacers in *Armeria* (Plumbaginaceae). *Annals of Botany*, **93**, 189–200.

Nieto Feliner G, Rosselló JA (2007) Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular Phylogenetics and Evolution*, **44**, 911–919.

Nixon K, Wheeler Q (1990) An amplification of the phylogenetic species concept. *Cladistics*, **6**, 211–223.

Padial JM, Aurélien M, De la Riva I, Vences M (2010) The integrative future of taxonomy. *Frontiers in Zoology*, **7**, 16.

Pepper M, Doughty P, Arculus R, Keogh JS (2008) Landforms predict phylogenetic structure on one of the world's most ancient surfaces. *BMC Evolutionary Biology*, **8**, 152.

Pepper M, Doughty P, Fujita MK, Moritz C, Keogh JS (2013a) Speciation on the rocks: Integrated systematics of the *Heteronotia spelea* species complex (Gekkota; Reptilia) from western and central Australia. *PLOS ONE*, **8**, e78110.

Pepper M, Doughty P, Keogh JS (2013b) Geodiversity and endemism in the iconic Australian Pilbara region: a review of landscape evolution and biotic response in an ancient refugium. *Journal of Biogeography*, **40**, 1225–1239.

Pepper M, Ho SYW, Fujita MK, Keogh JS (2011) The genetic legacy of aridification: climate cycling fostered lizard diversification in Australian montane refugia and left low-lying deserts genetically depauperate. *Molecular Phylogenetics and Evolution*, **61**, 750–759.

Peterson PM, Romaschenko K, Johnson G (2010) A classification of the Chloridoideae (Poaceae) based on multi-gene phylogenetic trees. *Molecular Phylogenetics and Evolution*, **55**, 580–598.

Pianka ER (1981) Diversity and adaptive radiations of Australian desert lizards. In: *Ecological Biogeography of Australia* (ed Keast A), pp. 1375–1392. Dr. W

Junk bv Publishers, The Hague.

Pimentel R (1981) A comparative study of data and ordination techniques based on a hybrid swarm of sand verbenas (*Abronia* Juss.). *Systematic Zoology*, **30**, 250–267.

Pisias N, Moore TJ (1981) The evolution of Pleistocene climate: a time series approach. *Earth and Planetary Science Letters*, **52**, 450–458.

Rambaut A (2014) FigTree. http://tree.bio.ed.ac.uk/software/figtree/.

Rambaut A, Suchard M, Xie D, Drummond A (2014) Tracer. http://beast.bio.ed.ac.uk/Tracer.

Ramírez-Soriano A, Ramos-Onsins SE, Rozas J, Calafell F, Navarro A (2008) Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. *Genetics*, **179**, 555–567.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092–2100.

Rieseberg L, Linder C, Seiler G (1995) Chromosomal and genic barriers to introgression in *Helianthus*. *Genetics*, **141**, 1163–1171.

Rieseberg L, Soltis D (1991) Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants*, **5**, 65–84.

Ronquist F, Teslenko M, van der Mark P *et al.* (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.

Sanger F, Nicklen S, Coulson A (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, **74**, 5463–5467.

Schlick-Steiner BC, Steiner FM, Seifert B *et al.* (2010) Integrative taxonomy: a multisource approach to exploring biodiversity. *Annual Review of Entomology*, **55**, 421–438.

Shoo LP, Rose R, Doughty P, Austin JJ, Melville J (2008) Diversification patterns of pebble-mimic dragons are consistent with historical disruption of important habitat corridors in arid Australia. *Molecular Phylogenetics and Evolution*, **48**, 528–542.

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

Sokal R, Sneath P (1963) *Principles of Numerical Taxonomy*. W. H. Freeman, San Francisco.

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.

Stanford A, Harden R, Parks C (2000) Phylogeny and biogeography of *Juglans* (Juglandaceae) based on matK and ITS sequence data. *American Journal of Botany*, **87**, 872–882.

Starr J, Harris S, Simpson D (2003) Potential of the 5' and 3' ends of the intergenic spacer (IGS) of rDNA in the Cyperaceae: new sequences for lower-level phylogenies in sedges with an example from *Uncinia* Pers. *International Journal of Plant Sciences*, **164**, 213–227.

Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *American Journal of Human Genetics*, **76**, 449–462.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.

Swofford DL (2002) *PAUP*: phylogenetic analysis using parsimony (*and other methods)*. Sinauer Associates, Sunderland, Massachusetts.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Templeton A, Crandall K, Sing C (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, **132**, 619–633.

Thiers B (2015) Index herbariorum: a global directory of public herbaria and associated staff. http://sweetgum.nybg.org/ih/.

Toon A, Crisp MD, Gamage H *et al.* (2015) Key innovation or adaptive change? A test of leaf traits using Triodiinae in Australia. *Scientific Reports*, **5**, 12398.

Walker J, Geissman J, Bowring S, Babcock L (2012) *Geological Time Scale v. 4.0*. Geological Society of America.

Western Australian Herbarium (1998) FloraBase—the Western Australian Flora. https://florabase.dpaw.wa.gov.au/.

White T, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: *PCR Protocols: a Guide to Methods and Applications* (eds Innis M, Gelfand D, Sninsky J, White T), pp. 315–322. Academic Press, Inc., San Diego.

Williams MAJ (1984) Cenozoic evolution of arid Australia. In: *Arid Australia: Proceedings of a Symposium on the Origins, Biota and Ecology of Australia's Arid Regions* (eds Cogger HG, Cameron EE), pp. 59–78. Australian Museum, Sydney.

Yeates DK, Seago A, Nelson L *et al.* (2011) Integrative taxonomy, or iterative taxonomy? *Systematic Entomology*, **36**, 209–217.

Chapter 3.

Insights into using genotyping by sequencing in a species complex: a case study using Australian hummock grasses (*Triodia*)

Abstract

Next-generation sequencing data sets are becoming increasingly accessible to researchers asking biosystematic questions, but current best practice in both choosing the type of data set and effectively analysing it is still being explored. We present a case study for the use of genotyping by sequencing (GBS) to resolve relationships in a species complex of Australian arid and semi-arid grasses (*Triodia* R.Br.), highlighting our solutions to methodological challenges in the use of GBS data. We merged overlapping paired-end reads then optimised locus assembly in the program PyRAD to generate GBS data sets for phylogenetic and distance-based analyses. In addition to traditional concatenation analyses in RAxML, we also demonstrate the novel use of summary species tree analyses with GBS loci. We found that while species tree analyses were relatively robust to variation in PyRAD assembly parameters, our RAxML analyses resulted in well-supported but conflicting topologies under different assembly settings. Despite this conflict, multiple clades in the complex were consistently supported as distinct across analyses. Our GBS data assembly and analyses improve the resolution of taxa and phylogenetic relationships in the *Triodia basedowii* complex compared to our previous study based on nuclear and chloroplast markers. The genomic results also both support and conflict with previous evidence for hybridization between species in the complex. Our approach to analysing GBS data should be useful to researchers using similar data to resolve phylogenetic relationships within species complexes.

Introduction

Next-generation sequencing data sets are becoming increasingly accessible for addressing phylogenetic and biosystematic questions. Approaches to generating these data sets (reviewed in McCormack *et al.* 2013) vary in their cost in terms of time and money, their requirement for existing genomic knowledge, their applicable evolutionary time (and hence taxonomic) scale, and the quality and sample coverage of the data. It is not always clear which approach will be most efficient and effective for addressing a given research question. In addition, analytical tools and approaches for resolving evolutionary relationships with genomic data are areas of ongoing research and testing (Bryant *et al.* 2012; Rubin *et al.* 2012; Huang & Knowles 2014; Leaché *et al.* 2015a; Mirarab & Warnow 2015; Rivers *et al.* 2016), empirical application (Eaton & Ree 2013; Pyron *et al.* 2014; Giarla & Esselstyn 2015; Gohli *et al.* 2015; Leaché *et al.* 2015b), and debate about best practice for phylogenetic inference (Edwards *et al.* 2016; Springer & Gatesy 2016).

Genotyping by sequencing (GBS; Elshire *et al.* 2011) is a next-generation sequencing approach based on restriction site-associated DNA (RAD) sequencing (Baird *et al.* 2008) and, like RAD sequencing, uses a restriction enzyme to generate thousands of loci as a reduced representation of a genome. Because it is more technically straightforward than RAD sequencing, GBS is relatively accessible for researchers lacking the equipment or protocols for custom library preparation and sequencing, and is available from commercial services. It can be used for non-model systems or in combination with a genomic reference (e.g. Alcaide *et al.* 2014). Refinements to the original method have been made by some studies (e.g. Alcaide *et al.* 2014; Grabowski *et al.* 2014), including normalising sample concentrations prior to pooling and adding size selection.

The assembly and analysis of GBS data pose bioinformatic challenges. GBS raw data (often Illumina HiSeq single-end 100 bp reads) have been commonly assembled using the UNEAK pipeline (Lu *et al.* 2013) in TASSEL (Bradbury *et al.* 2007), although some studies have used Stacks (Catchen *et al.* 2011; e.g. Sedeek *et al.* 2014) or PyRAD (Eaton 2014; e.g. Escudero *et al.* 2014) instead. Because GBS

adaptors are not specific to which end of a DNA fragment they bind to, there is the potential for read duplication, and bioinformatic assembly needs to account for this, particularly when using paired-end sequencing. Appropriate parameter values for assembly programs also need to be explored, with a recent approach (Mastretta-Yanes *et al.* 2015) advocating an exploration that aims to maximize recovered loci while minimizing error rates and intrapopulation genetic distances. Once the data set has been assembled, analyses have typically used the called single nucleotide polymorphisms (SNPs) in clustering/assignment or summary statistics methods (e.g. White *et al.* 2013; Alcaide *et al.* 2014; Baldassarre *et al.* 2014), and some studies have included a coalescent approach and/or concatenated entire loci for phylogenetic inference (e.g. Escudero *et al.* 2014; Hamlin & Arnold 2014; Gohli *et al.* 2015). A challenge at the analysis stage is handling the inherently high proportion of missing data (Fu 2014) caused by low coverage sequencing (Davey *et al.* 2011) and allelic drop-out (Arnold *et al.* 2013), and the optimal way to extract phylogenetic information from these data sets is not yet established. Newer summary species tree approaches (e.g. Liu *et al.* 2010; Mirarab & Warnow 2015; Vachaspati & Warnow 2015) have not typically been applied to reduced representation data sets, likely because these approaches require a gene tree from each locus and the loci recovered from GBS and RAD sequencing tend to be short (e.g. Leaché *et al.* 2015b). In this study of a species complex, we explore the analysis of GBS data using summary species tree approaches (among others) and address some of the methodological challenges mentioned above.

The genus *Triodia* R.Br. (Poaceae: Chloridoideae) comprises perennial hummock grasses found only in Australia (Lazarides 1997), typically in arid and semi-arid regions with < 350 mm mean annual precipitation and low-nutrient soils (Beadle 1981; Griffin 1990). As dominant components of the vegetation across much of central Australia, these grasses are ecologically important as food and habitat for animals (e.g. Kitchener *et al.* 1983; Morton & James 1988; Daly *et al.* 2008). The genus currently comprises 73 described species (Lazarides 1997; Barrett *et al.* 2005; Armstrong 2008; Barrett & Barrett 2011, 2015; Hurry *et al.* 2012; Crisp *et al.* 2015), with more awaiting description either from collections from relatively unexplored

areas or from in-depth study of taxonomically challenging complexes and broadly applied names.

The *Triodia basedowii* E.Pritz. species complex comprises two formally recognised species (*T. basedowii* and *T. lanigera* Domin) and five informally recognised taxa (see Table 3.1). Anderson *et al.* (2016; Chapter 2) have recently shown that the complex is most diverse in the Pilbara region of Western Australia, where there is at least one more unrecognised taxon (*T.* "shova") and multiple cases of putative hybridization between taxa. Morphological distinction of taxa can be difficult due to overlapping and variable character states and morphological plasticity, lack of readily available fertile material, and potential hybrid intermediates. Anderson *et al.* (2016; Chapter 2) found support for the distinction of most of the informally recognised taxa, but also cases of incongruence between nuclear and chloroplast markers and low support for relationships among some taxa in the complex. Putative taxa thought to be distinct based on morphology (*T.* "wcoast", *T.* "broad", *T.* "shovb") were not clearly supported by integrated data. Subsequent field work has identified two additional putative taxa (*T.* "Panna" and *T.* "nana") as part of the *T. basedowii* complex in Western Australia.

In this paper, we present a case study for assembling and analysing GBS data to resolve relationships among closely related species, using the *Triodia basedowii* species complex as our study system. We demonstrate a number of methodological improvements for handling and analysing GBS data that should assist researchers using this accessible and affordable technique in natural systems. In particular, we incorporate paired-end read merger as an initial step, employ the program PyRAD (and optimise assembly parameters) to reduce duplication in our data set, and apply new summary species tree analyses in addition to concatenation, coalescent and clustering analyses. We evaluate the robustness of the phylogenetic analyses to assembly parameter choices and highlight the differences between results from concatenation *vs*. summary coalescent approaches. We use our GBS results to test the distinction of species in the *T. basedowii* complex recognized in Anderson *et al.* (2016; Chapter 2) and improve the resolution of relationships among them, with implications for the taxonomy of the group and future studies of its evolution.

Table 3.1. Putative taxa in the *Triodia basedowii* species complex and their designation in this study. Names of informally recognised taxa with "sp." and a geographic epithet follow the Australian convention for undescribed taxa (see Barker 2005) and are on the Australian Plant Census, while putative taxa designated with " " were coined by Anderson *et al.* (2016; Chapter 2) or in this study.

Taxon	Designation in this study
T. basedowii E.Pritz.	*T. basedowii*
T. lanigera Domin	*T. lanigera*
T. sp. Little Sandy Desert (S. van Leeuwen 4935)	*T.* "LSandy"
T. sp. Shovelanna Hill (S. van Leeuwen 3835)	*T.* "Shov"
T. sp. Warrawagine (A.L. Payne PRP 1859)	*T.* "War"
T. sp. Peedamulla (A.A. Mitchell PRP1636)	*T.* "Peed"
T. sp. Pannawonica (B.M. Anderson & M.D. Barrett BMA 89)	*T.* "Panna"
T. "shova"	*T.* "shova"
T. "shovb"	*T.* "shovb"
T. "wcoast"	*T.* "wcoast"
T. "broad"	*T.* "broad"
T. "nana"	*T.* "nana"

Materials and Methods

Taxon sampling

We sampled 139 plants from the *Triodia basedowii* complex for sequencing (Fig. 3.1, Table S3.1), two of which were duplicated. Sample identification was based on a previous study (Anderson *et al.* 2016; Chapter 2) and the recognition of distinct morphology in the field. Two species (*T. concinna* N.T.Burb. and *T. plurinervata* N.T.Burb., 16 plants from four populations) closely related to the complex were included as outgroups, while additional outgroups (*T. wiseana* C.A.Gardner and *T. intermedia* Cheel, 8 plants from two populations) were obtained from a separate study sequenced in the same GBS run.

Ploidy determination

Ploidy was assessed for 28 populations (multiple samples) and 20 individuals (single samples) using flow cytometry. Flow cytometry has become the routine method of estimating plant ploidy variation (Dolezel *et al.* 2007; e.g. Waters *et al.* 2010). Freshly collected leaf tissue was stored at 4°C until processing (up to 3 weeks). The flow cytometry protocol closely followed the propidium iodide protocol of Doležel *et al.* (2007). Nuclei were released from fresh plant tissue by finely chopping 1–4 cm of leaf tissue with a new double-edged razor blade in 1 ml of cold Tris.$MgCl_2$-PVP buffer (200 mM Tris.HCl, 4 mM $MgCl_2$, 0.5% V/V Triton X-100, adjusted to pH 7.5). After gently aspirating the buffer to suspend nuclei, 600 µl of the homogenate was filtered through 22 µm nylon mesh (product number 03-22/14, Plastok, Birkenhead, UK), and propidium iodide (Sigma-Aldrich, Castle Hill, NSW, Australia) and RNase (Qiagen Pty Ltd, Chadstone, VIC, Australia) added, giving a final concentration of 20 µg/ml and 20 µg/ml, respectively. Samples were screened on a Becton Dickinson FACScan flow cytometer, and the fluorescence intensity of a minimum of 6000 particles was recorded.

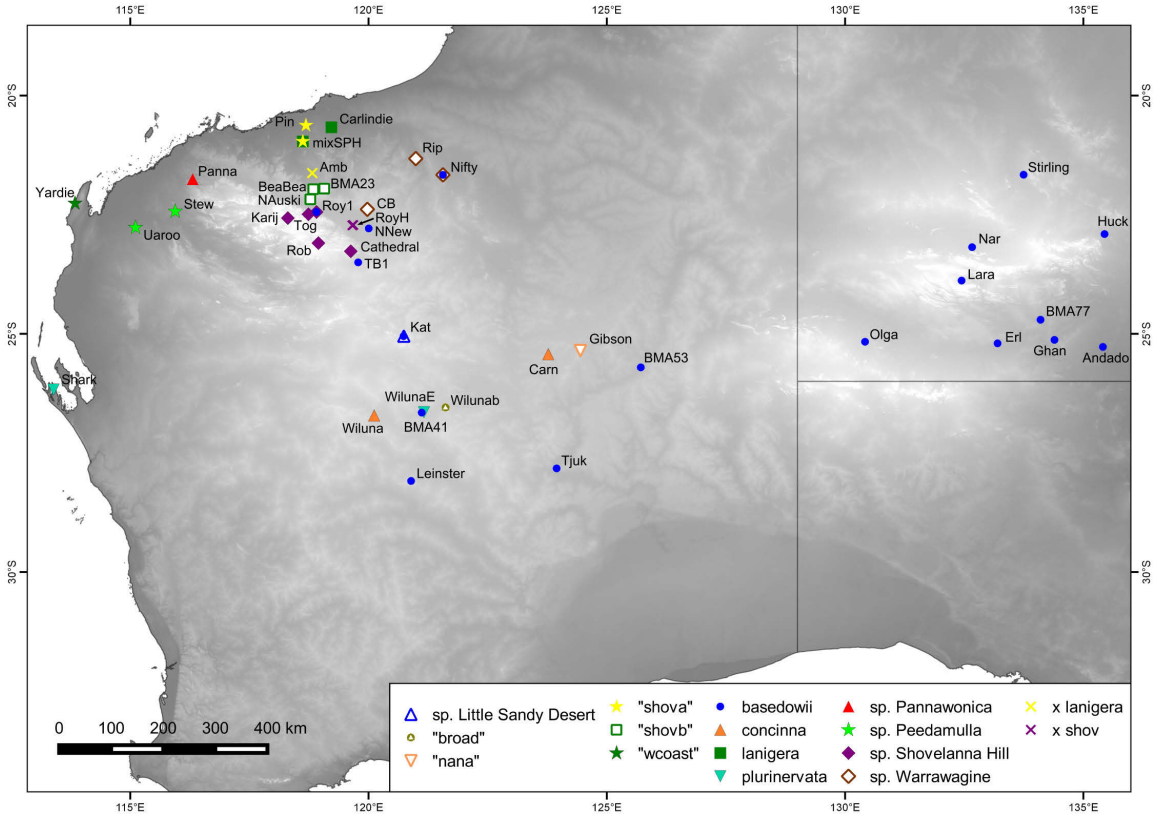

Figure 3.1. Sampling locations for populations from the *Triodia basedowii* complex and close relatives. The outgroups *T. concinna* and *T. plurinervata* are included. Map constructed with elevation base map information from (Jarvis *et al.* 2008).

One sample per population was chopped with an internal size standard (cultivated plants from *Triodia wiseana* MDB 3889 (diploid, 2C = 2.64 ± 0.04 s.e.) or *T. wiseana* MDB 3981 (tetraploid, 2C = 4.97 ± 0.07 s.e.); errors propagated through reference calculations) to obtain estimates of holoploid genome size (see Greilhuber *et al.* 2005). In most cases, these samples were prepared and measured three times to account for process and machine fluctuations, but one-off measurements are also included with notation to indicate this. *Solanum lycopersicum* L. 'Stupické polní rané' (Dolezel *et al.* 2007) was used as the reference standard (2C = 1.96 pg DNA) to establish the genome sizes of the *Triodia* reference plants, based on the average and standard deviation of five separate measurements. The remaining samples for each population were used to assess ploidy variation rather than measure genome size, so were chopped without a reference standard and the ploidy scored as diploid or tetraploid relative to an average fluorescence value of standards run separately. Observations with rarely-occurring and unusual ploidy levels were validated by repeating the measurement from new tissue.

DNA extraction, library preparation and next-generation sequencing

Silica-dried leaf material was ground in liquid nitrogen and genomic DNA extracted using a CTAB method (Doyle & Dickson 1987) as described in Anderson *et al.* (2016; Chapter 2) but with RNAse added prior to heating. DNA concentrations were quantified on a NanoDrop ND-1000 spectrophotometer and samples with concentrations between 139 ng/µl and 209 ng/µl sent to the Australian National University Biomolecular Resource Facility for library preparation and sequencing. Library preparation was modified from Elshire *et al.* (2011), with primers and barcodes from Justin Borevitz (Grabowski *et al.* 2014). Details of library preparation can be found in the supplementary material (S3.2). Briefly, 100–200 ng genomic DNA per sample was digested with the restriction enzyme PstI and custom adaptors, followed by ligation with T4 DNA ligase. Samples were purified with Ampure Beads and 80% ethanol washes, followed by individual PCR amplification and another purification step. Sample concentrations were assessed, and samples were pooled differentially to normalise concentrations. Pooled libraries were electrophoresed on an agarose gel and the 250–450 bp region cut out and purified for

sequencing. Libraries were run across three lanes of an Illumina HiSeq 2000 with 96 samples per lane (including samples from a separate *Triodia* study), using 100 bp paired-end reads.

Read processing and locus assembly

Raw paired-end reads from the ANU Biomolecular Resource Facility were demultiplexed with the sub-program *process_radtags* in Stacks v. 1.30 (Catchen *et al.* 2011) using the options -q (discard low quality reads) and -c (remove reads with uncalled bases); this program was used because it could handle combinatorial barcodes. To check for and merge overlapping reads, the demultiplexed reads were run through PEAR v. 0.9.8 (Zhang *et al.* 2014) using default settings and a 16 bp minimum overlap.

Merged and unmerged reads were assembled separately into loci for calling SNPs using PyRAD v. 3.0.6 (Eaton 2014) with the "merged" and "pairgbs" data types respectively. PyRAD has the advantage over Stacks of allowing indel variation within loci and performing reverse complement clustering. Because the primers used in paired-end GBS do not bind specifically to each end of the DNA fragments, the first and second reads are interchangeable, so further bioinformatic processing (i.e. reverse complementing the second read to join it to the first) will duplicate loci. Reverse complement clustering addresses this issue. To determine appropriate assembly settings for PyRAD, we followed the approach of Mastretta-Yanes *et al.* (2015) in using replicates and population samples to assess error rates and minimize genetic distances within populations. A subset of 40 samples, including two replicate pairs and ten populations (3–5 samples per population), was run through PyRAD for each value of six chosen parameters (Table 3.2), varying one parameter while keeping the other five at the default value.

Table 3.2. Parameters varied in PyRAD runs.

Parameter (designation)	Default value	Other values tested
#8 (d): minimum depth for a statistical base call	6	2, 3, 4, 8, 10
#9 (n): maximum number of low quality sites per read	4	2, 8
#10 (c): clustering threshold within and between samples	0.88	0.82, 0.85, 0.91, 0.94, 0.97
#13 (h): maximum number of individuals sharing a heterozygous site	3	5, 10
#25 (pl): ploidy filter = maximum number of alleles per locus	2	4, 8
#34 (mr): minimum number of reads for a dereplicate in an individual	1	2, 4, 8, 10

Parameter values were chosen to maximise the number of loci and SNPs recovered while minimising locus, allele and SNP error rates and genetic distances between samples from the same locality. Locus error is the number of loci missing in one replicate but not the other relative to the total loci; allele error is the number of shared loci differing in sequence between the replicates relative to the total shared loci; and SNP error is the number of SNPs differing between replicates relative to the total shared SNPs. Error rates and distances were calculated with custom scripts for R v. 3.2.1 (R Development Core Team 2015) and the package APE v. 3.3 (Paradis *et al.* 2004), following scripts from Mastretta-Yanes *et al.* (2015).

After choosing parameter values for PyRAD, the full sample sets were assembled using those parameter values, and error rates calculated again using the two replicate sample pairs. In order to assess the impact on downstream phylogenetic inference of clustering threshold (a parameter that was observed to have a relatively large effect on the number of loci generated), we generated additional assemblies at thresholds 0.82 and 0.97. For each of the assemblies, we generated output at 25, 50, and 75% minimum sample coverage for loci.

Phylogenetic analyses

A subset of 57 samples (4 per taxon where possible, 3 for *T.* "LSandy", 2 for *T.* "wcoast") were selected to give more even taxonomic coverage for phylogenetic inference. Unpartitioned PHYLIP files for the merged and unmerged data sets and combinations of clustering threshold and minimum sample coverage (2 data sets x 3 clustering thresholds x 3 minimum sample coverages = 18 alignments) were each analysed with RAxML v. 8.1.21 (Stamatakis 2014) using the GTRGAMMA model and a rapid bootstrap analysis (Stamatakis *et al.* 2008) with 100 bootstrap replicates and maximum likelihood (ML) search in one run (option "-f a"). The resulting trees (Figs. S3.3 and S3.4) were visually compared for conflict, and two sets of assembly settings were chosen for further analysis. First, using the chosen clustering threshold (0.88/0.91), we selected coverage levels that produced congruent topologies for merged and unmerged data sets. Second, we selected clustering thresholds and coverage levels that produced a topology incongruent with the first. This was done to

extract the most discordant phylogenomic signals observed when varying these parameters. The two assembly settings (A: 0.88/0.91 at 50% minimum sample coverage, and B: 0.88/0.82 at 25% minimum sample coverage) were used to select input PHYLIP files, which were combined (merged + unmerged) for each assembly with a custom Python v. 2.7.2 (Python Software Foundation 2016) script and analysed in RAxML with the same settings as for the separate alignments. Trees from this analysis and subsequent analyses were visualised using FigTree v. 1.4.3pre (Rambaut 2014).

Analyses on data sets with high proportions of missing data using RAxML may produce inflated support values (Simmons 2014), so we also analysed the concatenated alignments using a parsimony analysis in TNT v. 1.5beta (Goloboff *et al.* 2008). PHYLIP files from the RAxML runs were converted to TNT format with a custom Python script. Gaps were treated as missing data rather than as a fifth character. Analyses used the "xmult" command to run 10 searches for the best tree from different starting points, implementing parsimony ratchet, drift and fuse, holding up to 100 trees for each search and keeping all found trees. The resulting tree(s) was then subjected to tree bisection-reconnection branch-swapping, and a strict consensus of the resulting trees generated. Branch support was computed with 100 bootstrap replicates, each replicate consisting of 10 searches holding up to 100 trees.

Clustering and ordination

SNPs in variant call format files for the merged and unmerged data assemblies using the chosen PyRAD parameter values (data set A; data set B produced similar results and is not shown) were combined into a matrix using a custom Python script. The matrix was filtered to remove invariant and low coverage (< 50%) SNPs in R v. 3.2.1 (R Development Core Team 2015). The full SNP data set (multiple SNPs per locus) was used to calculate genetic distances between samples using the dist.dna function in the R package APE v. 3.3 (Paradis *et al.* 2004), with the default "K80" model and pairwise deletion of missing data. The resulting distance matrix was used

to build a neighbour-joining tree (NJ tree; APE bionj function) and conduct principal coordinates analysis (PCoA; built-in cmdscale function) in R.

Summary species tree analyses

We applied summary coalescent approaches, which take gene trees as input rather than DNA alignments and are able to handle large data sets. Our individual assembled GBS loci often had low coverage across samples and low information content for resolving relationships between samples, but the coalescent methods we used are able to handle polytomies in the input trees and missing taxa to some degree. We first created a PHYLIP file for each locus from the PyRAD output for the merged and unmerged data using a custom Python script. For many of the loci, RAxML (run with the GTRGAMMA model) reported a warning that there was likely not a requirement for rate heterogeneity, so we ran jModelTest v. 2.1.7 (Guindon & Gascuel 2003; Darriba *et al.* 2012) on the loci from the first chosen data set and found that 80% of loci fit a model without rate heterogeneity. To speed computation, we ran RAxML on each locus using the GTRCAT model without rate heterogeneity (-V), with 20 searches for the ML tree. For multi-locus bootstrapping (MLBS; Seo 2008), we generated the site resampling component by running 150 bootstrap replicates per locus in RAxML using the GTRCAT model without rate heterogeneity. These bootstrap trees were then sampled without replacement during 100 MLBS reps.

We used the sets of locus trees as input for ASTRAL v. 4.8.0 (Mirarab & Warnow 2015), which searches for a species tree that maximises a score reflecting the number of quartets induced by the tree that are present in the input tree set. ASTRAL was run using default settings and a name map file connecting samples to putative species. For MLBS, ASTRAL was run with the "-k bootstraps_norun" option to generate bootstrap input data sets without analysing them sequentially and the "-g" option to perform gene resampling. The resulting ASTRAL input data sets were executed multiple times in parallel to greatly speed computation. ASTRAL bootstrap trees were concatenated and summarised (using RAxML) on the tree from the first search, which was modified using a custom Python script to add branch lengths.

We also ran ASTRID v. 1.1 (Vachaspati & Warnow 2015), which computes distances between taxa averaged over the input gene trees and creates a distance matrix, which is used with a distance-based method to construct a species tree. ASTRID was run with default settings and only the gene trees as input. For MLBS, we created a custom Python script to generate random sampling of loci with replacement for 100 replicates, coupled with selecting a RAxML bootstrap tree (without replacement) for each sampled locus to emulate site resampling. The resulting 100 sets of trees were used as input to ASTRID runs to generate a set of 100 ASTRID bootstrap trees. The bootstrap trees were concatenated and summarised (using RAxML) on the tree from the first search, again modified to add branch lengths.

We also analysed our locus trees using MP-EST v. 1.5 (Liu *et al.* 2010), which estimates branch lengths in coalescent units in addition to topology (ASTRAL and ASTRID only estimate topology). MP-EST uses the topologies of rooted input gene trees to calculate a pseudo-likelihood for the topology and branch lengths of the species tree based on the sets of rooted triplets in the species tree and the input gene trees. MP-EST assumes one lineage is sampled per species and thus does not estimate terminal branch lengths, and may perform poorly with non-random missing lineages in the input gene trees. To generate rooted trees (RAxML trees are unrooted), we used a custom R script and the package APE to examine each tree for an appropriate outgroup and root the tree on that outgroup if it was present or discard the tree if it was not. The preferred outgroup was any sample of *T. wiseana*, then *T. intermedia*, then either *T. plurinervata* or *T. concinna* if the other two were not present. MP-EST was executed with three runs for each of two configurations: 1) each sample was treated as a distinct lineage, and 2) groups of samples were assigned to species. A custom R script was used to extract and convert the resulting tree from the run with the highest likelihood. MP-EST creates trees with very long terminal branches to indicate they are not estimated, so a custom Python script was used to shorten terminals for visual clarity. For MLBS, the RAxML bootstrap trees were rooted as for the first run using a custom R script (or discarded if no outgroups were present), and the rooted trees used with the same custom Python script used to generate MLBS input tree sets for ASTRID. The resulting MLBS input tree sets

were used for 100 replicates of three MP-EST runs for each of the two configurations, and the highest likelihood tree chosen from each set of the three runs as the MP-EST bootstrap tree for that replicate. The bootstrap trees were concatenated and summarised (using RAxML) on the best tree from the first run, again with branch lengths added.

Unlinked SNP species tree analyses

We applied a coalescent approach in SNAPP v. 1.2.2 (Bryant *et al.* 2012) as implemented in BEAST v. 2.3.0 (Drummond *et al.* 2012). SNAPP takes unlinked biallelic SNPs as input rather than gene trees or alignments, but it also requires that SNPs be present in every species, which presents challenges when using GBS data (with large amounts of missing data). To maximise the chance of sampling a SNP for a species, we included up to seven samples per species where possible (fewer in some species due to limited sampling). Again, data sets were generated for two clustering threshold combinations: 0.88/0.91 and 0.88/0.82. After finding that unlinked SNPs sampled by PyRAD contained a high proportion of singletons, we also created an input set of SNPs using a custom Python script, biasing the SNP selection toward biallelic SNPs which had multiple occurrences of the rare allele. Custom R scripts were used to filter SNPs (biallelic, not invariant, present in at least one sample per taxon) and create an alignment Nexus file for BEAUti (Drummond *et al.* 2012) in order to generate the xml file for SNAPP. We used default values for the priors on alpha and beta, calculated forward and reverse starting mutation rates, and set these to be sampled from the MCMC chain. SNAPP was executed with three parallel runs, each sampling every 400 generations from an MCMC chain, and was stopped manually when effective sample size (ESS) of the posterior was > 100 for most of the runs. ESS and convergence for the three runs was inspected with Tracer v. 1.6.0 (Rambaut *et al.* 2014) to confirm that runs had reached stationarity on the same likelihood value and that ESS for most parameters was well above 200 for the combined runs. We summarised a maximum clade credibility tree (with 10%, or more if the chain hadn't reached stationarity by that point, of trees discarded as burn-in) using mean heights in TreeAnnotator 2.3.0 (part of BEAST) and a cloudogram using DensiTree v. 2.2.2 (Bouckaert 2010).

We also used the unlinked SNPs as input to SVDquartets (Chifman & Kubatko 2014) as implemented in PAUP* v. 4.0a147 (Swofford 2002). SVDquartets can handle missing taxa, so we broadened the sampling compared to SNAPP by including *T. plurinervata* and *T. wiseana*. The broader sampling and the tolerance for missing taxa resulted in a larger number of SNPs for the SVDquartets analyses compared to the SNAPP analyses. We again generated two unlinked SNP input data sets: a randomly selected SNP from each locus (PyRAD output), and biased selection toward biallelic SNPs with multiple occurrences of the rare allele. We ran two SVDquartets analyses per data set, one to estimate a species tree with multiple samples assigned to each species, and the other with all samples as individual lineages. Both analyses used default settings, randomly sampled 100,000 quartets, and included 100 bootstrap replicates. *Triodia wiseana* was used as the outgroup.

Results

Ploidy determination

Samples from the *Triodia basedowii* complex and close relatives were found to be diploid or tetraploid (Table 3.3), in some cases with widely dispersed populations of tetraploids within otherwise diploid taxa.

Next-generation sequencing

Sequencing generated 123–201 million paired-end reads per lane. Across two lanes, 21 samples from the *T. basedowii* complex failed to produce enough reads for further processing, leaving 120 samples for analysis. Of the 144 samples (120 *T. basedowii* complex samples plus 24 outgroup samples) that were successfully sequenced, there were 4.56 million reads per sample on average (1.28–8.06 million; two reads per fragment).

Table 3.3. Ploidy measurements for the *T. basedowii* complex and close relatives. Standard errors based on three replicates of a single individual (not given when only one or two measurements were made). (table continued on next page)

Taxon	Population	Voucher	Mean (pg DNA)	Std. Error	Additional individuals	Ploidy
T. concinna	Carn	BMA 45	4.11	0.07	-	tetraploid
T. plurinervata	Shark	Mayence s.n.	2.42	0.05	-	diploid
T. "ipluri"	WilunaE	BMA 42	4.08	0.06	-	tetraploid
T. basedowii	NNew	BMA 31	2.21	single	8 (diploid)	diploid
T. basedowii	Niftyb	BMA 37	1.97	single	-	diploid
T. basedowii	Tjuk	BMA 58	1.85	0.33	-	diploid
T. basedowii	-	BMA 64	4.58	0.07	-	tetraploid
T. basedowii	Huck	BMA 72	2.42	single	-	diploid
T. basedowii	-	BMA 73	4.77	single	-	tetraploid
T. basedowii	Nar	BMA 74	4.45	single	-	tetraploid
T. basedowii	Lara	BMA 75	2.22	single	-	diploid
T. basedowii	-	BMA 77	2.34	0.03	2 (diploid)	diploid
T. basedowii	Ghan	BMA 79	2.34	0.03	2 (diploid)	diploid
T. basedowii	Andado	BMA 80	4.84	single	-	tetraploid
T. basedowii	Erl	BMA 81	2.35	0.04	2 (diploid)	diploid
T. basedowii	Olga	BMA 83	2.36	0.06	2 (diploid)	diploid
T. basedowii	Roy1b	MDB 3932	2.15	single	4 (diploid)	diploid
T. basedowii	Wit	MDB 4127	2.29	0.04	9 (diploid)	diploid
T. basedowii	Stirling	MDB 4520	2.19	single	-	diploid
T. "broad"	Wilunab	BMA 43	2.35	0.04	-	diploid
T. "LSandy"	Kat	BMA 62	2.14	0.03	-	diploid
T. lanigera	Pea	BMA 13	4.09	0.07	9 (tetraploid)	tetraploid
T. lanigera	SPH	BMA 14	2.08	single	8 (diploid)	diploid
T. lanigera	mixSPH	BMA 16	2.08	0.03	9 (diploid)	diploid
T. lanigera	Carlindie	MDB 4099	2.34	single	9 (diploid)	diploid
T. "shovb"	NAuski	BMA 9, 10, 11	1.92	0.03	9 (diploid)	diploid
T. "shovb"	BeaBea	MDB 4113	3.78	0.06	8 (tetraploid)	tetraploid
T. x *lanigera*	Amb	BMA 20	3.95	0.08	5 (tetraploid)	tetraploid
T. x *lanigera* shova	Amb	BMA 21	3.84	0.06	7 (tetraploid), 1 (diploid)	tetraploid
T. x shova	Amb	BMA 19	1.89	0.03	4 (diploid)	diploid
T. "shova"	mixSPH	BMA 15	1.97	0.05	8 (diploid)	diploid
T. "shova"	-	MDB 4437	1.87	double	-	diploid
T. "Peed"	Stew	MDB 3978	1.77	0.03	9 (diploid)	diploid
T. "Peed"	Uaroo	MDB 4120	1.75	0.03	9 (diploid)	diploid
T. "Peed"	-	MDB 4130	1.74	0.03	-	diploid
T. "wcoast"	Yardie	BMA 90	4.66	single	5 (tetraploid), 3 (diploid)	mixed
T. "wcoast"	-	BMA 91	4.30	0.08	4 (tetraploid)	tetraploid
T. "Panna"	Panna	BMA 89	5.15	0.11	9 (tetraploid)	tetraploid

Taxon	Population	Voucher	Mean (pg DNA)	Std. Error	Additional individuals	Ploidy
T. "nana"	Gibson	BMA 49	2.13	0.03	-	diploid
T. "Shov"	-	BMA 6	2.16	double	-	diploid
T. "Shov"	Roy1s	MDB 3933	1.95	single	9 (diploid)	diploid
T. "Shov"	Karijini	MDB 4110	2.16	0.04	8 (diploid)	diploid
T. "Shov"	-	MDB 4456	2.09	single	-	diploid
T. x shov war	RoyH2	MDB 4069, 70, 71	2.13	0.04	9 (diploid)	diploid
T. "War"	Niftyw	BMA 36	1.55	single	-	diploid
T. "War"	CB	MDB 3944	1.84	single	7 (diploid)	diploid
T. "War"	-	MDB 4073	1.73	single	4 (diploid)	diploid
T. "War"	Rip	MDB 4082	1.72	0.03	7 (diploid)	diploid

Read processing and locus assembly

Stacks demultiplexing and filtering retained on average 95.6% (93.1–97.1%) of reads per sample. PEAR merged on average 56.2% (44.6–68.4%) of demultiplexed reads, indicating size selection was not as effective as intended and that there was potentially a short fragment amplification bias. Statistics from PyRAD assembly showed that the merged data set had on average twice the read depth of the unmerged data, consistent with an amplification bias.

Choosing an optimal parameter value set was not straightforward due to the trade-offs between the number of loci/SNPs recovered, error rates (Fig. S3.5) and intra-population distances (Figs. S3.6–S3.12). Parameters with the greatest impact were clustering threshold (c) and minimum depth for a statistical base call (d). The chosen parameter values (Table 3.4) reflect a compromise for low but not the lowest error rates and intra-population distances, and an intermediate number of recovered loci/SNPs. Although polyploids are present in the data set, adjusting the ploidy filter (pl) only impacted the merged data set, possibly because the unmerged data set had insufficient read depth to recover additional alleles in polyploids. Because PyRAD does not output the additional alleles when the ploidy filter is relaxed (allowing more heterozygosity in loci instead), we chose to filter the merged data set to exclude loci with more than two recovered alleles per sample (pl=2).

Recovered loci/SNPs and error rates (Table 3.5) assessed using the two replicate samples (*T. basedowii* pop. Andado / *T.* "shova" pop. Pin) for the full data set illustrate the impact of clustering threshold and minimum taxon coverage. Higher allele error in the unmerged data probably reflects the longer loci (i.e. a longer sequence is more likely to have a sequencing error and can retain more differences when clustering based on a percentage similarity than a short sequence).

Table 3.4. Chosen parameters for PyRAD assembly of two data sets.

Parameter (designation)	Chosen merged	Chosen unmerged	Comment
#8 (d)	6	8	lower allele/SNP error; little effect on distances
#9 (n)	8	8	very little impact
#10 (c)	0.88	0.91	higher mean loci per sample, lower errors
#13 (h)	0.25 (10/40)	0.25 (10/40)	little impact
#25 (pl)	2	2	little impact (none for unmerged)
#34 (mr)	2	2	slightly lower error than mr=1 (default)

Table 3.5. Recovered loci/SNPs and error rates for various PyRAD assemblies of the full data sets. Error rates were assessed using two replicated sample pairs (*T. basedowii*, designated "bas", and *T.* "shova", designated "shova")

Clustering threshold / read set	Min. taxon cover	Loci	SNPs	Locus error (bas / shova)	Allele error (bas / shova)	SNP error (bas / shova)
0.82 / merged	25%	3867	165,307	0.070 / 0.10	0.069 / 0.049	0.0045 / 0.0028
0.82 / unmerged	25%	5092	309,890	0.084 / 0.13	0.24 / 0.20	0.0048 / 0.0044
0.82 / merged	50%	1674	80,555	0.068 / 0.11	0.079 / 0.050	0.0053 / 0.0029
0.82 / unmerged	50%	2141	147,544	0.062 / 0.12	0.23 / 0.18	0.0049 / 0.0044
0.82 / merged	75%	478	21,568	0.050 / 0.052	0.066 / 0.038	0.0034 / 0.0022
0.82 / unmerged	75%	599	40,120	0.045 / 0.053	0.19 / 0.15	0.0052 / 0.0043
0.88 / merged	25%	3505	133,643	0.082 / 0.11	0.060 / 0.044	0.0035 / 0.0024
0.91 / unmerged	25%	3690	182,838	0.081 / 0.13	0.17 / 0.13	0.0023 / 0.0019
0.88 / merged	50%	1351	58,703	0.078 / 0.11	0.059 / 0.045	0.0033 / 0.0024
0.91 / unmerged	50%	1344	78,591	0.053 / 0.12	0.15 / 0.11	0.0021 / 0.0016
0.88 / merged	75%	386	15,481	0.067 / 0.049	0.044 / 0.027	0.0018 / 0.00078
0.91 / unmerged	75%	406	23,990	0.042 / 0.067	0.12 / 0.072	0.0015 / 0.00062
0.97 / merged	25%	1271	24,284	0.15 / 0.11	0.033 / 0.037	0.0024 / 0.0031
0.97 / unmerged	25%	722	18,347	0.11 / 0.13	0.096 / 0.095	0.0028 / 0.0037
0.97 / merged	50%	287	7013	0.20 / 0.14	0.0062 / 0.025	0.00052 / 0.0011
0.97 / unmerged	50%	92	3423	0.14 / 0.21	0.077 / 0.083	0.0017 / 0.0013
0.97 / merged	75%	74	1907	0.18 / 0.095	0.020 / 0.015	0.0016 / 0.00062
0.97 / unmerged	75%	22	889	0.045 / 0.091	0 / 0	0 / 0

Phylogenetic analyses

The three clustering thesholds and three minimum sample coverage levels had substantial impact on the size of the resulting data sets and the resolution and topologies of RAxML trees for the merged and umerged data (Figs. S3.3 & S3.4). A higher clustering threshold tended to decrease the number of recovered loci; this was exacerbated by increasing the minimum sample coverage, as more variable loci were split into multiple loci that were then not present in enough taxa to pass the coverage filter. Within each clustering threshold, resolution of the RAxML trees improved with more loci and more missing data. RAxML analyses of the two chosen data sets with combined loci (merged + unmerged) produced well-supported and conflicting topologies (Fig. 3.2). The topological conflict results from the placement of two taxa, *T.* "nana" and *T.* "Panna" (shaded in the trees). Other groups in the trees remain well-supported between the two data sets, such as the sister relationship between *T.* "Shov" and *T.* "War" (100%), which was not recovered previously (Anderson *et al.* 2016; Chapter 2). A close relationship between *T. lanigera* and *T.* "shovb", with *T.* "shova" sister, is also consistently recovered with strong support (100%). Corresponding TNT parsimony trees (Fig. 3.3) support the same topologies as the RAxML trees, though with lower support for some clades in data set A and similar support to RAxML for data set B.

Clustering and ordination

After filtering, 119,449 SNPs were retained from 118 samples (excluding the two duplicates) within the *T. basedowii* complex (no outgroups). Both the NJ tree (Fig. 3.4) and PCoA (Fig. 3.5A) clearly show distinct clusters of samples with strong correspondence to *a priori* taxon identifications. The geographically widespread *T. basedowii* forms a distinct cluster. *T.* "LSandy" is distant from *T. basedowii*, with the *T.* "broad" samples clustering between them. *T.* "Shov" and *T.* "War" are clearly separated, with putative hybrids between them clustering with *T.* "Shov" rather than between the two taxa as might be expected. Samples from a population of mixed *T. basedowii* and *T.* "Shov" group distinctly with other samples from those species as they were identified in the field (with one exception shown by an asterisk). *T.* "Peed"

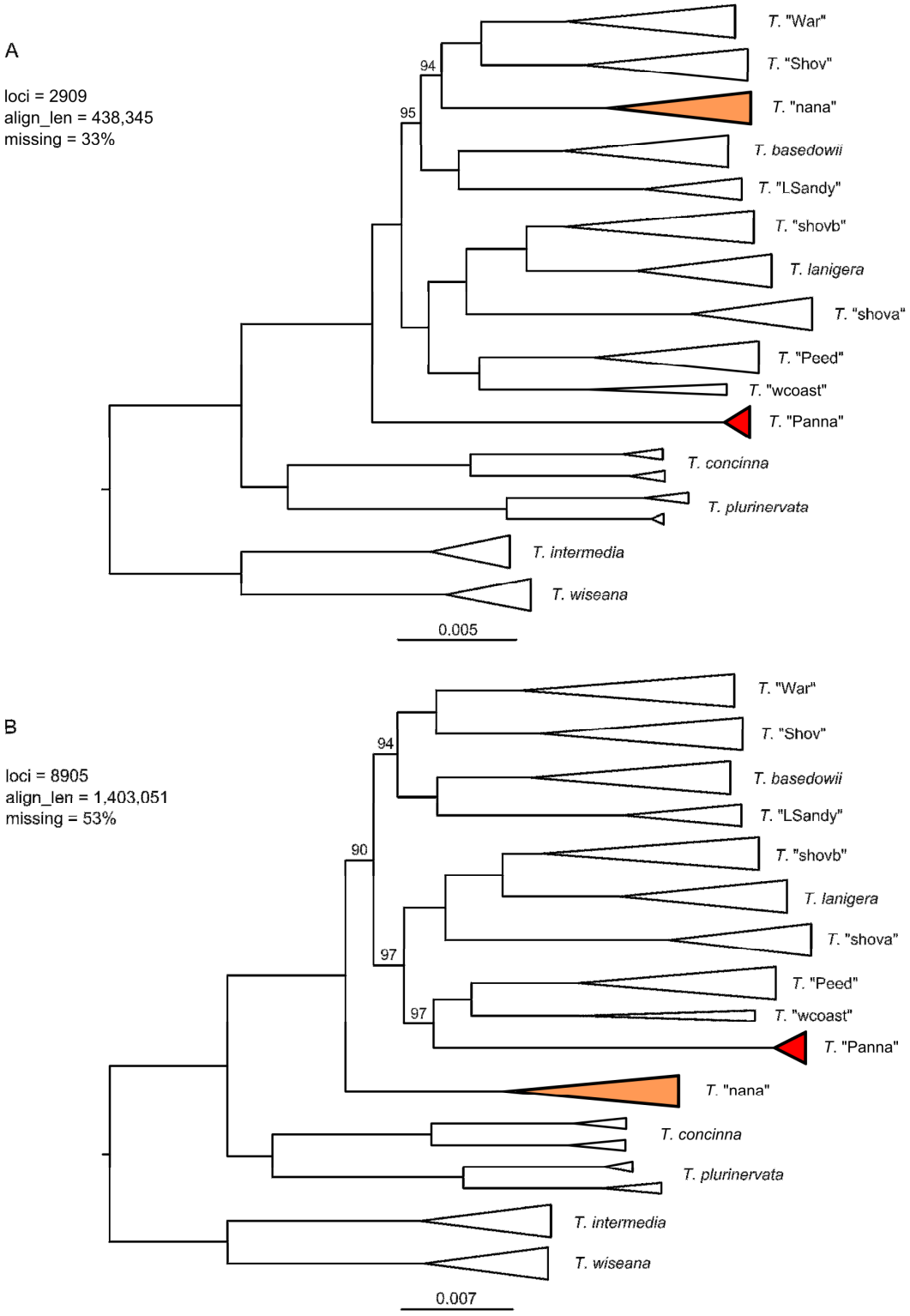

Figure 3.2. RAxML trees for the combined data set and two assemblies. A) 0.88/0.91 (merged/unmerged) clustering threshold, 50% minimum taxon coverage; B) 0.88/0.82 clustering threshold, 25% minimum taxon coverage. Support values from 100 bootstrap replicates are only shown for branches with <100% support. Scale bar units are RAxML branch lengths.

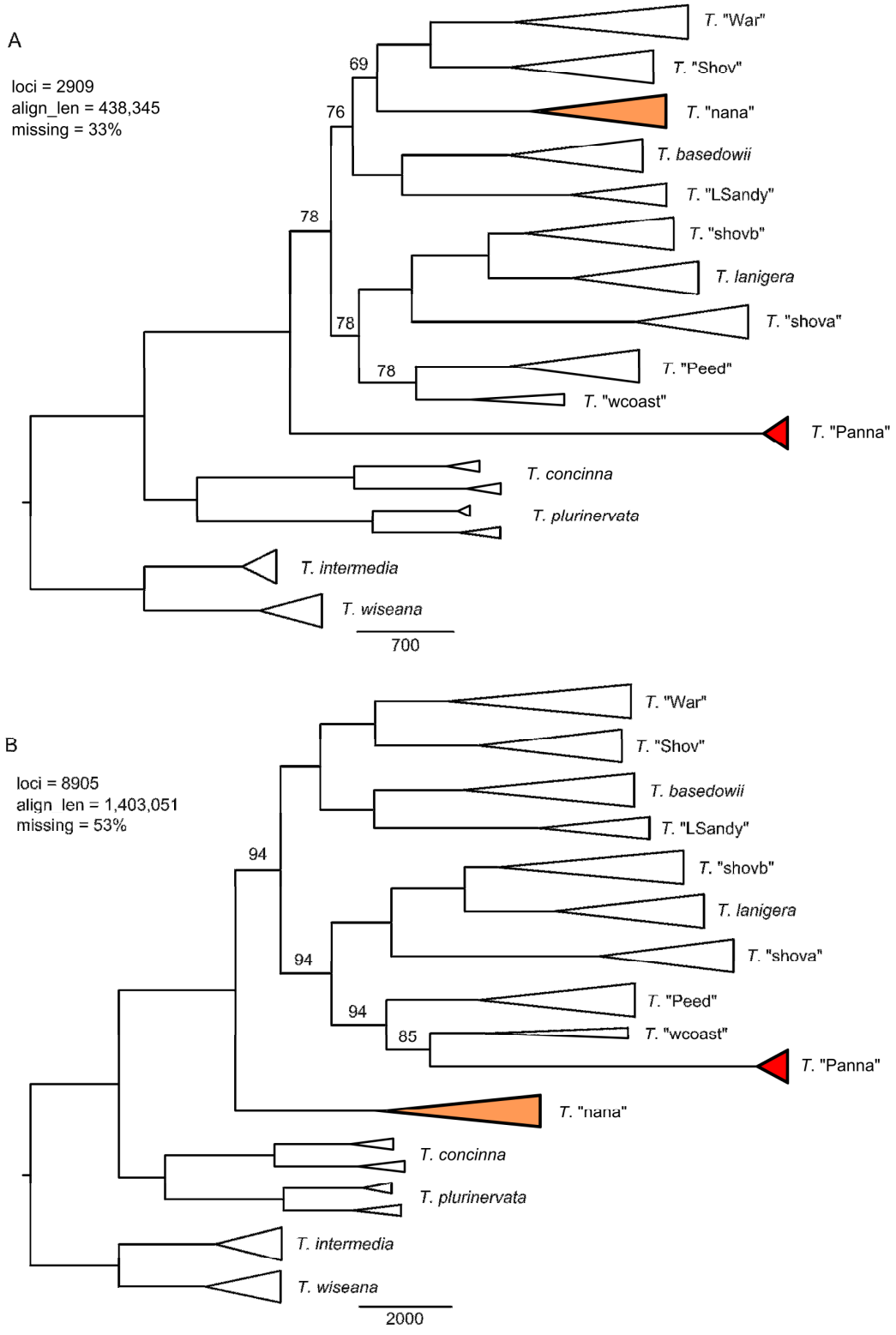

Figure 3.3. TNT trees for the combined data set and two assemblies. A) 0.88/0.91 (merged/unmerged) clustering threshold, 50% minimum taxon coverage; B) 0.88/0.82 clustering threshold, 25% minimum taxon coverage. Support values from 100 bootstrap replicates are only shown for branches with <100% support. Scale bar units are differences.

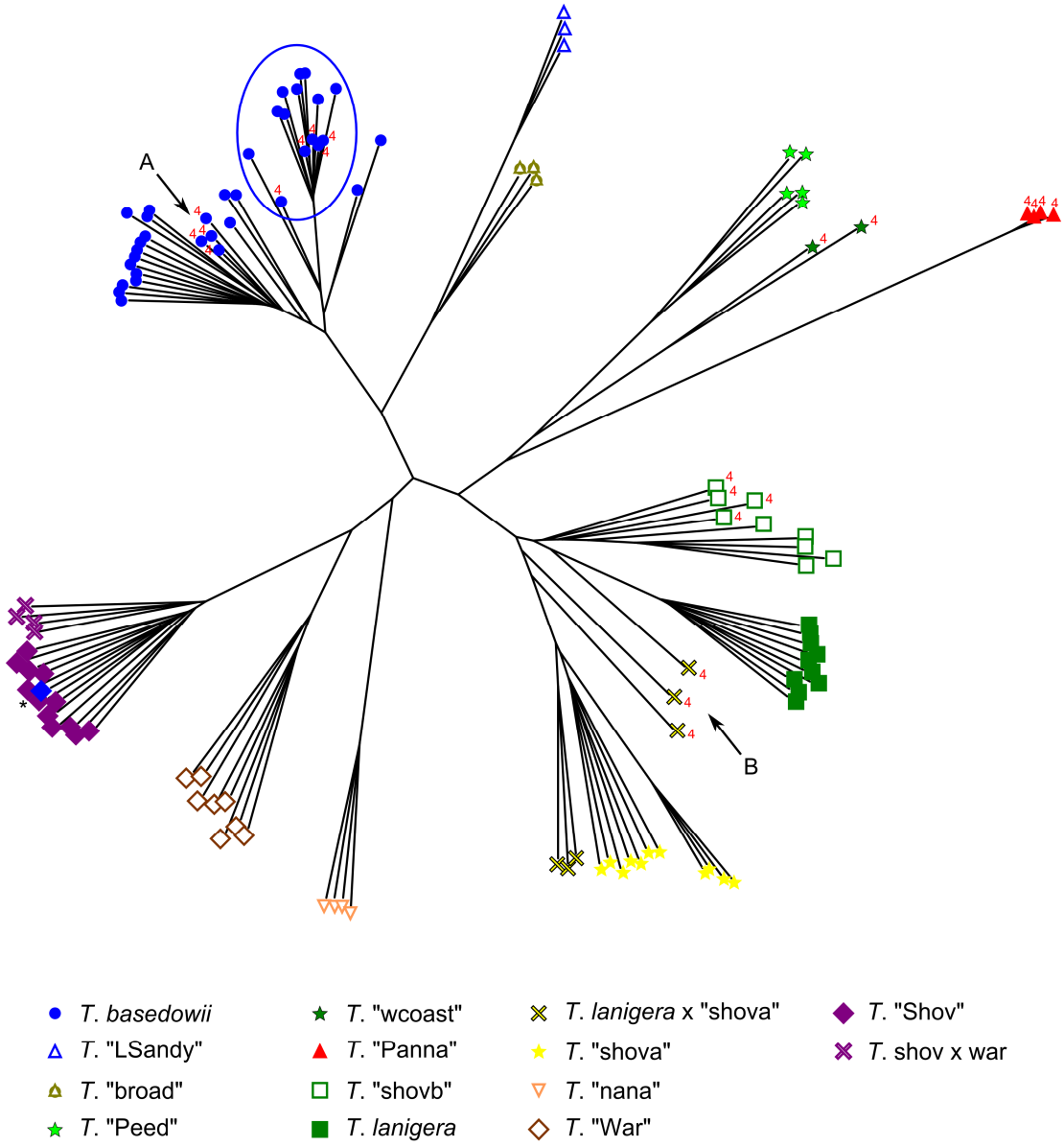

Figure 3.4. Neighbour-joining tree for samples from the *Triodia basedowii* complex. Distances are based on 119,449 SNPs, after filtering for 50% minimum taxon coverage and removing invariant SNPs. *T. basedowii* populations from central Australia are circled in blue. The asterisked sample is likely mislabelled as *T. basedowii* (co-occurring). Samples that are likely to be putative tetraploids are indicated with a red "4". (**A**) These putative tetraploid *T. basedowii* are samples from the general area of the samples used to measure ploidy. (**B**) These putative tetraploid hybrids come from a population with mixed ploidy, and are assigned ploidy based on morphology and heterozygosity. Putative tetraploids had higher heterozygosity than closely-related diploids, supporting the designations here.

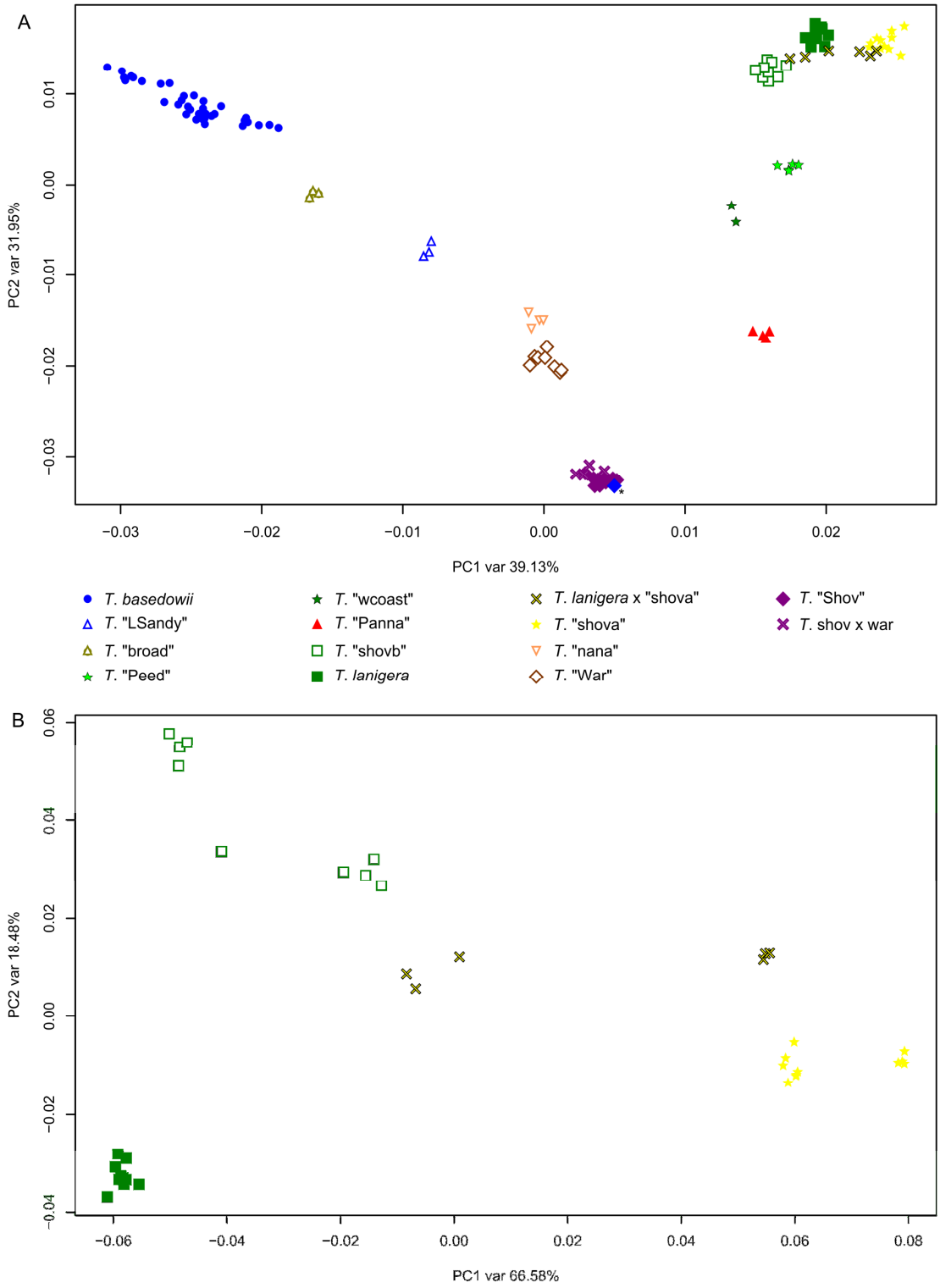

Figure 3.5. Principal coordinates analysis of samples from the *Triodia basedowii* complex. A) all samples, using 119,449 SNPs following filtering for 50% minimum taxon coverage and removing invariant SNPs; the asterisked sample is likely mislabelled as *T. basedowii* (co-occurring); B) samples of *T. lanigera*, *T.* "shovb", *T.* "shova" and putative hybrids between them, using 44,365 SNPs after filtering as in A.

and *T.* "wcoast" cluster closely but with some distinction, whereas *T.* "Panna" is quite distinct. *T. lanigera* groups closely with *T.* "shovb" and *T.* "shova" and their putative hybrids, which cluster between the parental taxa as expected (shown in greater detail in the PCoA in Fig. 3.5B, based on a subset of 36 samples and 44,365 SNPs after filtering).

Summary species tree analyses

The ASTRAL analyses of data sets A and B produced trees (Fig. 3.6) with the same topology but different levels of bootstrap support. The recovered topology is the same as the concatenated RAxML/TNT analyses of data set A. Both ASTRAL trees have 100% support for clades supported in the concatenated analyses: *T.* "Shov" sister to *T.* "War", and the clade comprising *T. lanigera* and *T.* "shovb" with *T.* "shova" sister to them, among others. Low bootstrap support is associated with the position of *T.* "nana", and to a lesser degree with the sister relationship of *T.* "Panna" to the rest of the complex.

As with the ASTRAL analyses, the ASTRID analyses of both data sets produced trees (Fig. 3.7) with the same topology and differing bootstrap support values. The topology is incongruent, however, with the concatenation analysis of data set A and the ASTRAL trees, having different placements for *T.* "nana" (now sister to the clade containing *T.* "War", *T.* "Shov", *T. basedowii* and *T.* "LSandy") and *T.* "Panna" (now sister to the clade containing *T. lanigera*, *T.* "shovb", *T.* "shova", *T.* "Peed" and *T.* "wcoast"). These shifted positions are poorly supported for data set A (75% support for the clade sister to *T.* "nana"; 69% support for the clade containing *T.* "Panna"), but have good support for data set B (90% and 89%, respectively). The lower support for the clade sister to *T.* "nana" for data set A reflects the tendency in some bootstrap replicates for *T.* "nana" to be recovered sister to *T.* "War" and *T.* "Shov" (as in the ASTRAL and concatenation trees).

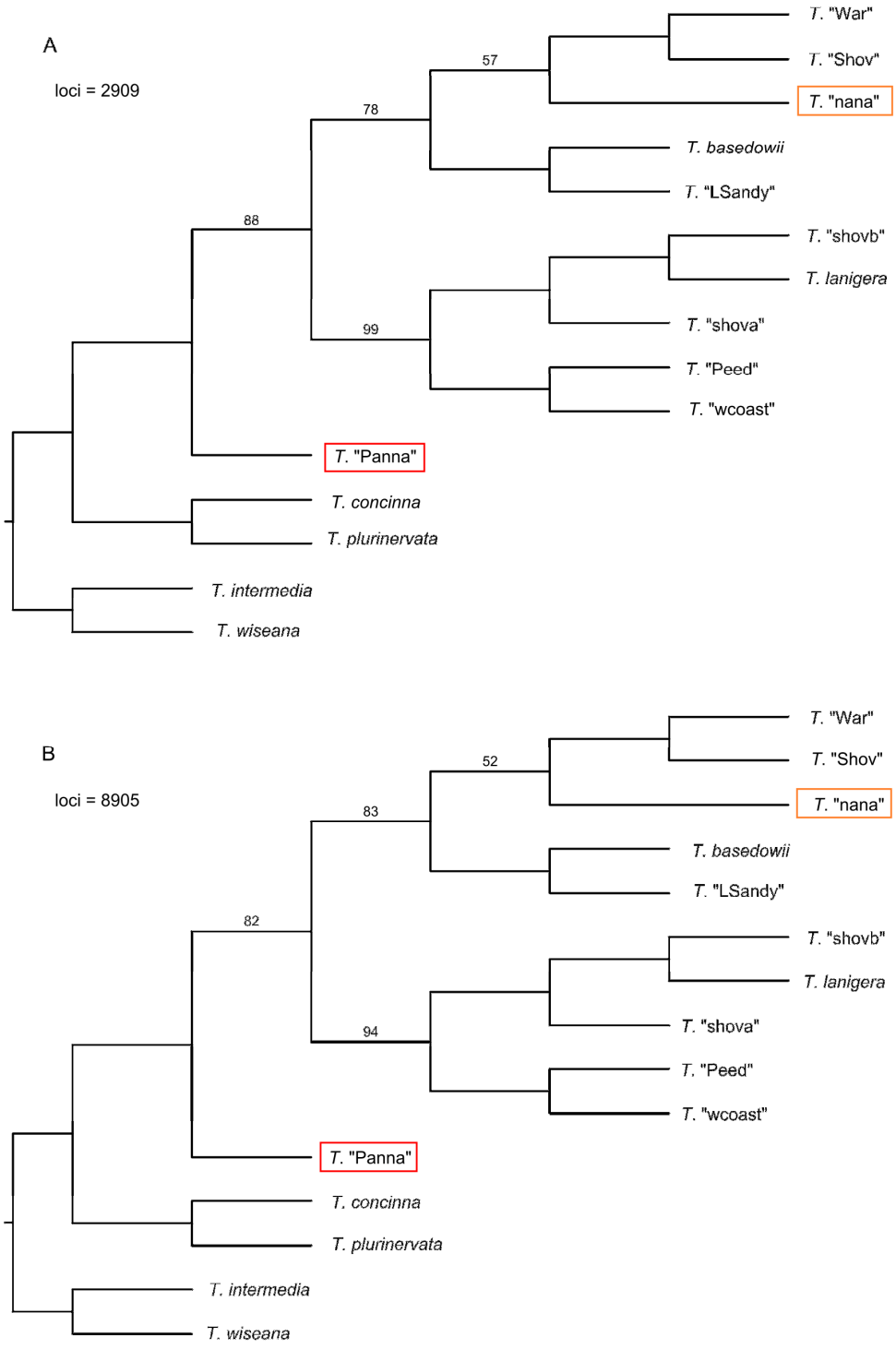

Figure 3.6. ASTRAL trees for the combined data set and two assemblies. A) 0.88/0.91 (merged/unmerged) clustering threshold, 50% minimum taxon coverage; B) 0.88/0.82 clustering threshold, 25% minimum taxon coverage. Support values from 100 multi-locus bootstrap replicates are only shown for branches with <100% support. Branch lengths are meaningless.

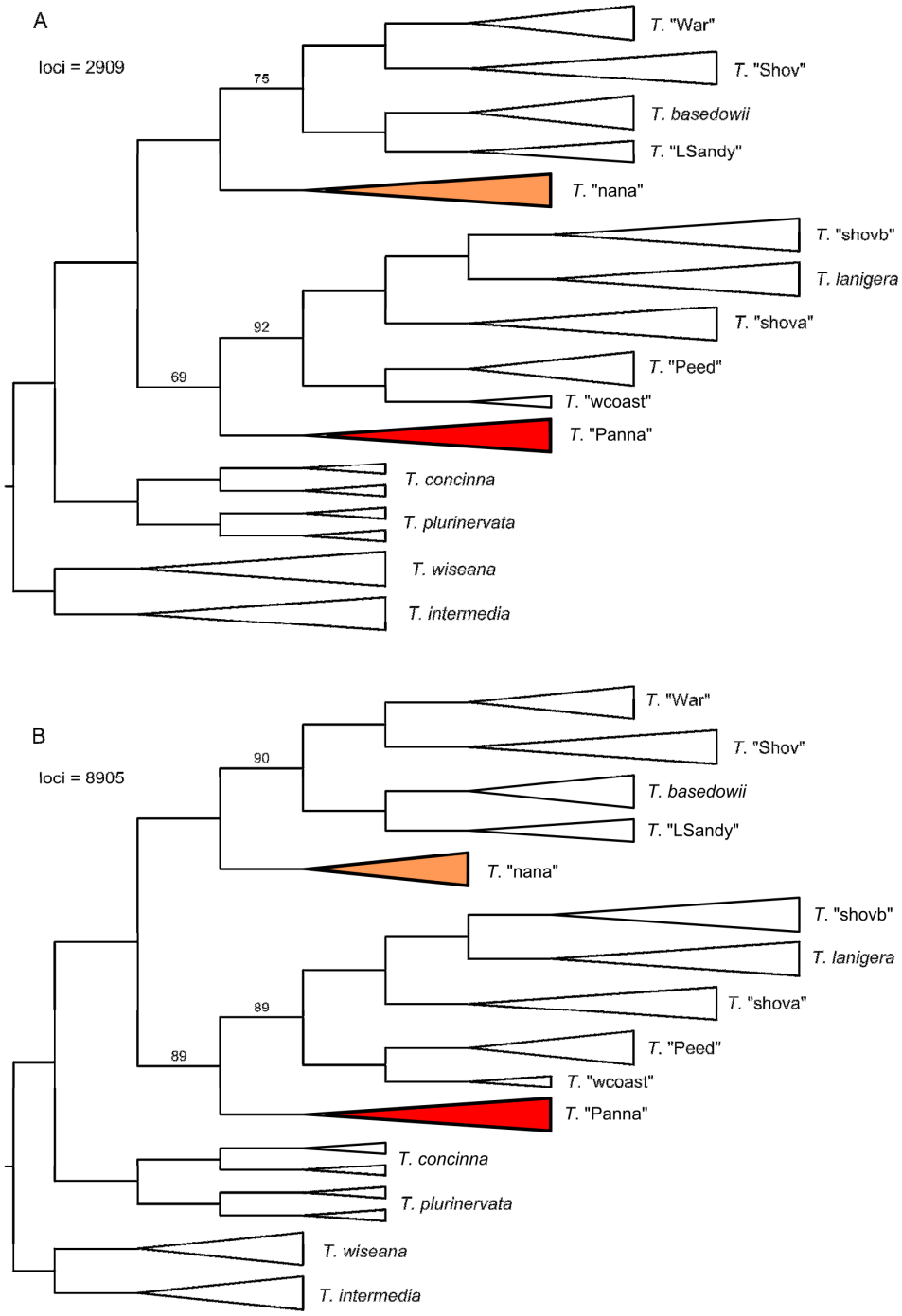

Figure 3.7. ASTRID trees for the combined data set and two assemblies. A) 0.88/0.91 (merged/unmerged) clustering threshold, 50% minimum taxon coverage; B) 0.88/0.82 clustering threshold, 25% minimum taxon coverage. Support values from 100 multi-locus bootstrap replicates are only shown for branches with <100% support. Branch lengths are meaningless.

The requirement for rooted gene trees for MP-EST resulted in the exclusion of 171 loci from data set A and 1643 loci from data set B. MP-EST analyses of both data sets resulted in largely congruent trees (Fig. 3.8), the primary difference being the placement of *T.* "nana", although with < 50% bootstrap support. Otherwise, the topology is the same as was found in the ASTRAL analyses and the concatenation analyses of data set A. Again there was strong support (100%) for the sister relationship between *T.* "Shov" and *T.* "War", and for the clade *T. lanigera* + *T.* "shovb" sister to *T.* "shova". There is marginal support (81% data set A, 76% data set B) for the sister relationship of *T.* "Panna" to the rest of the complex. Topologies and bootstrap supports were highly similar between analyses using individual samples as distinct lineages (trees in Fig. 3.8) and using groups of samples as species, the only difference in topology being the placement of *T.* "nana" for data set B (with < 50% bootstrap support). Using individual samples as species allowed MP-EST to estimate branch lengths subtending each species, whereas analysing the data with samples assigned as groups to species allowed only the estimation of branch lengths subtending two or more species.

Unlinked SNP species tree analyses

SNAPP

After filtering the randomly sampled SNPs, 983 (72% singletons) and 1335 (67% singletons) unlinked biallelic SNPs shared across species (67 samples) were retained for SNAPP for clustering thresholds 0.88/0.91 and 0.88/0.82 respectively. The SNAPP analyses resulted in highly unresolved trees (Fig. S3.13), the topologies representing less than 0.2% and 1% of the 95% HPD sets for the two clustering threshold combinations. The cloudograms show the massive phylogenetic uncertainty. Only the sister relationship of *T. concinna* to the complex and the clade *T. lanigera* + *T.* "shovb" are supported (PP=1.0) in the 0.88/0.91 tree, while the 0.88/0.82 tree also has support for the sister relationship of *T.* "shova" and *T. lanigera* + *T.* "shovb".

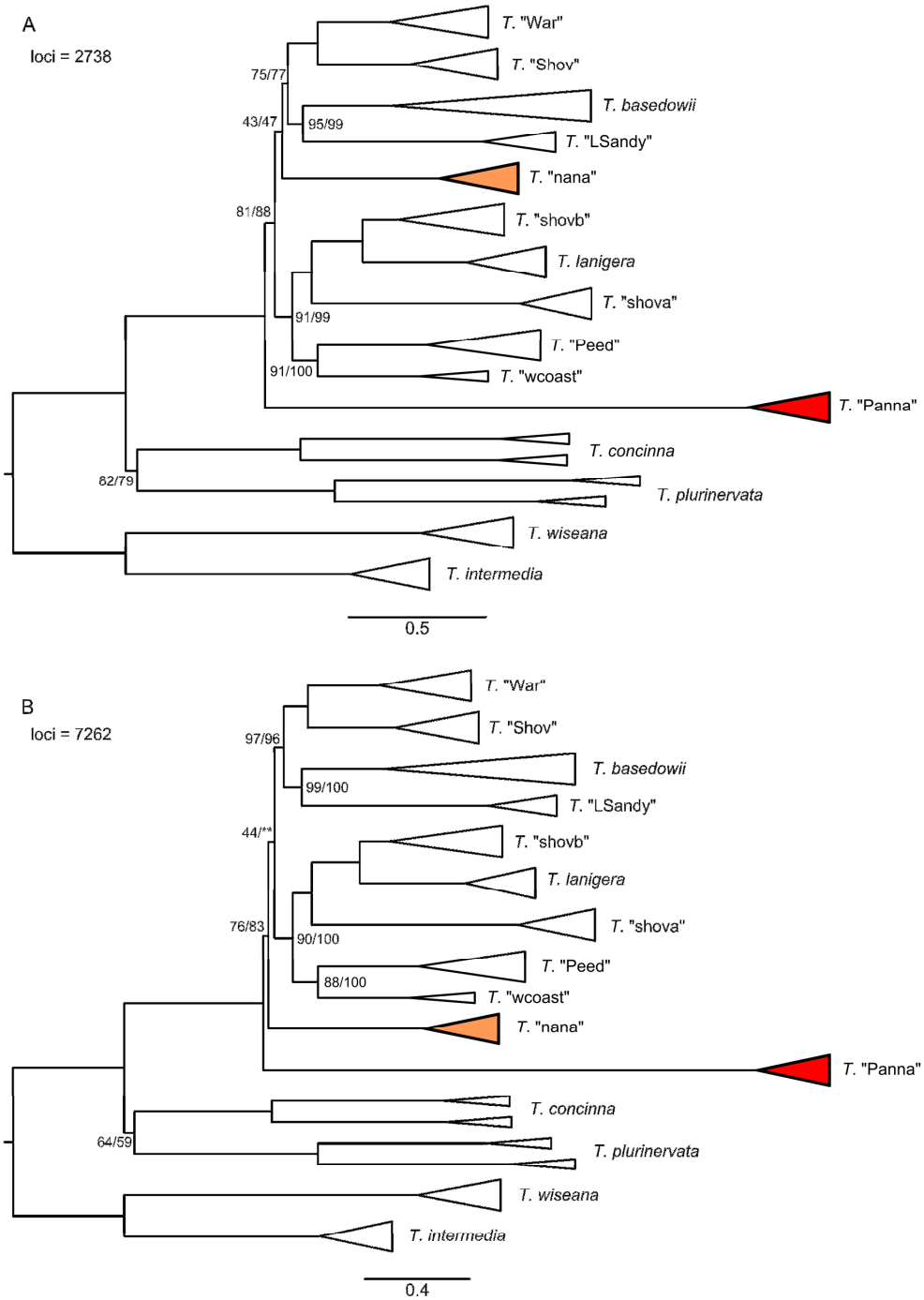

Figure 3.8. MP-EST trees (samples as species) for the combined data set and two assemblies. A) 0.88/0.91 (merged/unmerged) clustering threshold, 50% minimum taxon coverage; B) 0.88/0.82 clustering threshold, 25% minimum taxon coverage. Scale bars are in coalescent units. Terminal branch lengths are not estimated, so the length of collapsed groups (triangles) is not reliable. All shown branch lengths are estimated by MP-EST. Support values from 100 multi-locus bootstrap replicates are only shown for branches with <100% support, for both analyses (samples as species / groups of samples as species). **In the analysis with groups of samples as species, *T.* "nana" grouped with the top four taxa (as in A) with 46% support.

After filtering the biased SNPs, 986 (59% singletons) and 1293 (60% singletons) unlinked biallelic SNPs shared across species (67 samples) were retained for SNAPP for clustering thresholds 0.88/0.91 and 0.88/0.82 respectively. The resolution of the SNAPP trees (see Fig. S3.14) improved over the randomly selected SNP trees, but were again largely unresolved. The 0.88/0.91 tree (9.8% of the 95% HPD) has support (PP=1.0) for the clade of *T. lanigera*, *T.* "shovb" and *T.* "shova" but without support for the sister relationship between *T. lanigera* and *T.* "shovb". There is support (PP=0.99) for the sister relationship of *T.* "Shov" and *T.* "War" and marginal support (PP=0.97) for the sister relationship of *T. basedowii* and *T.* "LSandy". The 0.88/0.82 tree (8.5% of the 95% HPD) similarly has support (PP=1.0) for the clade of *T. lanigera*, *T.* "shovb" and *T.* "shova" with very low support (PP=0.8) for the sister relationship between *T. lanigera* and *T.* "shovb".

SVDquartets

After filtering the randomly sampled SNPs, 14,411 (65% singletons) and 15,120 (62% singletons) unlinked SNPs from 78 samples were retained for SVDquartets for clustering thresholds 0.88/0.91 and 0.88/0.82 respectively. The species and lineage trees (Fig. S3.15) are partly resolved along the backbone of the complex, and there is support for some clades, but support values are not consistent between the species and lineage trees. For example, there is 100% bootstrap support for the sister relationship between *T.* "War" and *T.* "Shov" in the 0.88/0.91 species tree, but only 76% in the lineage tree. There is stronger and more consistent support for both the above relationship and the clade of *T. lanigera* + *T.* "shovb" (99/95%) sister to *T.* "shova" (100/90%) in the 0.88/0.82 species and lineage trees, respectively. While *T.* "Peed" was assigned samples of *T.* "wcoast" in species tree analyses, the *T.* "wcoast" samples sometimes grouped separately from other *T.* "Peed" samples in the lineage analyses, though without strong support.

After filtering the biased SNPs, 15,268 (52% singletons) and 16,001 (50% singletons) unlinked SNPs from 78 samples were retained for SVDquartets for clustering thresholds 0.88/0.91 and 0.88/0.82 respectively. The species and lineage trees (Fig. S3.16) are again only partly resolved, with slightly more resolution along

the backbone for the 0.88/0.82 trees compared to the randomly sampled SNP analysis. There is strong support (98/98% and 96/91%) for the sister relationship of *T. basedowii* and *T.* "LSandy" for both clustering thresholds, in contrast to the randomly sampled SNP analysis which did not recover support for this relationship. There is again support for the clade of *T. lanigera* + *T.* "shovb" (100/95% and 100/100%) sister to *T.* "shova" (100/98% and 97/79%), but only marginal support (88/88% and <75/80%) for the sister relationship of *T.* "War" and *T.* "Shov".

Discussion

Our approach to assembling and analysing GBS data highlights common bioinformatic challenges and reveals multiple ways to extract biologically meaningful signal from these data sets, especially in non-model systems. Our results also have implications for the biosystematics of the *Triodia basedowii* complex, with support for the recognition of multiple new species and greater resolution of relationships among species than was obtained in a previous study (Anderson *et al.* 2016; Chapter 2) based on Sanger sequencing of a few loci.

Methodological insights for using GBS in a species complex

Using GBS to address phylogenomic questions presents methodological challenges that have implications for downstream analyses. While each study will face different obstacles, depending especially on the biology of the study system, we provide some common potential challenges and solutions that may be useful for researchers working with GBS data.

Sample targeting and taxon selection

When using GBS data in a species complex, preliminary sample targeting is important, as is getting a sense of what the data indicate might be taxa for more detailed analysis. While sequencing technologies are likely to continue to improve to the point that it is cost efficient and feasible to sequence massive numbers of samples from a given study system, we are not there yet, and researchers, especially those from labs with limited funding, have to carefully target samples for a GBS study. If

the study system is poorly known, effectively targeting lineages is problematic, so knowledge of the system becomes vital for generating meaningful data, underscoring the importance of natural history collections and field-based observations. We were able to efficiently target taxa in the *T. basedowii* complex for sequencing only because of previous work (Anderson *et al.* 2016; Chapter 2) using herbarium specimens and extensive field observations. For detecting potential genomic entities, we found that computationally quick distance-based analyses and even RAxML runs on a subset of the data provided a guide for whether taxa were clustering distinctly and the relative genomic divergence. Some of the species tree methods we used (e.g. ASTRAL, SVDquartets, SNAPP) rely on the researcher to specify the species membership of samples, and we recommend relatively simple and fast clustering analyses for this *a priori* sample assignment process.

Paired-end sequencing and locus duplication

One of the advantages of GBS is the technical simplicity of adaptor design, but this presents a challenge for bioinformatic assembly. Because GBS libraries are constructed with adaptors that are not specific to which end of the digested genomic DNA fragments they bind to, sequencing of very short fragments or, particularly, paired-end sequencing will lead to the duplication of loci in the data set. This duplication becomes less of an issue with single-end GBS data with effective size selection, but the benefits of added read length (increased information content per locus, which is important for phylogenomic analyses) from paired-end sequencing is lost. We addressed this duplication by using PyRAD, which, unlike UNEAK or Stacks, performs reverse complementation clustering during assembly and merges loci that had been constructed from fragments sequenced starting from opposite ends. A second source of duplication specific to paired-end sequencing occurs when there is poor size selection and small fragments are sequenced, resulting in read overlap. We observed this in our data set, where a large portion of our raw reads (56% on average) overlapped. This issue does not seem to be commonly addressed in GBS studies, but it also has the potential to lead to SNP duplication if reads overlap substantially. We addressed this overlap by merging those reads in PEAR prior to assembly with PyRAD.

Read depth and missing data

A challenge for using GBS is the large amount of missing data partly due to low coverage sequencing. In our study, this low coverage was more evident in the longer loci from unmerged reads, which had roughly half the read depth compared to the shorter loci from merged reads. We initially analysed the two data sets separately to assess consistency of phylogenetic and distance-based signals before combining them. Complicating this, our samples came from plants with differing ploidy levels, increasing the number of alleles that might be missed by the low coverage sequencing. While ploidy could be a considerable problem in our system (with both diploid and tetraploid populations in the same species), recent work (Ilut *et al.* 2014) suggests it is a minor concern for these data sets and is largely addressed by ploidy filters such as we applied through PyRAD. Our study focused on detecting genomic divergence and phylogenetic signal (potentially less impacted by missing alleles within an individual), but other researchers may be interested in population genetics questions that require more accuracy in detecting all alleles for a given locus. Due to the low coverage, GBS may not be ideal for accurate detection of heterozygotes (see Davey *et al.* 2011), needed for generating population genetics summary statistics, and in non-model systems GBS data sets might be better substituted with an approach with greater sequencing depth, such as RAD sequencing (Baird *et al.* 2008).

The missing data seem to have particularly affected our use of SNAPP. Because SNAPP requires that a SNP be present in each terminal taxon and we were analysing eleven taxa, the data set of available SNPs was massively decreased, possibly contributing to the poor resolution observed. The poor resolution may also be due to the limitation to a single SNP per locus, a limitation shared by SVDquartets, which also produced largely unresolved trees. It may be that while the analyses that can use multiple parsimony informative sites from each locus (e.g. RAxML, ASTRAL) can recover strong genomic signal, the restriction to a single SNP per locus erodes signal and SNAPP is unable to resolve relationships, especially in areas of the tree with short branch lengths characteristic of rapid divergence. SNAPP has been used successfully with GBS (e.g. Gohli *et al.* 2015), so it may also be that idiosyncracies

of relationships in the *T. basedowii* complex contributed to the lack of resolution (i.e. real biological incongruence). For GBS data sets with a large amount of missing data and study systems with closely related species, we recommend including analyses that can use multiple SNPs per locus as an alternative to SNAPP.

Optimising assembly parameters

When first approaching bioinformatic assembly of loci, it is not always clear what value is most appropriate for a given parameter in a software program like PyRAD or Stacks. We found that the assembly parameter optimisation approach based on Mastretta-Yanes *et al.* (2015) was useful in helping to select reasonable parameter values and get a sense of how our data were impacted by each parameter. The inherent trade-offs between the number of loci and SNPs recovered and error rates and genetic distances made deciding on optimal parameters subjective, and researchers should explore the impact of assembly parameters to tailor the selection for their purposes (e.g. minimising error rates may be more important for a population genetic study). In our study, assembly parameters (particularly clustering threshold and minimum sample coverage) impacted downstream phylogenetic and species tree analyses, especially concatenation approaches. Our preliminary RAxML runs across a broad range of parameter values (Figs. S3.3 & S3.4) clearly show that there are conflicting signals in the genomic data sets for different values. We recommend selecting an optimal parameter set following the approach of (Mastretta-Yanes *et al.* 2015), but also exploring how key parameters impact phylogenetic inference to get a better sense of uncertainty in the data set.

Phylogenetic signal and conflict

Bioinformatic assembly of GBS loci may produce data sets with conflicting phylogenetic signals. We specifically looked for phylogenetic conflict across assembly parameters to evaluate how various analyses handled it. For the combined data, RAxML analyses produced well-supported but conflicting topologies depending on the amount of missing data and loci, while TNT analyses better reflected (indicated less support) the uncertainty in the tree recovered for data set A, though they still recovered conflicting topologies between data sets. In contrast,

while the species tree analyses differed in their support values between data sets, they produced largely congruent trees. This difference between methods might reflect the better performance of coalescent approaches that account for gene tree conflict in data sets with large amounts of incomplete lineage sorting (likely in a group of closely related species) compared to concatenation (Kubatko & Degnan 2007; Edwards *et al.* 2007). While species tree approaches typically require more computational time (partly because gene trees have to be generated for every locus prior to the analysis), we found them valuable to incorporate alongside traditional phylogenetic approaches for comparison, and useable with reduced representation genomic data of decent read lengths.

While we recovered some degree of resolution for relationships in the complex (Fig. 3.9), the topological conflicts and low support for some nodes in the trees may reflect real biological conflict caused by incomplete lineage sorting and/or hybridization between recently diverged species. Even large genomic data sets may be unable to resolve bifurcating patterns when the real process is substantially reticulate or rapid; in such cases, the lower support reflects valuable biological information about divergence patterns in the study system. Favouring results that give high topological support (e.g. RAxML with concatenated data) is therefore potentially problematic for understanding natural systems, and researchers would benefit from the inclusion of methods that may better reflect real uncertainty (e.g. lower support values in species tree approaches).

Despite differences, the various analyses consistently supported certain clades (highlighted in Fig. 3.9) regardless of assembly parameters or method of analysis, suggesting there is strong genomic signal supporting their distinction. Across all the well-supported topologies, there is a common trend of more resolution for data set B, which had more missing data and more loci than data set A. Greater resolution with more missing data (and therefore more loci) has been previously highlighted for large data sets (Streicher *et al.* 2015). Applying strict filtering for sample coverage quickly reduces the size and phylogenetic informativeness of GBS data sets, so researchers are encouraged to consider low sample coverage thresholds when inferring phylogenies. We also found that setting a strict filtering level for the

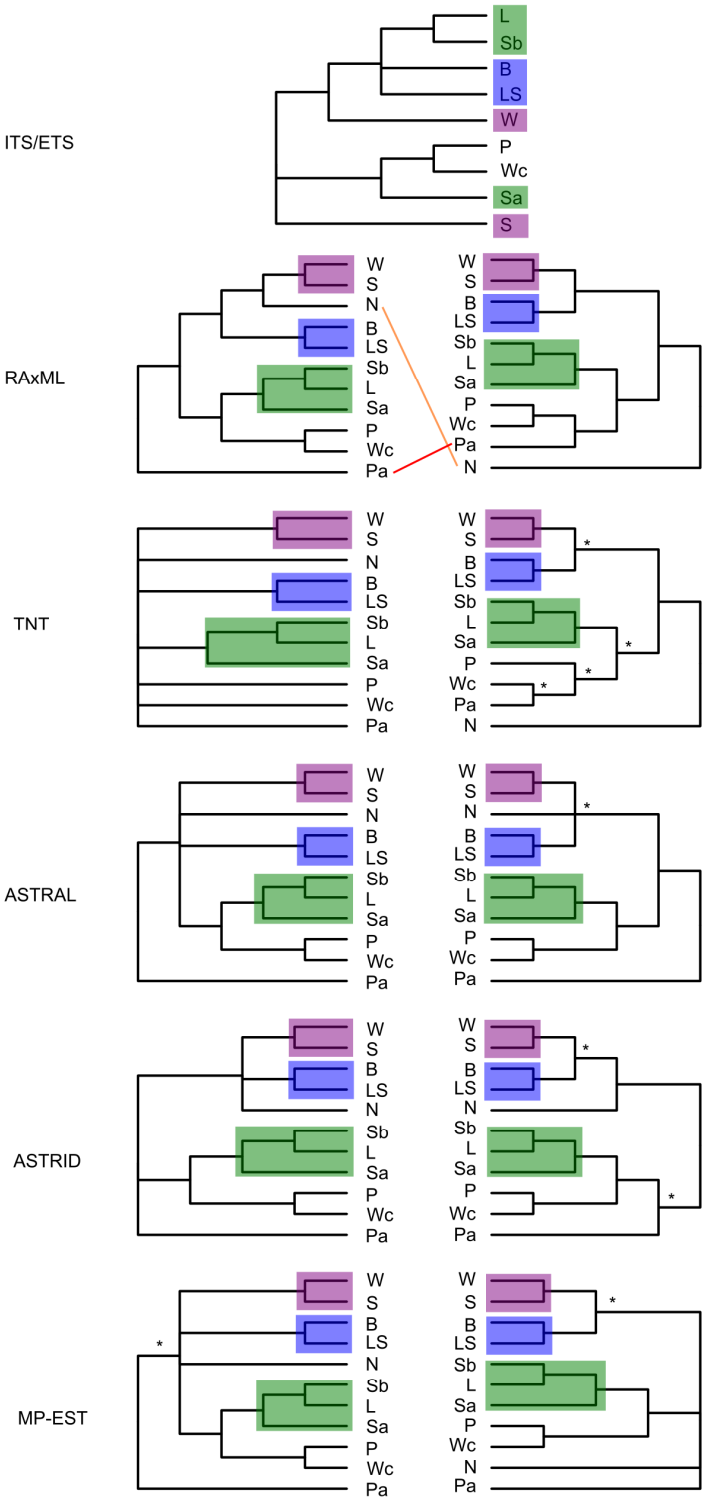

Figure 3.9. Topologies for the *Triodia basedowii* complex across analyses for two assemblies. (caption continued on next page)

Figure 3.9. (cont.) Left) 0.88/0.91 (merged/unmerged) clustering threshold, 50% minimum taxon coverage; right) 0.88/0.82 clustering threshold, 25% minimum taxon coverage. Branches with less than 80% support are collapsed. Asterisks designate clades not present in the alternate assembly. Clades consistent across analyses are highlighted. B: *T. basedowii*, L: *T. lanigera*, LS: *T.* "LSandy", N: *T.* "nana", P: *T.* "Peed", Pa: *T.* "Panna", S: *T.* "Shov", Sa: *T.* "shova", Sb: *T.* "shovb", W: *T.* "War", Wc: *T.* "wcoast". ITS/ETS results from (Anderson *et al.* 2016; Chapter 2).

number of SNPs in a locus (to remove poorly aligned or repetitive regions) removed good loci and phylogenetic signal. While a small number of SNPs per locus is a reasonable filter for studies on one or two species, it can be detrimental to phylogenomic studies with multiple divergent species and outgroups. In our study system, we found that loci in one species might differ from another at a few positions, but differ from a different relative at other positions, often resulting in many SNPs (and phylogenetically informative positions) recovered per locus. As such, researchers working in similarly multi-species systems would benefit from not setting strict limits on the number of SNPs in a final locus.

The Triodia basedowii *species complex*

Our GBS data provide a first measure of genomic variation in the *Triodia basedowii* species complex and new insights into the distinction of taxa and their evolutionary relationships. There is a close correspondence between genomic entities and *a priori* taxon identifications, across large geographical ranges and even where taxa co-occur, consistent with the hypothesis that these taxa are distinct species.

Well-supported clades recovered in phylogenetic and species tree analyses are both congruent and incongruent with those found in Anderson *et al.* (2016; Chapter 2). In some cases, GBS results provide support where previous results were ambiguous. For example, Anderson *et al.* (2016; Chapter 2) hypothesised a sister relationship between *T.* "Shov" and *T.* "War" based on morphology and the presence of shared ITS copies, but found no or ambiguous support for this in molecular phylogenies using ITS/ETS and a chloroplast marker. Our GBS analysis strongly supports this sister relationship, with phylogenetic and species trees always recovering it and clustering analyses showing the two entities to be close but distinct. Similarly, Anderson *et al.* (2016; Chapter 2) were unable to resolve *T.* "LSandy" as distinct from *T. basedowii* despite morphology and field observations suggesting that it is. In our GBS trees, *T.* "LSandy" and *T. basedowii* were consistently recovered as sisters, with clustering analyses and branch lengths in concatenation and MP-EST trees suggesting substantial divergence between them. Anderson *et al.* (2016; Chapter 2) also found insufficient evidence to recognize the putative taxon *T.* "shovb" as

distinct from *T. lanigera*, and the ITS/ETS analysis placed the combined taxon as part of a polytomy with *T. basedowii*. Our GBS results consistently recover their close relationship and limited divergence, supporting the former conclusion (we continue to treat them as divergent populations of a single species), but place them as sister to *T.* "shova" rather than *T. basedowii*. This agrees with chloroplast data (Anderson *et al.* 2016; Chapter 2), suggesting that the ITS result may represent introgression from *T. basedowii*.

The new putative taxa *T.* "nana" and *T.* "Panna" were consistently distinct across analyses but unstably placed in the GBS trees. Branch lengths subtending these two taxa are consistently long (especially for *T.* "Panna"), suggesting that these morphologically distinct entities should be recognised as distinct new species in the complex. *T.* "Panna" is of particular importance as it is known from a single population, and, despite being a tetraploid, has the lowest heterozygosity for taxa in our study based on statistics from the PyRAD assembly. This suggests that the population has experienced a severe bottleneck, which together with its restricted distribution makes it of conservation concern.

Ploidy variation in the *Triodia basedowii* complex does not follow taxonomic boundaries, suggesting that polyploidy events represent sporadic cases of autopolyploidy without subsequent long-term persistence and divergence. Our GBS analyses show that the widespread *T. basedowii* forms a distinct genomic cluster, with some distinction between populations but without tetraploid populations clustering distinctly from diploid populations (e.g. central Australian populations cluster together despite different ploidy levels). Similarly, a population of *T.* "wcoast" comprises mixed diploid and tetraploid plants, suggesting random autopolyploidy events rather than an established polyploid lineage. Multiple ploidy levels are also found in *T. lanigera* (including some of its *T.* "shovb" populations) and in a population of putative hybrids between tetraploid *T. lanigera* and diploid *T.* "shova", suggesting that differing ploidy is not acting as a strong reproductive barrier in this case. Outside the complex, *T. plurinervata* also comprises diploid and tetraploid populations, and heterozygosity statistics for PyRAD loci (elevated in

tetraploids in the unmerged data) suggest there may be diploid plants in the tetraploid *T. concinna* population.

Our species tree analyses are helpful in assessing the distinction of entities that are dispersed or not sister in the trees (e.g. *T.* "Shov" from *T. basedowii*), but because we designated taxa *a priori*, these analyses do not help to assess whether sisters are distinct (e.g. *T.* "Peed" from *T.* "wcoast"). Our decision to recognise taxa in these cases depends on assessing the degree of genomic divergence (from clustering analyses and phylogenetic branch lengths) along with supporting evidence (from e.g. morphology or distribution). For example, the geographically disjunct putative taxa *T.* "Peed" (diploid) and *T.* "wcoast" (tetraploid/diploid) have some morphological distinction but are indistinguishable in ITS/ETS (Anderson *et al.* 2016; Chapter 2). Our GBS results suggest moderate divergence between the two entities, and they do not always form a single clade (e.g. paraphyletic in Fig. 3.3B, Figs. S3.13 & S3.14), but our sampling of *T.* "wcoast" is relatively poor. With these ambiguous results, we conservatively treat them as divergent populations of one species, pending more extensive sampling. This example contrasts with that of *T. basedowii* and *T.* "LSandy", where there is more clearly morphological distinction, a more substantial degree of genomic divergence, and co-occurrence without evident mixing.

Localities where putative taxa co-occur provide a useful test for species distinction and an opportunity to assess the frequency of hybridization. Anderson *et al.* (2016; Chapter 2) provided evidence of hybridization between some taxa in the *T. basedowii* complex, and our GBS results both support and contradict that evidence. GBS results from one population ("Amb") support previous evidence for hybridization between *T. lanigera* and *T.* "shova". Individuals with intermediate morphologies clearly form genomic clusters between the putative parental species in the PCoA (Fig. 3.4B), with those identified as more like *T.* "shova" clustering closer to that taxon and similarly for *T. lanigera*. In contrast, a population ("mixSPH") containing both *T. lanigera* and *T.* "shova" that showed evidence of some ITS copy sharing (Table 3.6) did not show evidence of genomic mixing in our results. Multiple *T.* "shova" individuals possess both *T. lanigera* and *T.* "shova" ITS copies, while the same cannot be said for *T. lanigera* individuals, and our GBS results place all

Table 3.6. ITS copy types for co-occurring *Triodia lanigera* and *T.* "shova". Sequences sometimes had heterozygous positions so that they differed slightly from the references, but were still identifiably either of the two copy types. Samples with "SPortHedland", "mixLSPH" or "mixSPH" in their names are from the "mixSPH" mixed population.

Sample name	Taxon	ITS copy type	GenBank reference	Note
T_lan_Barrett_4107_GillamCk	*T. lanigera*	A	KU173242	*T. lanigera* reference
T_lan_Anderson_16_SPortHedland	*T. lanigera*	A		pop. voucher
T_lan_Anderson16_mixLSPH_2	*T. lanigera*	A		
T_lan_Anderson16_mixLSPH_6	*T. lanigera*	A		
T_lan_Anderson16_mixLSPH_8	*T. lanigera*	A		
T_lan_Anderson16_mixLSPH_10	*T. lanigera*	A		
T_lan_Anderson16_mixLSPH_14	*T. lanigera*	A		
T_shova_Barrett_4106_GillamCk	*T.* "shova"	B	KU173287	*T.* "shova" reference
T_shova_Anderson15_mixSPH_3	*T.* "shova"	B		
T_shova_Anderson15_mixSPH_7	*T.* "shova"	B		
T_shova_Anderson15_mixSPH_9	*T.* "shova"	B		
T_shova_Anderson15_mixSPH_17	*T.* "shova"	B		
T_shova_Anderson_15_SPortHedland	*T.* "shova"	A + B		pop. voucher
T_shova_Anderson15_mixSPH_11	*T.* "shova"	A + B		
T_shova_Anderson15_mixSPH_15	*T.* "shova"	A + B		
T_shova_Anderson_19_TwoCamelCk	*T. lanigera* x "shova"	A + B	KU173284, KU173285	hybrid reference

samples from this population with their respective taxa and without suggestions of genomic intermediacy.

The contrasting results suggest that there is limited introgression of rDNA despite relatively strong barriers across the rest of the genome. Similarly, a population ("RoyH") of individuals found to possess ITS copies characteristic of *T.* "Shov" and *T.* "War" showed no evidence of genomic mixing in our results following resampling. Samples from the population clustered with *T.* "Shov", suggesting that introgression may be restricted to rDNA or that our second sampling for GBS (only four plants) comprised parental plants rather than hybrids, if the population contained both. In other populations with co-occurring taxa (e.g. "Roy1" comprising *T.* "Shov" and *T. basedowii*, with one probable misidentification, and "Nifty" comprising *T.* "War" and *T. basedowii*) there is no evidence of genomic mixing or rDNA introgression, which suggests these taxa are distinct.

Conclusion

By addressing GBS bioinformatic limitations, such as overlapping reads following imperfect size selection and paired-end read locus duplication, a biologically meaningful phylogenetic signal from across the genome can be extracted for study systems lacking prior genomic knowledge. The continuing improvement in next-generation sequencing read length holds promise for using analytical methods that rely on extracting phylogenetic signal from each locus, such as new species tree approaches. Here we have demonstrated their use with our assembled GBS loci and shown that they are more robust to variation in assembly parameters than commonly used concatenation approaches. Our GBS data and analyses have improved the resolution of relationships in the *T. basedowii* complex and provided new insights into processes influencing evolution of the group (e.g. polyploidy and partial introgression). These results will support an upcoming taxonomic revision of the complex, with the recognition of new species including one of conservation significance, and complement the findings of Anderson *et al.* (2016; Chapter 2) with regard to the high diversity of the complex in the Pilbara region of Western Australia. We encourage other researchers working in similarly difficult taxonomic

systems to use the relatively affordable and accessible GBS approach for generating and analysing genomic data.

Acknowledgements

We thank S. van Leeuwen, N. Gibson, M. Langley, K. Brown, P. Jobson, and R. Davis for assistance with fieldwork, and E. Mayence for collecting a population of *T. plurinervata* from the Shark Bay area. We thank K. Ottewell, T. Hopley and R. Binks for allowing us to join them in a project using the Pawsey Supercomputing Centre facilities, and we thank the staff at the Pawsey Supercomputing Centre for help using the system and script troubleshooting. We thank D. Eaton for helpful responses to our questions about PyRAD, and A. Chuah at the ANU Biomolecular Resource Facility for advice on processing our data and for helpful scripts.

Supplementary Information

Supplementary information can be found in Appendix B, along with some of the scripts written for data handling and analysis.

References

Alcaide M, Scordato ESC, Price TD, Irwin DE (2014) Genomic divergence in a ring species complex. *Nature*, **511**, 83–85.

Anderson BM, Barrett MD, Krauss SL, Thiele K (2016) Untangling a species complex of arid zone grasses (*Triodia*) reveals patterns congruent with co-occurring animals. *Molecular Phylogenetics and Evolution*, **101**, 142–162.

Armstrong G (2008) *Triodia caelestialis* (Triodieae: Chloridoideae: Poaceae), a new species from the central Kimberley, Western Australia. *Journal of the Royal Society of Western Australia*, **91**, 313–317.

Arnold B, Corbett-Detig RB, Hartl D, Bomblies K (2013) RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, **22**, 3179–3190.

Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLOS ONE*, **3**, e3376.

Baldassarre DT, White TA, Karubian J, Webster MS (2014) Genomic and morphological analysis of a semipermeable avian hybrid zone suggests asymmetrical introgression of a sexual signal. *Evolution*, **68**, 2644–2657.

Barker B (2005) Standardising informal names in Australian publications. *Australian*

Systematic Botany Society Newsletter, **122**, 11–12.

Barrett RL, Barrett MD (2011) Two new species of *Triodia* (Poaceae: Triodieae) from the Kimberley region of Western Australia. *Telopea*, **13**, 57–67.

Barrett RL, Barrett MD (2015) Twenty-seven new species of vascular plants from Western Australia. *Nuytsia*, **26**, 21–87.

Barrett RL, Wells GB, Dixon KW (2005) New taxa and combinations: Subfam. Chloridoideae, Trib. Triodieae, *Triodia*. In: *Flora of Australia* (ed Mallett K), pp. 458–459. CSIRO Publishing, Melbourne.

Beadle NCW (1981) *The Vegetation of Australia*. Cambridge University Press, Cambridge.

Bouckaert RR (2010) DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics*, **26**, 1372–1373.

Bradbury PJ, Zhang Z, Kroon DE *et al.* (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, **23**, 2633–2635.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A (2012) Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, **29**, 1917–1932.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci *de novo* from short-read sequences. *G3: Genes, Genomes, Genetics*, **1**, 171–182.

Chifman J, Kubatko L (2014) Quartet inference from SNP data under the coalescent model. *Bioinformatics*, **30**, 3317–3324.

Crisp MD, Mant J, Toon A, Cook LG (2015) Australian spinifex grasses: new names in *Triodia* for *Monodia* and *Symplectrodia*. *Phytotaxa*, **230**, 293–296.

Daly BG, Dickman CR, Crowther MS (2008) Causes of habitat divergence in two species of agamid lizards in arid central Australia. *Ecology*, **89**, 65–76.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772.

Davey JW, Hohenlohe PA, Etter PD *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews. Genetics*, **12**, 499–510.

Dolezel J, Greilhuber J, Suda J (2007) Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols*, **2**, 2233–2244.

Doyle JJ, Dickson EE (1987) Preservation of plant samples for DNA restriction endonuclease analysis. *Taxon*, **36**, 715–722.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

Eaton DAR (2014) PyRAD: assembly of *de novo* RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1844–1849.

Eaton DAR, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). *Systematic Biology*, **62**, 689–706.

Edwards SV, Liu L, Pearl DK (2007) High-resolution species trees without concatenation. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 5936–5941.

Edwards SV, Xi Z, Janke A *et al.* (2016) Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. *Molecular Phylogenetics and Evolution*, **94**, 447–462.

Elshire RJ, Glaubitz JC, Sun Q *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLOS ONE*, **6**, e19379.

Escudero M, Eaton DAR, Hahn M, Hipp AL (2014) Genotyping-by-sequencing as a tool to infer phylogeny and ancestral hybridization: A case study in *Carex* (Cyperaceae). *Molecular Phylogenetics and Evolution*, **79**, 359–367.

Fu Y-B (2014) Genetic diversity analysis of highly incomplete SNP genotype data with imputations: an empirical assessment. *G3: Genes, Genomes, Genetics*, **4**, 891–900.

Giarla TC, Esselstyn JA (2015) The challenges of resolving a rapid, recent radiation: empirical and simulated phylogenomics of Philippine shrews. *Systematic Biology*, **64**, 727–740.

Gohli J, Leder EH, Garcia-Del-Rey E *et al.* (2015) The evolutionary history of Afrocanarian blue tits inferred from genomewide SNPs. *Molecular Ecology*, **24**, 180–191.

Goloboff PA, Farris JS, Nixon KC (2008) TNT, a free program for phylogenetic analysis. *Cladistics*, **24**, 774–786.

Grabowski PP, Morris GP, Casler MD, Borevitz JO (2014) Population genomic variation reveals roles of history, adaptation and ploidy in switchgrass. *Molecular Ecology*, **23**, 4059–4073.

Greilhuber J, Doležel J, Lysák MA, Bennett MD (2005) The origin, evolution and proposed stabilization of the terms “genome size” and “C-value” to describe nuclear DNA contents. *Annals of Botany*, **95**, 255–260.

Griffin GF (1990) Characteristics of three spinifex alliances in Central Australia. *Journal of Vegetation Science*, **1**, 435–444.

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696–704.

Hamlin JAP, Arnold ML (2014) Determining population structure and hybridization for two iris species. *Ecology and Evolution*, **4**, 743–755.

Huang H, Knowles LL (2014) Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. *Systematic Biology*.

Hurry CR, Walsh NG, Murphy DJ (2012) A taxonomic review of *Triodia bunicola*

and *T. scariosa* (Poaceae: Chloridoideae), based on morphological and molecular data. *Australian Systematic Botany*, **25**, 304–312.

Ilut DC, Nydam ML, Hare MP (2014) Defining loci in restriction-based reduced representation genomic data from nonmodel species: Sources of bias and diagnostics for optimal clustering. *BioMed Research International*, **2014**, 1–9.

Jarvis A, Reuter H, Nelson A, Guevera E (2008) *Hole-filled seamless SRTM data V4*. International Centre for Tropical Agriculture (CIAT). http://srtm.csi.cgiar.org.

Kitchener DJ, Stoddart J, Henry J (1983) A taxonomic appraisal of the genus *Ningaui* Archer (Marsupialia: Dasyuridae), including description of a new species. *Australian Journal of Zoology*, **31**, 361–379.

Kubatko LS, Degnan JH (2007) Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology*, **56**, 17–24.

Lazarides M (1997) A revision of *Triodia* including *Plectrachne* (Poaceae, Eragrostideae, Triodiinae). *Australian Systematic Botany*, **10**, 381–489.

Leaché AD, Banbury BL, Felsenstein J, Nieto-Montes de Oca A, Stamatakis A (2015a) Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Systematic Biology*, **64**, 1032–1047.

Leaché AD, Chavez AS, Jones LN *et al.* (2015b) Phylogenomics of phrynosomatid lizards: conflicting signals from sequence capture versus restriction site associated DNA sequencing. *Genome Biology and Evolution*, **7**, 706–719.

Liu L, Yu L, Edwards SV (2010) A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evolutionary Biology*, **10**, 302.

Lu F, Lipka AE, Glaubitz J *et al.* (2013) Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLOS Genetics*, **9**, e1003215.

Mastretta-Yanes A, Arrigo N, Alvarez N *et al.* (2015) Restriction site-associated DNA sequencing, genotyping error estimation and *de novo* assembly optimization for population genetic inference. *Molecular Ecology Resources*, **15**, 28–41.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**, 526–538.

Mirarab S, Warnow T (2015) ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*, **31**, i44–i52.

Morton S, James S (1988) The diversity and abundance of lizards in arid Australia: a new hypothesis. *American Naturalist*, **132**, 237–256.

Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.

Pyron RA, Hendry CR, Chou VM *et al.* (2014) Effectiveness of phylogenomic data and coalescent species-tree methods for resolving difficult nodes in the phylogeny of advanced snakes (Serpentes: Caenophidia). *Molecular Phylogenetics and Evolution*, **81**, 221–231.

Python Software Foundation (2016) Python Language Reference, version 2.7. https://docs.python.org/2.7/.

R Development Core Team (2015) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rambaut A (2014) FigTree. http://tree.bio.ed.ac.uk/software/figtree/.

Rambaut A, Suchard M, Xie D, Drummond A (2014) Tracer. http://beast.bio.ed.ac.uk/Tracer.

Rivers DM, Darwell CT, Althoff DM (2016) Phylogenetic analysis of RAD-seq data: examining the influence of gene genealogy conflict on analysis of concatenated data. *Cladistics*. DOI: 10.1111/cla.12149.

Rubin BER, Ree RH, Moreau CS (2012) Inferring phylogenies from RAD sequence data. *PLOS ONE*, **7**, e33394.

Sedeek KEM, Scopece G, Staedler YM *et al.* (2014) Genic rather than genome-wide differences between sexually deceptive *Ophrys* orchids with different pollinators. *Molecular Ecology*, **23**, 6192–6205.

Seo T-K (2008) Calculating bootstrap probabilities of phylogeny using multilocus sequence data. *Molecular Biology and Evolution*, **25**, 960–971.

Simmons MP (2014) Limitations of locally sampled characters in phylogenetic analyses of sparse supermatrices. *Molecular Phylogenetics and Evolution*, **74**, 1–14.

Springer MS, Gatesy J (2016) The gene tree delusion. *Molecular Phylogenetics and Evolution*, **94**, 1–33.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology*, **57**, 758–771.

Streicher JW, Schulte JA, Wiens JJ (2015) How should genes and taxa be sampled for phylogenomic analyses with missing data? An empirical study in iguanian lizards. *Systematic Biology*, **65**, 128–145.

Swofford DL (2002) *PAUP*: phylogenetic analysis using parsimony (*and other methods)*. Sinauer Associates, Sunderland, Massachusetts.

Vachaspati P, Warnow T (2015) ASTRID: Accurate Species TRees from Internode Distances. *BMC Genomics*, **16**, S3.

Waters C, Murray B, Melville G *et al.* (2010) Polyploidy and possible implications for the evolutionary history of some Australian Danthonieae. *Australian Journal of Botany*, **58**, 23–34.

White TA, Perkins SE, Heckel G, Searle JB (2013) Adaptive evolution during an

ongoing range expansion: the invasive bank vole (*Myodes glareolus*) in Ireland. *Molecular Ecology*, **22**, 2971–2985.

Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, **30**, 614–620.

Chapter 4.

Diversification timing for Australian arid zone grasses (*Triodia*) and evidence for recent range expansion across sandy deserts

Abstract

Aim The Australian arid zone has experienced intensifying aridification and landscape changes from the Pliocene, but relatively little is known about the diversification and response of the flora in relation to these climate and landscape changes. We use a group of closely-related arid zone grasses, the *Triodia basedowii* E.Pritz. species complex, to investigate 1) the relative timing of divergences in the group in relation to major climate changes in the arid zone and 2) the evidence for, and origin of, a recent range expansion in the widespread *T. basedowii*, which occupies sandy dunefields thought to have begun to form within the last *c*. 1 Ma.

Location Arid regions in western and central Australia.

Methods We aligned newly-sequenced and existing chloroplast genomes from across Poaceae, and used these to construct a fossil-calibrated phylogeny using BEAST. Denser sampling of Chloridoideae and *Triodia* was then used to construct a second time-calibrated phylogeny to estimate divergence times for the complex. Genomic SNPs previously generated using genotyping by sequencing were used to test the signal for and geographic origin of a range expansion.

Results The *T. basedowii* complex began to diversify 1.3–2.8 Ma, prior to the formation of sandy dunefields. Various sampling schemes for the range expansion analysis consistently recovered a signal of expansion for *T. basedowii* with an origin in central or eastern Australia, not the Pilbara, where the bulk of lineage richness in the complex is found.

Main conclusions Lineages in the *T. basedowii* complex likely diversified during the Pleistocene and persisted through the mid-Pleistocene shift in global glacial cycle amplitude and the concurrent formation of sandy dunefields in Australia. The persistence of these lineages suggests an evolutionary resilience in Australian arid zone plants to climatic and landscape changes.

Introduction

The Australian flora comprises lineages that have been present since the continent became isolated at the end of the Eocene (Veevers *et al.* 1991; McLoughlin 2001) *c*. 34 Ma (Walker *et al.* 2012), as well as more recent immigrants that have arrived by long-distance dispersal (Crisp & Cook 2013). In the Australian arid zone, which has formed over the past *c*. 15 Ma (Byrne *et al.* 2008) and covers more than 60% of the continent, there are examples of both older lineages and newer arrivals. Older lineages that have invaded or persisted in the developing arid zone include Goodeniaceae (Jabaily *et al.* 2014), *Acacia* (Ariati *et al.* 2006), and *Callitris* (Sakaguchi *et al.* 2013). More recent immigrants that have arrived and radiated in the arid zone include *Lepidium* (Mummenhoff *et al.* 2004), chenopods from Salicornioideae (Shepherd *et al.* 2004) and Camphorosmoideae (Kadereit & Freitag 2011), and *Triodia* (Toon *et al.* 2015). Both older lineages and newer arrivals have been subjected to intensifying aridification in the arid zone, including the extreme climatic variability of Pleistocene glacial cycles and the formation of sandy dunefields (Fujioka *et al.* 2009). A better understanding of the timing of diversification in these groups can provide insights into how specific climatic changes may have influenced plant evolution, such as through increased extinction or speciation.

The perennial grass genus *Triodia* R.Br. (Poaceae: Chloridoideae) is an ideal study system for exploring evolution in arid Australia. *Triodia* species are iconic Australian plants and dominant components of hummock grasslands, which cover more than 18% of the continent (Department of the Environment 2006). These grasses show a wide ecological adaptability and are distributed across some of the

driest parts of the continent, extending into tropical savannahs in northern Australia and semi-arid temperate regions in the south (Lazarides 1997). In a recent divergence dating study of a broad sampling of *Triodia*, Toon *et al.* (2015) used two molecular markers (ITS and *matK*) to provide a temporal framework for investigating the evolution of leaf traits, focusing on the ecological transition from the arid zone to the northern savannahs. Here, we provide more focused sampling of a group of arid zone *Triodia* species, the *T. basedowii* E.Pritz. species complex, using full chloroplast genomes to estimate the timing of their diversification.

The *T. basedowii* species complex includes two named taxa (*T. basedowii* and *T. lanigera* Domin), five informally recognised taxa (*T.* sp. Shovelanna Hill (S. van Leeuwen 3835), *T.* sp. Little Sandy Desert (S. van Leeuwen 4935), *T.* sp. Peedamulla (A.A. Mitchell PRP 1636), *T.* sp. Warrawagine (A.L. Payne PRP 1859) and *T.* sp. Pannawonica (B.M. Anderson & M.D. Barrett BMA 89)) and two as yet unnamed entities (*T.* "shova" and *T.* "nana"). Our previous work on the complex revealed higher lineage richness in the Pilbara region of Western Australia and lower richness across the inland sandy deserts (Anderson *et al.* 2016; Chapter 2). We proposed two potential explanations for the pattern of lineage richness: 1) that the Pilbara allowed lineages to persist while lineages outside the Pilbara went extinct in adverse climatic conditions, and/or 2) that the Pilbara has been a centre of diversification while lineages outside the Pilbara have not diversified (Anderson *et al.* 2016; Chapter 2). Evaluating these and other hypotheses about the history of the complex requires an estimate of divergence times as well as improved resolution of phylogenetic relationships.

Of the taxa in the *T. basedowii* complex, only *T. basedowii* is widespread outside of Western Australia, inhabiting sandy desert environments including most of the Australian dunefields. The sandy dunefields in central Australia have been dated to *c*. 1 Ma (Fujioka *et al.* 2009), corresponding to a time of global climatic transition as glacial cycles shifted frequency and amplitude (Clark *et al.* 1999). Dune formation and activity may have continued during subsequent dry Pleistocene glacial cycles (Hesse *et al.* 2004; Hesse 2010). Given the current habitat of *T. basedowii*, the formation of the dunefields may have provided new or expanded existing habitat

onto which *T. basedowii* spread. There is molecular evidence for a signal of expansion in *T. basedowii* (Anderson *et al.* 2016; Chapter 2), consistent with the hypothesis that the dunefields provided new habitat, but the timing or direction of that expansion is unknown. Given the high lineage richness of the complex in the Pilbara, populations of *T. basedowii* there might represent a source for the demographic expansion.

This study has two main aims: 1) to provide a temporal framework for diversification in the *T. basedowii* complex to compare with the timing of climate and landscape changes in Australia, and 2) to determine if, and from where, *T. basedowii* has recently expanded its range. We address the first aim through fossil-calibrated molecular dating of alignments from coding regions of full chloroplast genomes of representative lineages from across Poaceae to provide an estimate of the age of crown Chloridoideae (the grass subfamily to which *Triodia* belongs; (Peterson *et al.* 2010)). We then use this estimate to calibrate a second dating analysis with more detailed sampling of Chloridoideae, *Triodia* and the *T. basedowii* complex, and a broader sampling of the chloroplast genome. We address our second aim by using previously sequenced genomic SNPs from populations of *T. basedowii* across its range in a method designed to detect both a signal for expansion and the most likely geographic location of the origin for that expansion (Peter & Slatkin 2013, 2015).

Materials and Methods

Sampling

For chloroplast divergence dating across Poaceae, we newly sequenced and assembled (13) or downloaded from GenBank (26) genomes from members of all subfamilies of grasses (see Table S4.1). Depending on available genomes, we aimed for two to three members of each subfamily (four for Oryzoideae), with greater sampling of Chloridoideae (14). Sample classification and subfamililial taxonomy follows Soreng *et al.* (2015). We were able to obtain only one representative for each of Anomochloideae, Pharoideae and Puelioideae. For a second analysis with denser sampling of Chloridoideae and *Triodia*, we included an additional ten *Triodia* and five other Chloridoideae genomes newly sequenced and assembled.

For the test of range expansion in *Triodia basedowii*, we used genotyping by sequencing (GBS) data from a previous study (Chapter 3). We included 36 samples of *Triodia basedowii* from 17 populations across its range (see Table S4.2), along with three samples of *T.* sp. Little Sandy Desert, four of *T.* "nana" and five of *T.* sp. Peedamulla to be used as outgroups for identifying ancestral and derived SNPs.

Chloroplast genome sequencing and assembly

Silica-dried leaf material was ground in liquid nitrogen and genomic DNA extracted using a CTAB method (Doyle & Dickson 1987) as described in Anderson *et al.* (2016; Chapter 2) but with RNAse added prior to heating. DNA was also extracted from herbarium samples using the commercial DNeasy Plant Mini Kit (Qiagen, USA) following manufacturer's instructions, with DNA eluted in 100 µl of AE buffer. DNA concentration was quantified on a NanoDrop ND-1000 spectrophotometer and samples with concentrations of 3–20 ng/µl sent to the Australian Genome Research Facility node in Melbourne, Victoria. 200 ng of DNA was sheared in a volume of 50 µl using a Coavris E220 Focused ultrasonicator. Following shearing, sequencing libraries were prepared using Illumina's TruSeq Nano DNA Library preparation kit (350 bp median insert) following the manufacturer's protocol. Libraries were assessed by gel electrophoresis (Agilent D1000 ScreenTape Assay) and quantified by qPCR (KAPA Library Quantification Kits for Illumina). Sequencing was performed on the Illumina HiSeq 2500 system with 2×125 bp paired-end reads using the HiSeq PE Cluster Kit v. 5 and HiSeq SBS Kit v. 4 (250 cycles).

For 19 genomes, adapter sequences ("AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC" and "AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT") in raw paired-end reads were removed using the software cutadapt v. 1.9.1 (Martin 2011). Sequencing read errors were corrected with SPAdes v. 3.6.1 (Bankevich *et al.* 2012). The filtered reads were normalised for sequencing depth based on k-mer counts and merged into single reads using BBNorm and BBMerge, respectively, from the BBMap package v. 35.82 (https://sourceforge.net/projects/bbmap/ ; visited

June 2016). The processed reads were assembled into contigs using Velvet v. 1.2.09 (Zerbino & Birney 2008) with k-mer values 51, 71, 91 and 111, and coverage cut-off values 7, 10, 15 and 20. Assembled contigs were aligned to the chloroplast genome of *Spinacia oleracea* (GenBank accession NC_002202) using MUMmer v. 3.1 (Kurtz *et al.* 2004). Based on the alignments, assembled chloroplast contigs were identified, ordered and then merged into a single circular draft genome for each specimen using a custom script. Reads were mapped back to the assembly using BWA v. 0.7.5a-r405 (Li & Durbin 2009) and the assembly further refined with Pilon v. 1.16 (Walker *et al.* 2014). The *Setaria viridis* genome (GenBank accession NC_028075) was used as a reference for gene annotation.

One chloroplast genome (*T.* sp. Shovelanna Hill) was sequenced using 454 shotgun sequencing by M. Gardner, using the method outlined in Gardner *et al.* (2011), and assembled by M. Duvall, S. Burke and W. Wysocki, (Northern Illinois University) using the NGS plastome assembly and verification methods outlined in Wysocki *et al.* (2014) and Duvall *et al.* (2016).

The remaining 8 genomes were assembled using a beta-test version of the de novo assembler NOVOplasty (https://github.com/ndierckx/NOVOPlasty), using various conserved seed sequences where necessary. Genomes were validated by mapping raw reads onto the assembly in Geneious v. 6.1.8 (Kearse *et al.* 2012) to check for ambiguously aligned regions. The NOVOplasty method was applied to three of the genomes assembled using the first method above, to demonstrate its utility in *Triodia*. The three genomes assembled using the two methods differed in, at most, a few heterozygous/homozygous calls and the number of units in repeat regions.

Phylogenetic analysis

We created two data sets for phylogenetic and dating analyses: A) 39 samples from across Poaceae with focused sampling in Chloridoideae, and B) 29 samples of Chloridoideae with focused sampling in *Triodia* plus two samples of Danthonioideae as an outgroup. Data set A comprised only chloroplast coding regions, while data set B included coding regions as well as introns and intergenic spacers. Whole chloroplast genomes were aligned in Geneious v. 6.1.8 (Kearse *et al.* 2012) using the

MAFFT (Katoh & Standley 2013) plugin with gap creation 2.0 and gap extension 0.12. Target regions were manually extracted from the alignment and checked by eye for alignment ambiguities, which were removed or adjusted. A custom Python v. 2.7.2 (Python Software Foundation 2016) script was used to remove positions in the alignment with more than one missing or 'N' state.

To determine the optimum partitioning and model scheme, the loci for each data set were run through PartitionFinder v. 2.0.0-pre13 (Lanfear *et al.* 2012), using models available for MrBayes, the Bayesian information criterion to select the optimal model, and a greedy search. The optimal partitioning scheme was used in phylogenetic inference with RAxML v. 8.1.21 (Stamatakis 2014), which implements the GTRGAMMA model for all partitions. RAxML was run with a 100 replicate rapid bootstrap (Stamatakis *et al.* 2008) followed by a search for the maximum likelihood tree ('-f a' option). We also ran the partitions and best models in MrBayes v. 3.2.6 (Ronquist *et al.* 2012). MrBayes was executed for three runs, each using four chains and two million generations, sampling every 500, and 10% of samples discarded as burn-in. Trees from each run were combined into a single file with a custom script, and a maximum clade credibility tree constructed using TreeAnnotator v. 2.4.0 (Bouckaert *et al.* 2014) and visualized in FigTree v. 1.4.3pre (Rambaut 2014).

Divergence dating

In order to link molecular divergences to time, we included five fossils (Table 4.1) in Bayesian analyses of data set A using BEAST v. 2.4.0 (Bouckaert *et al.* 2014). The placement of the Prasad *et al.* (2005, 2011) phytoliths (hereafter simply 'phytoliths') is somewhat controversial (see Christin *et al.* 2014), so they were included in two separate analyses in alternate positions: stem Oryzeae (common ancestor of *Oryza* and *Microlaena*), and stem BOP+PACMAD (Bambusoideae, Oryzoideae, Pooideae + Panicoideae, Aristidoideae, Chloridoideae, Micrairoideae, Arundinoideae, and Danthonioideae). In setting the prior probability distributions, maximum ages were arbitrarily chosen as 1.5 times the minimum age except for the stem of Chloridoideae, which was extended (making a less informative prior) to account for

Table 4.1. Fossils used to calibrate the molecular dating analysis of data set A and their parameterization in BEAST.

Number (Reference)	Tree placement	Age	log-norm (m, mean of ln)	log-norm (s, std dev of ln)
1 (Prasad *et al.* 2005, 2011)	(1) stem BOP+PACMAD (2) stem Oryzeae	65 Ma (max 97.5 Ma)	4.43846	0.085454
2 (Zucol *et al.* 2010)	stem Pooideae	40 Ma (max 60 Ma)	3.95295	0.085454
3 (MacGinitie 1953; Manchester 2001)	common ancestor of *Stipa* and *Poa*	34 Ma (max 51 Ma)	3.79043	0.085454
4 (Strömberg 2005)	crown Bambusoideae (Piperno & Pearsall 1998)	35 Ma (max 52.5 Ma)	3.81942	0.085454
5 (Strömberg 2005)	stem Chloridoideae	19 Ma (max 50 Ma)	3.57461	0.203923

the evidence of C_4 grasses from the Oligocene (Urban *et al.* 2010). Node priors for the placement of the fossils were implemented using a log-normal distribution set such that the minimum and maximum ages comprised the 0.1% and 95.1% quantiles of the probability distribution, in order to make the minimum ages more strict than standard 2.5% and 97.5% quantiles. Values for the mean and standard deviation of the natural logarithm for the log-normal distribution were set using ParameterSolver v. 3.0.0 (https://biostatistics.mdanderson.org/SoftwareDownload/ ; visited May 2016). The prior on the root of the tree was set as a uniform distribution from 0 to a hard maximum age of 125 Ma, under the assumption that lineages of Poaceae were not present prior to the earliest eudicot fossil (Doyle 1992). *Anomochloa* was fixed as the outgroup (see Bouchenak-Khelladi *et al.* 2008).

For data set A, duplicated to evaluate both placements of the phytoliths, BEAST was initially run with 20 million generations, sampling every 500, starting from a UPGMA tree and using the Yule process to model speciation under an uncorrelated relaxed log-normal clock (UCLN; Drummond *et al.* 2006). Rather than fixing models for the optimum partitions, we allowed the substitution models to be sampled as part of the analysis using the bModelTest v. 0.2.0 package (Bouckaert 2015). Log files were examined with Tracer v. 1.6.0 (Rambaut *et al.* 2014) and had low effective sample sizes (ESS) for some parameters and unstable traces. After three runs of 20 million generations and consistently low ESS for some parameters, we incorporated two additional runs of 60 million generations. Burn-in percentages were chosen based on evaluating traces for a number of parameters such that the retained samples were from traces at stationarity. After removing the variable burn-in proportions, the runs were combined with LogCombiner v. 2.4.0 (part of BEAST) and comprised *c.* 125 million generations, sampled every 500, resulting in ESS > 200 for most parameters, with a few parameters having 200 > ESS > 100. Trees were combined with LogCombiner using the same burn-in percentages used for the log files, and a maximum clade credibility tree constructed with TreeAnnotator and visualized in FigTree.

For data set B, the 95% highest posterior density (HPD) of the node age for crown Chloridoideae found in analysis of data set A was used as a secondary calibration.

For each of the two placements of phytoliths, the 95% HPDs were input into ParameterSolver for normal, gamma, and log-normal distributions, and the distribution that produced a mean closest to the node height was used to set the secondary calibration with 2.5% and 97.5% quantiles set to the 95% HPD values (Table 4.2). A recent molecular dating analysis of *Triodia* (Toon *et al.* 2015) found that a random local clocks model (RLC; Drummond & Suchard 2010) provided a better fit to account for rate changes in *Triodia* than the UCLN clock (see also Crisp *et al.* 2014), which produced much younger ages in *Triodia*. After encountering a bug in BEAST 2 that prevented us from implementing a RLC with our data set, we instead used BEAST v. 1.8.4pre (Drummond & Rambaut 2007) to run our analyses of data set B under an RLC. We set the models from PartitionFinder for each partition explicitly, and again ran BEAST with a UPGMA tree and a Yule process. For both secondary calibrations (from two phytolith placements), we initially ran BEAST for 60 million generations, sampling every 500. We examined log files in Tracer to ensure most parameters had ESS > 200, but found that a few had low ESS, including treeModel.rootHeight, which might be expected to be unstable given we had not restricted it with a prior. We ran a second run for *c*. 60 million generations and a third run for 120 million generations sampling every 1000 generations. Burn-in percentages were again chosen to target samples from traces at stationarity. The three runs for each calibration appeared to converge on slightly different values for the posterior and prior (more noticeable for the calibration with phytoliths at stem BOP+PACMAD), but maximum clade credibility trees from each run showed very similar node ages. We combined trees from the three runs with the burn-in percentages determined in Tracer for each calibration using LogCombiner and created a maximum clade credibility tree using TreeAnnotator and visualized in FigTree.

To assess the impact of clock model, calibration choice and locus effects, we ran analyses of 60 million generations for these variations. Because we were unable to implement a RLC in BEAST 2 to compare with the UCLN clock, or to implement a UCLN in BEAST 1.8.4pre to compare with the RLC using path sampling (e.g. Baele *et al.* 2012), we ran a UCLN analysis in BEAST 2 to illustrate that, as in the analysis in Toon *et al.* (2015), the UCLN model produces younger ages in *Triodia* compared

Table 4.2. Secondary calibrations used for analysis of data set B. Placement of phytoliths in analysis of data set A produced different 95% HPDs on the crown age of Chloridoideae. HPD = highest posterior density interval.

Source of calibration	Age (Ma)	95% HPD	Distribution (mean closest to age)	parameter 1 (gamma: shape; normal: mean)	parameter 2 (gamma: scale; normal: standard dev)
phytoliths at stem BOP-PACMAD	crown Chloridoideae: 34.86	26.21, 44.04	gamma	57.544388	0.60052641
phytoliths at stem Oryzeae	crown Chloridoideae: 38.2	28.78, 47.78	normal	38.28	4.8470278
(Toon *et al.* 2015)	crown *Triodia*: 14.7	11.4, 18.3	gamma	69.086871	0.21204279

to the RLC model. Given that we are looking at divergences within a genus, it is likely there is some correlation of rates of evolution between species, which makes a RLC model more likely to be accurate than a UCLN model (see Crisp *et al.* 2014). Since our recovered node age for crown *Triodia* differed from that of Toon *et al.* (2015), we assessed how using their dates would affect the crown age of the *T. basedowii* complex by running analyses under the RLC model using their recovered age for crown *Triodia* as a secondary calibration, both with and without the crown Chloridoideae secondary calibration. As Toon *et al.* (2015) used the ITS region and the chloroplast gene *matK*, we also ran two runs using only the *matK* gene rather than the full chloroplast alignment to determine if the length of the alignment might be partly causing the differences.

Range expansion

Peter and Slatkin (2015) produced a program that takes biallelic SNPs from a set of populations of known location, uses outgroup samples to polarize the SNPs, and computes the likely origin for the expansion. If a source population extends its range through a series of founder events, it is expected that populations further from the origin of the expansion will have experienced more genetic drift (Peter & Slatkin 2013). This greater drift will produce clines in the frequencies of neutral alleles (as alleles are lost with each subsequent founder event), leading to populations further away from the origin having higher frequencies of derived alleles (Peter & Slatkin 2013). Peter and Slatkin (2013, 2015) developed a measure to account for the difference in allele frequencies between two populations, and after observing that it tended to increase linearly with distance from an origin, used this measure in a time-difference of arrival method (Gustafsson & Gunnarsson 2003) to identify the most likely location of the origin for an expansion. Dr. Peter kindly provided the scripts for running their program, along with scripts for producing graphical output, which we adjusted to fit our data and geographic area of interest. The scripts use the R packages 'geosphere' v. 1.5-1 (Hijmans 2015), 'sp' v. 1.2-2 (Pebesma & Bivand 2005), 'rworldmap' v. 1.3-6 (South 2011), 'maps' v. 3.1.0 (Becker *et al.* 2016), and 'mapproj' v. 1.2-4 (McIlroy *et al.* 2015) in R v. 3.2.5 (R Development Core Team 2015).

We used SNPs from a previous study (Chapter 3) that used PyRAD (Eaton 2014) to assemble loci from GBS data. We generated SNP output at the optimal clustering thresholds for loci assembled or unassembled by PEAR (Zhang *et al.* 2014) (see Chapter 3 for details) using only the target *T. basedowii* samples and alternatively samples of *T.* sp. Little Sandy Desert, *T.* "nana" or *T.* sp. Peedamulla as outgroups, keeping SNPs present in at least four samples. We used custom Python scripts to select a single SNP per locus either randomly or with a bias toward biallelic SNPs that had multiple copies of the rare allele. We then used a custom R script to filter the resulting SNPs to keep only biallelic SNPs and only those present in at least one outgroup sample, and to generate the input files for use with the range expansion scripts.

We ran the expansion program for various subsets of the data and with each outgroup to assess the consistency of the signal for expansion. Potential challenges for this analysis include the large amount of missing data (Fu 2014) and low sequencing coverage (Davey *et al.* 2011) of GBS data, which makes detecting heterozygous alleles in a population problematic, as well as polyploidy and unequal population sizes. We analysed the assembled SNPs separately from the unassembled SNPs, since the former had roughly twice the read depth and would therefore be more likely to pick up heterozygous alleles. We also analysed some subsets of our samples excluding tetraploids to avoid potentially higher heterozygosity in polyploids. Finally, as our populations were of different sizes, we chose a random sample from each population as an input subset, as well as including all samples as input and letting the range expansion scripts downsample larger populations.

The range expansion analysis assumes a single point of origin, which likely oversimplifies real demographic history. In the case of multiple origins, Peter and Slatkin (2013) recommend determining which samples are likely to have come from each origin and then applying their method to each group of samples separately. Samples can be divided into groups using population genetic structure or clustering. To assess genetic structure in *T. basedowii*, we used 8663 randomly-selected assembled biallelic SNPs in our *T. basedowii* samples generated with *T.* sp. Little Sandy Desert as the outgroup, in the Bayesian clustering program Structure v. 2.3.4

(Pritchard *et al.* 2000). Structure was run five times for each value of K from 1 to 5, each run with 100,000 generations and a 10,000 generation burn-in. The optimum K was chosen using the method of Evanno *et al.* (2005). We also used principal components analysis (PCoA; built-in cmdscale function) based on Euclidean distance (built-in dist function) in R to visualise population structure with the same set of SNPs as were run in Structure. These approaches detected three subsets of the samples (designated "west", "central" and "interm"), for which we separately ran the range expansion analysis in addition to running all samples together ("ALL").

The accuracy of Peter and Slatkin's range expansion origin detection is reduced if the origin is near or beyond the edge of the sampled area (Peter & Slatkin 2013). We expanded the geographic area for detecting the origin when we found that the origin was often recovered at the edge of the area; the detection of an origin outside the area bounded by the samples should be interpreted cautiously.

Results

Chloroplast genomes

The 28 newly-assembled grass chloroplast genomes were on average *c*. 135,000 bp (132,643–137,308 bp) in length. There were no major discrepancies in gene content or order.

Phylogenetic analysis

After filtering the aligned genomic regions, data set A comprised 79 loci with a combined length of 62,005 bp, while data set B comprised 163 loci with a combined length of 101,510 bp. Optimal partitioning and models for both data sets are included in Table S4.3.

Phylogenetic inference for both data sets resulted in highly resolved trees (see supplementary Figs. S4.1–S4.4), with high (100% / 1.00 posterior probability) support for all grass subfamilies except Pharoideae and Anomochloideae, and high support for *Triodia* as a genus and for the *Triodia basedowii* species complex. Topologies were identical between RAxML and MrBayes trees.

Divergence dating

The chronogram for data set A with the phytoliths placed conservatively at stem BOP+PACMAD is shown in Fig. 4.1. Divergence dates for important nodes across the grasses (Table 4.3) are similar to some previous analyses (Prasad *et al.* 2011; Christin *et al.* 2014), and estimates for crown Chloridoideae (in our sampling, Chloridoideae did not include the divergent genus *Centropodia*, so equivalent nodes were taken from previous studies) are fairly similar for both placements of the phytoliths. With the phytoliths placed at stem Oryzeae (= crown Oryzoideae), our node age for crown Chloridoideae (38.2 Ma) falls between equivalent estimates from the two previous studies. The greatest effect of the phytolith placement is on node ages within Oryzoideae, with little impact on other subfamilies (e.g. Pooideae).

The chronogram for data set B using the secondary calibration from the analysis of data set A that placed the phytoliths at stem BOP+PACMAD is shown in Fig. 4.2. Table 4.4 shows variation in node ages based on different secondary calibrations, using RLC vs. UCLN, using the dates of Toon *et al.* (2015), and using only *matK*. Across both secondary calibrations, our analyses indicate a crown age for *Triodia* in the late Miocene (range 6.4–7.1 Ma; 4.6–9.0 Ma 95% HPDs) and a crown age for the *T. basedowii* complex in the early Pleistocene (range 1.9–2.1 Ma; 1.3–2.8 Ma 95% HPDs).

Range expansion

We detected spatial genetic structure within *T. basedowii*, with western samples ("west") grouping together in the PCoA (Fig. 4.3) and central samples ("central") grouping separately, along with a group of intermediately distributed samples ("interm"). The Structure runs for K >2 sometimes produced outlying estimated ln(likelihood) values that we were hesitant to use. With the outlier runs removed, optimum K=4; with them retained, optimum K=2. Across runs for K=4, Structure typically indicated a central Australian group of samples ("central") and a group of central Western Australian samples ("interm") sometimes with admixture, but inconsistently treated the "west" samples as two groups (results not shown). For K=2 (one run shown in Fig. 4.4, the other four essentially the same), Structure recognized

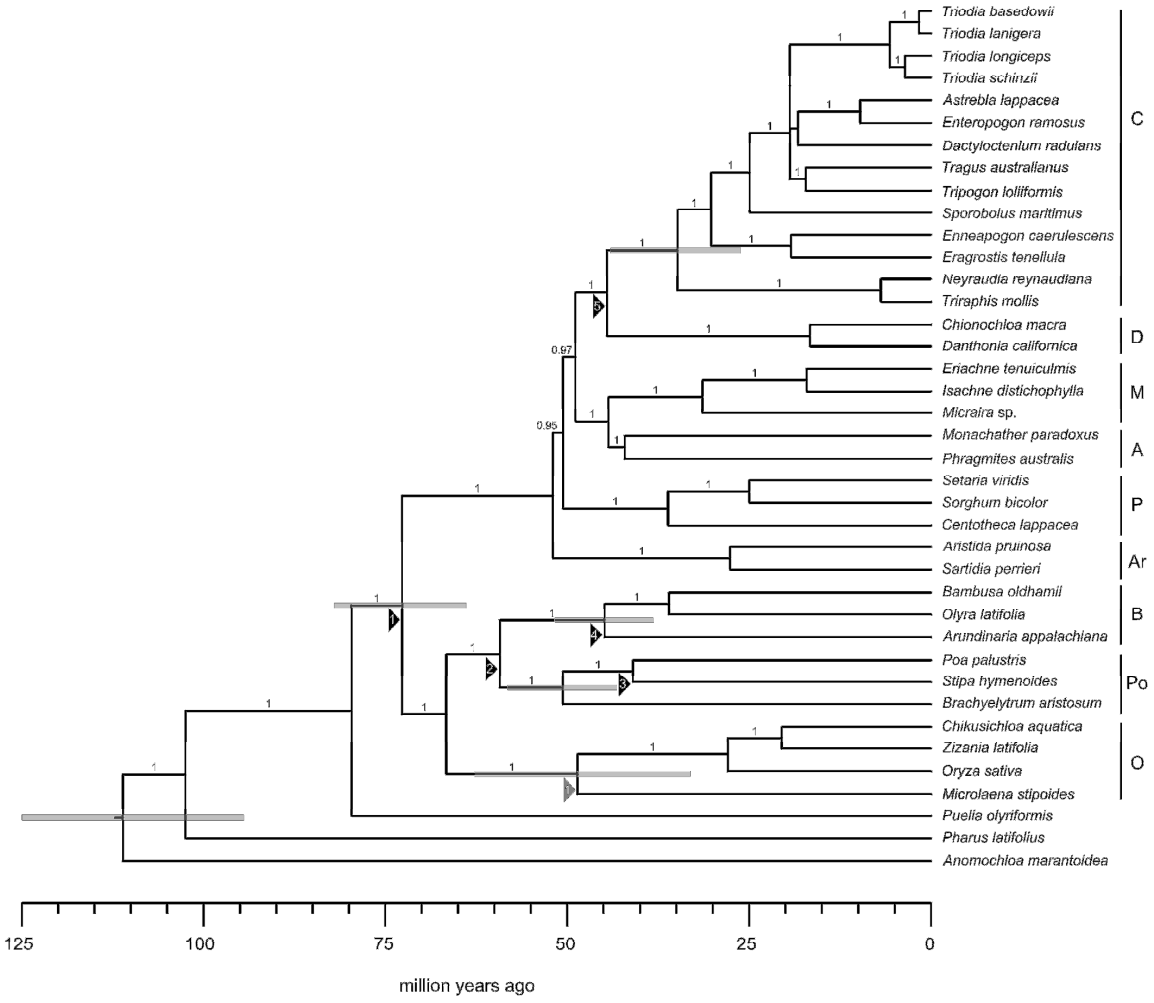

Figure 4.1. Chronogram from the BEAST analysis of data set A with phytoliths placed at stem BOP+PACMAD. Fossil calibration points are shown with black triangles (numbers correspond to Table 4.1), with the grey triangle showing the alternate placement for the phytoliths. Node bars for nodes of interest are 95% HPD intervals. Branch support values are posterior probabilities and are shown for values > 0.95. Grass subfamilies in BOP+PACMAD are indicated at right; O: Oryzoideae, Po: Pooideae, B: Bambusoideae, Ar: Aristidoideae, P: Panicoideae, A: Arundinoideae, M: Micrairoideae, D: Danthonioideae, C: Chloridoideae.

Table 4.3. Node ages (Ma) for data set A compared to previous molecular dating of the grasses. Node ages from Prasad *et al.* (2011) are with placement of the phytoliths at stem Oryzeae (their H1), while those from Christin *et al.* (2014) are based on their BEAST analysis of chloroplast data from across angiosperms, which included the placement of the phytoliths at stem Oryzeae. HPD = highest posterior density interval.

Study	**This study, phytoliths at stem BOP+PACMAD**		**This study, phytoliths at stem Oryzeae**		**Prasad et al. (2011)**		**Christin et al. (2014)**	
Node	**Age**	**95% HPD**	**Age**	**95% HPD**	**Age**	**95% HPD**	**Age**	**95% HPD**
crown Poaceae	111.1	94.46, 125	113.72	98.04, 125	121.31	95.86, 148.87	88.46	80.94, 97.78
crown BOP+PACMAD	72.72	63.92, 82.08	78.89	68.05, 89.98	81.64	69.55, 93.83	74.48	70.26, 80
crown Oryzoideae	48.6	33.11, 62.72	67.88	58.72, 77.29	67.14	56.85, 76.98	67.99	67, 70.764
crown Bambusoideae	44.9	38.16, 51.71	44.98	38.06, 51.93	47.4	36.46, 59.72	34.21	19.83, 56.22
crown Pooideae	50.59	43.26, 58.23	51.26	42.64, 59.97	57.8	48.23, 67.59	59.86	51.35, 68.45
crown Chloridoideae	34.86	26.21, 44.04	38.2	28.78, 47.78	33.64	24.54, 42.54	41.21	33.21, 48.96

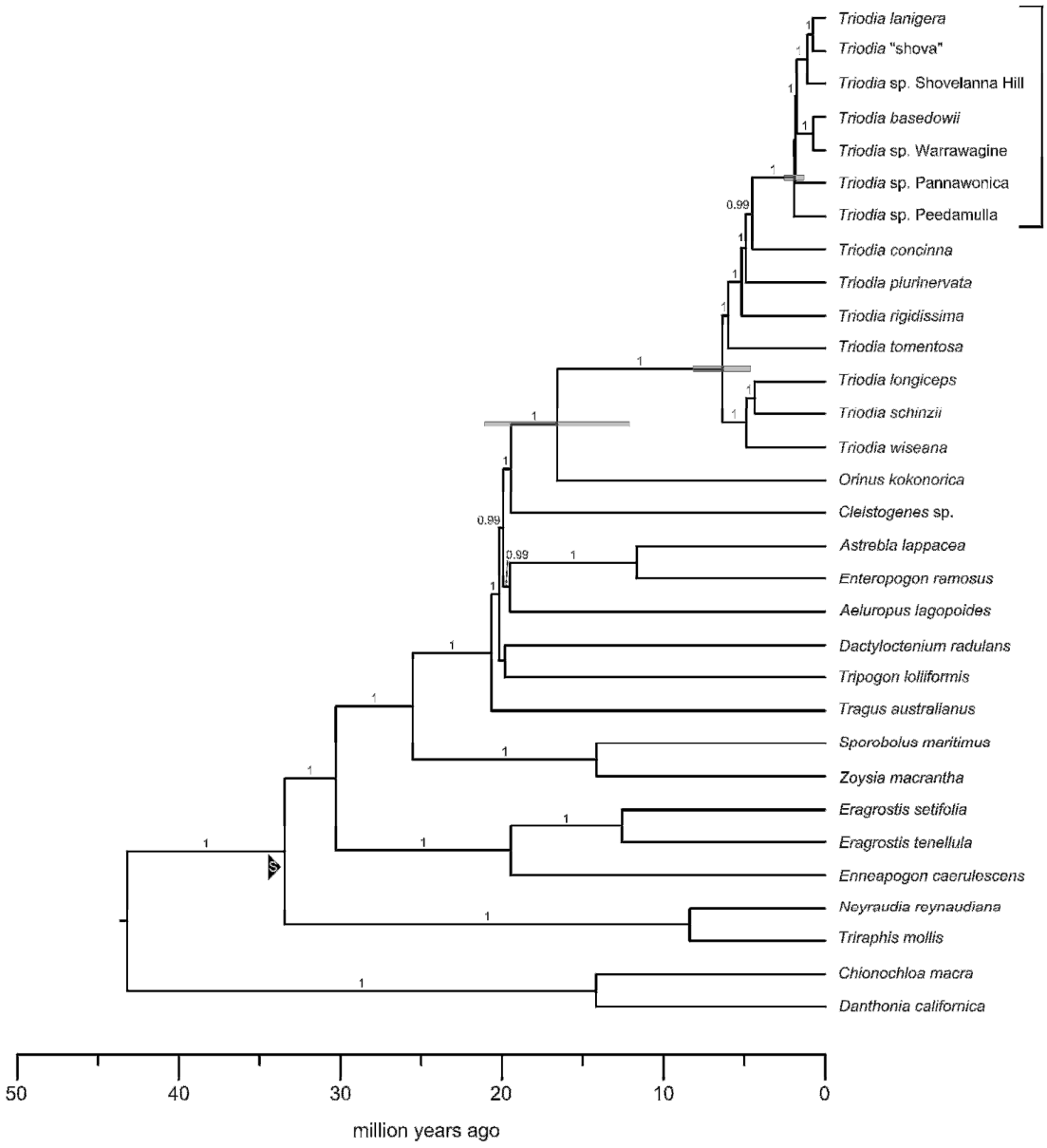

Figure 4.2. Chronogram from the BEAST analysis of data set B with secondary calibration from the first analysis when phytoliths are placed at stem BOP+PACMAD. The secondary calibration point is shown with a black triangle. Node bars for nodes of interest are 95% HPD intervals. Branch support values are posterior probabilities and are shown for values > 0.95. The *Triodia basedowii* complex is indicated at the top right.

Table 4.4. Node ages (Ma) for data set B across runs and using different model assumptions and calibration points. All runs used a random local clock model in BEAST except where indicated.

Node	stem *Triodia*		crown *Triodia*		crown *T. basedowii* complex	
Analysis	**Age**	**95% HPD**	**Age**	**95% HPD**	**Age**	**95% HPD**
BOP+PACMAD secondary calibration, combined runs	16.6	12.1, 21.1	6.35	4.62, 8.18	1.91	1.31, 2.54
Oryzeae secondary calibration, combined runs	18.4	13.5, 23.3	7.10	5.15, 9.04	2.14	1.48, 2.82
UCLN in BEAST 2 with BOP+PACMAD secondary calibration	11.6	6.61, 16.8	4.57	2.73, 6.72	1.03	0.550, 1.57
Toon *et al.* (2015)	20.9	17.9, 23.5	14.7	11.4, 18.3	4.58	2.60, 6.86
BOP+PACMAD secondary calibration plus crown *Triodia* calibrated to Toon *et al.* (2015)	25.4	20.8, 30.0	10.0	8.22, 11.9	2.98	2.28, 3.74
Oryzeae secondary calibration plus crown *Triodia* calibrated to Toon *et al.* (2015)	24.4	20.5, 28.2	9.65	8.10, 11.2	2.87	2.25, 3.51
crown *Triodia* calibrated to Toon *et al.* (2015) (no other calibrations)	37.4	28.4, 47.0	14.3	10.9, 17.7	4.33	3.11, 5.60
matK only, BEAST 2 with BOP+PACMAD secondary calibration	15.8	8.69, 24.0	8.81	4.21, 14.7	3.29	1.03, 6.31

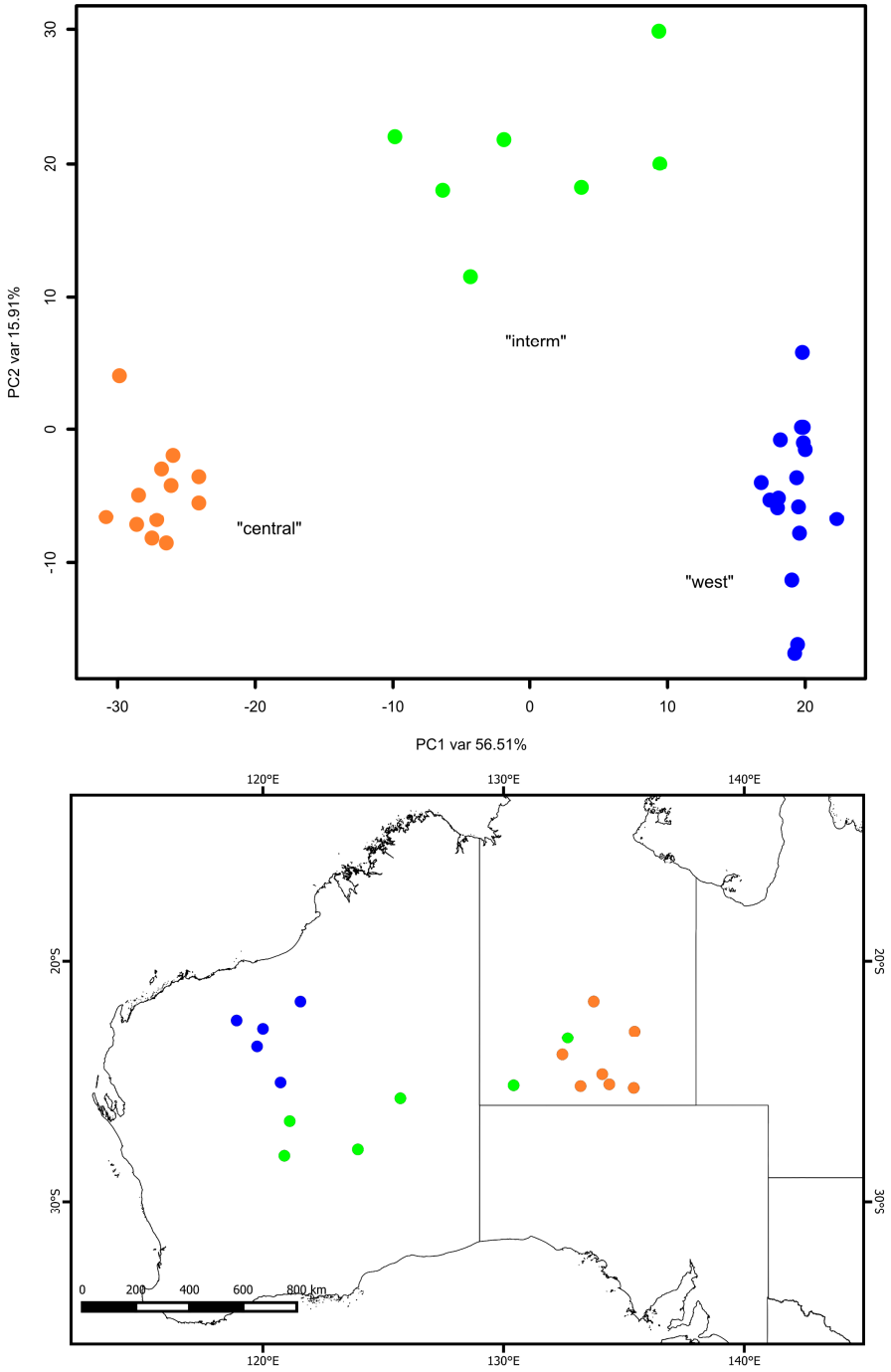

Figure 4.3. Principal components analysis for *Triodia basedowii* SNPs used in the Structure analyses. Labelled clusters correspond to groups mentioned in the text. A map of population locations is included below with populations coloured by group membership.

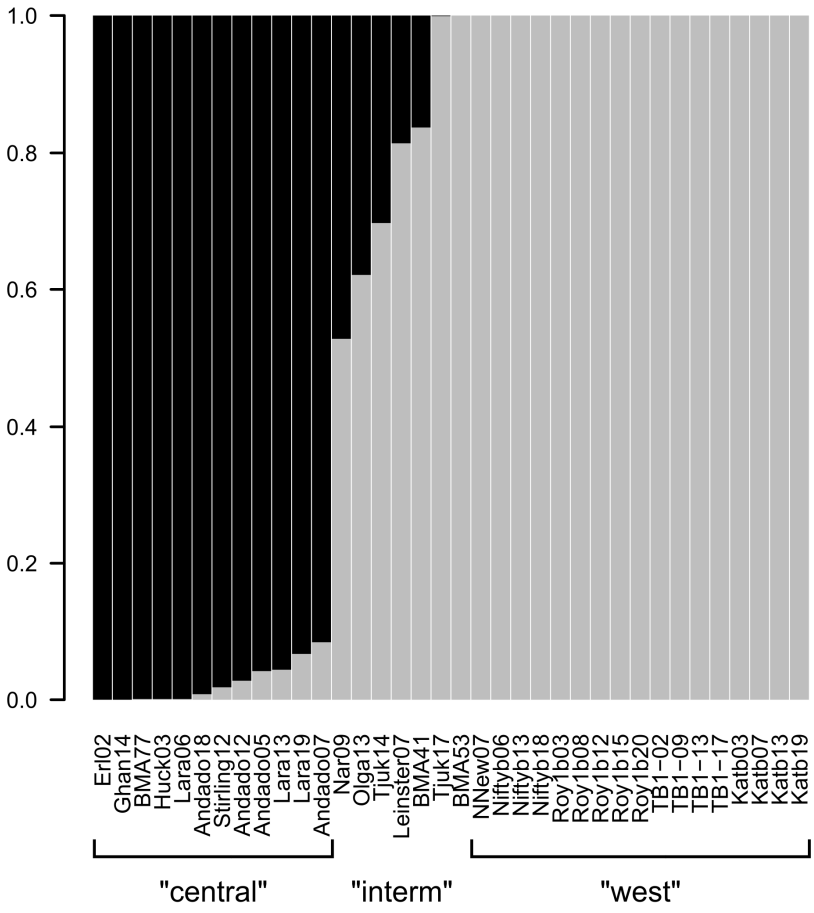

Figure 4.4. Structure results for *Triodia basedowii* from one of five runs for K=2. Groups identified from the PCoA (Fig. 4.3) are indicated at the bottom.

the "central" group and the "west" group, showing most of the "interm" samples as being admixed between them. The three groups detected in the PCoA were used for separate runs of the range expansion analyses, in addition to analysing all samples together.

Analyses based on the various data sets and choices of SNPs consistently resolved an expansion origin in or beyond the eastern portion of the current range of *T. basedowii* (Table 4.5; Figs. 4.5 & 4.6). Separate analyses of subsets corresponding to the three groups identified in the PCoA were typically not significantly different from a signal of isolation by distance, but running all samples together ("ALL") was consistently significant regardless of choice of SNP set or sampling. A subset of range expansion analyses are shown (Figs. 4.5 & 4.6) to illustrate the variation in results and the impact of choosing different SNP subsets. While the origin of expansion was consistently recovered in central or eastern Australia, the signal of expansion recovered between pairs of populations varied greatly across the different choices of included populations, SNP subset and outgroup.

Discussion

Using full chloroplast genome sequences we have provided a new estimate for divergence times across Poaceae and for *Triodia* and the *T. basedowii* species complex. Our results suggest that the *T. basedowii* complex began to radiate prior to the formation of sandy dunefields (*c*. 1 Ma) and that lineages persisted through Pleistocene glacial cycles and associated landscape changes. One of those lineages, *T. basedowii*, shows evidence of a recent range expansion originating from central or eastern Australia.

A temporal context for diversification in the T. basedowii *complex*

Our divergence dating provides new estimates for the age of crown *Triodia* (4.6–9.0 Ma 95% HPDs) and the *T. basedowii* complex (1.3–2.8 Ma 95% HPDs), but without a fossil calibration for our ingroup, it is difficult to be confident of these ages.

Table 4.5. Results from running the range expansion origin detection analyses with various choices of SNPs and samples. SNP sets are polarized using various outgroups ("E" *T.* sp. Little Sandy Desert, "F" *T.* "nana", "G" *T.* sp. Peedamulla) and use reads assembled or unassembled by PEAR or a combination of the two. SNPs were chosen randomly or with a bias toward SNPs with multiple copies of the rare allele. The number of SNPs indicates how many remain after the scripts remove those that have an ambiguous ancestral state. Sampling subsets are I: one per population (random), II: one per population (alternate choice), III: one per population excluding tetraploids, IV: all samples (populations downsampled), V: all samples excluding tetraploids (populations downsampled). Regions are "ALL", "west", "central" and "interm", and were deemed significant if $P < 0.01$. (table continued on next page)

SNP set	SNP choice	Scheme	# Samples	# SNPs	Sig. region	P-value	Location of origin
"E" assembled	random	I	18	2406	ALL	1.46E-08	NE Australia
		II	18	2392	-		
		III	15	2126	ALL	1.26E-12	NE Australia
		IV	36	3377	ALL	1.22E-13	NE NT
		V	27	2840	ALL	2.92E-15	NE Australia
"E" assembled	biased to biallelic	I	18	2409	ALL	2.14734E-07	SE NT
		II	18	2415	ALL	0.000182671	NW NT
		III	15	2138	ALL	5.62929E-08	central NT
		IV	36	3390	ALL	3.81E-18	S central NT
		V	27	2844	ALL	1.32E-16	SE NT
"E" unassembled	random	I	18	3009	-		
		II	18	2974	ALL	0.000230605	NE Australia
		III	15	2622	ALL	0.005937197	N NT
		IV	36	4222	ALL	7.15311E-13	S central NT
		V	27	3472	ALL	4.55E-17	E Australia
"E" unassembled	biased to biallelic	I	18	2993	central	0.005443	central NT
		II	18	3008	-		
		III	15	2644	-		
		IV	36	4239	ALL	1.17E-06	S central NT
		V	27	3542	ALL	1.97E-06	E Australia
"E" combined	random	I	18	5415	ALL	2.78E-08	NE Australia
		II	18	5366	ALL	5.96E-07	NE Australia
		III	15	4748	ALL	5.50E-15	NE Australia
		IV	36	7599	ALL	1.36E-15	NE NT
		V	27	6312	ALL	5.49E-21	E Australia
"E" combined	biased to biallelic	I	18	5402	ALL, central	0.000174, 0.00177	NE Australia, central NT
		II	18	5423	ALL	0.002025	NE NT
		III	15	4782	ALL	1.23E-06	N NT

SNP set	SNP choice	Scheme	# Samples	# SNPs	Sig. region	P-value	Location of origin
		IV	36	7629	ALL	1.27E-13	S central NT
		V	27	6386	ALL	2.05E-14	E Australia
"F" assembled	random	I	18	2297	-		
		II	18	2332	ALL	9.64E-13	NE NT
		III	15	2045	ALL	0.000329	NW NT
		IV	36	3251	ALL	1.23E-20	NE NT
		V	27	2752	ALL	1.54E-18	NE Australia
"F" assembled	biased to biallelic	I	18	2321	-		
		II	18	2357	-		
		III	15	2081	-		
		IV	36	3272	ALL	9.73E-08	S NT
		V	27	2770	ALL	5.73E-07	SE Australia
"F" combined	biased to biallelic	I	18	5169	-		
		II	18	5154	ALL	2.64E-06	S central NT
		III	15	4573	-		
		IV	36	7151	ALL	5.74E-10	S central NT
		V	27	6002	ALL	1.35E-10	E Australia
"G" assembled	random	I	18	2152	-		
		II	18	2167	-		
		III	15	1917	-		
		IV	36	2990	ALL	3.54E-11	N SA
		V	27	2560	ALL	5.95E-11	SE SA
"G" assembled	biased to biallelic	I	18	2081	ALL	0.000683	E NT
		II	18	2152	ALL	0.002865	SE NT
		III	15	1874	ALL	1.46E-07	E NT
		IV	36	2952	ALL	2.2E-14	S NT
		V	27	2541	ALL	2.26E-19	SE NT
"G" combined	biased to biallelic	I	18	4854	-		
		II	18	4921	ALL	0.006716	S central NT
		III	15	4339	ALL	0.00015	S central NT
		IV	36	6672	ALL	5.63E-13	S central NT
		V	27	5698	ALL	6.97E-17	SE NT

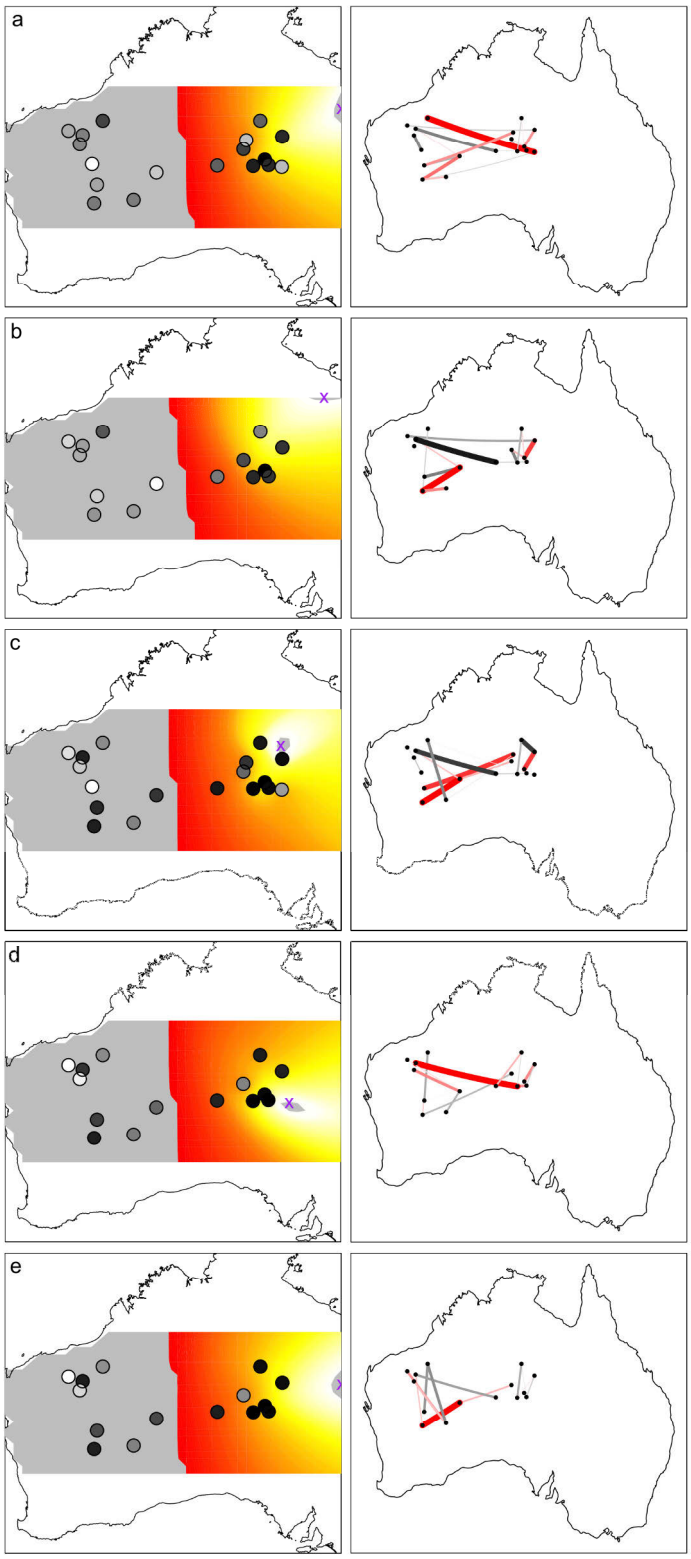

Figure 4.5. Origin of range expansion analyses for *Triodia basedowii* for different SNP subsets. Ancestral state of SNPs determined using *T.* sp. Little Sandy Desert as an outgroup. Subsets are (**a**) assembled, random, one per population; (**b**) assembled, random, one per population excluding tetraploids; (**c**) assembled, random, all samples; (**d**) assembled, biased to biallelic, all samples excluding tetraploids; and (**e**) combined, random, all samples excluding tetraploids. Figures on the left show a heat map and probable location of the origin of a range expansion (purple 'X'). Populations are indicated with circles, where darker shades show higher heterozygosity. Figures on the right show psi value comparisons between populations, with the thickness of the lines proportional to magnitude, and direction of expansion shown by colour (red indicates a westward expansion and black eastward).

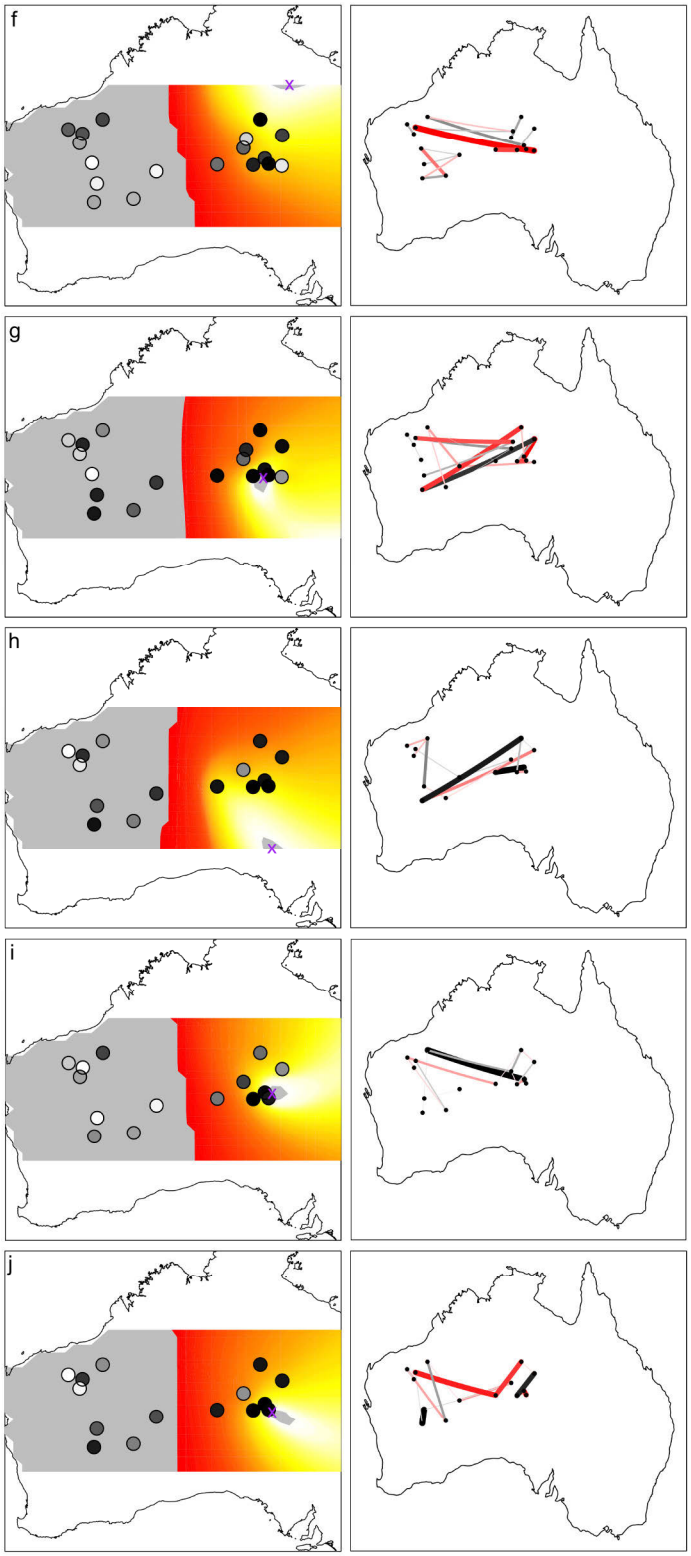

Figure 4.6. Origin of range expansion analyses for *Triodia basedowii* for different SNP subsets. Ancestral state of SNPs determined using either *T.* "nana" ((**f**) and (**g**)) or *T.* sp. Peedamulla ((**h**), (**i**), (**j**)) as an outgroup. Subsets are (**f**) assembled, random, one per population (alternate choice); (**g**) assembled, biased to biallelic, all samples; (**h**) assembled, random, all samples excluding tetraploids; (**i**) assembled, biased to biallelic, one per population excluding tetraploids; (**j**) combined, biased to biallelic, all samples excluding tetraploids. Figures on the left show a heat map and probable location of the origin of a range expansion (purple 'X'). Populations are indicated with circles, where darker shades show higher heterozygosity. Figures on the right show psi value comparisons between populations, with the thickness of the lines proportional to magnitude, and direction of expansion shown by colour (red indicates a westward expansion and black eastward).

Indeed, our estimate for the age of crown *Triodia* is significantly younger than the 11.4–18.3 Ma estimate obtained in a recent study (Toon *et al.* 2015). This is surprising, especially given Toon *et al.* (2015) used a younger secondary calibration (*c*. 32 Ma) for crown Chloridoideae. There are a few reasons why our dates might be younger. Firstly, we used a larger data set (chloroplast genome *vs*. ITS + *matK*), which we expect to be more informative for inferring rates of evolution. When we used the highly variable *matK* region alone, we obtained an older estimate for crown *Triodia* (4.2–14.7 Ma), although still not as old as that of Toon *et al.* (2015). Secondly, our sampling of *Triodia* is sparser, which may have produced a node density effect (see Heath *et al.* 2008; Simon Ho, pers. comm.), limiting the number of inferred substitutions. While undersampling across a tree has been shown to result in younger ages (and our sampling of Chloridoideae was sparser), undersampling of a specific clade was not observed to have the same effect on the age of the subtending node (Linder *et al.* 2005), suggesting that undersampling in *Triodia* may not fully explain the effect. Thirdly, our sampling of *Triodia* lacks species from northern tropical regions (included in Toon *et al.* (2015)), which may have higher rates of evolution and when included could bias lower nodes to be older (see Beaulieu *et al.* 2015). Finally, given Toon *et al.* (2015) found a rate shift in *Triodia* relative to other chloridoids, our analysis may have inferred that shift to have occurred nearer the base of the stem of *Triodia* (creating a relatively longer stem and younger crown).

Regardless of whether the *T. basedowii* complex began to diversify *c*. 2 Ma (this study) or *c*. 4.6 Ma (Toon *et al.* 2015), we can exclude the possibility that it represents a rapid recent radiation within the last 1 Ma, and we can therefore conclude that some lineages in the complex have persisted through the mid-Pleistocene climate shift and the formation of dunefields *c*. 1 Ma (Fujioka *et al.* 2009). While arid zone lineages have been suggested to have mostly diversified into current species by the onset of the Pleistocene 2.5 Ma (Byrne *et al.* 2008), our estimate for the crown of the complex suggests that this group of arid *Triodia* lineages has continued to diversify through the Pleistocene.

Expansion of T. basedowii *across the sandy dunefields*

Our results across different sampling schemes consistently indicate that *T. basedowii* has undergone a recent range expansion and that the most likely origin for that expansion is central or eastern Australia. The location of the origin does not support a hypothesis that the Pilbara region of Western Australia functioned to maintain lineages during adverse climatic conditions while they went extinct elsewhere, as might be predicted by the higher lineage richness there. While the pairwise signal for expansion varied across sampling schemes, the consistency of an east to west signal suggests the Pilbara did not function to maintain the bulk of *T. basedowii* prior to its expansion across sandy dunefields. Previous work (Anderson *et al.* 2016; Chapter 2) found no evidence for a refugium in central Australia, but was unable to look at diversity within *T. basedowii* populations. Our current analyses suggest the Central Ranges or some other region of central Australia may have indeed maintained populations of *T. basedowii* during the formation of the sandy dunefields and/or through Pleistocene glacial cycles.

As the location of the origin is variable across sampling schemes, we interpret our results cautiously with regards to a precise location. The limitations of the program are still relatively unexplored with regards to the impacts of missing data (as in our GBS data), and few studies have tested the program. A study on tropical skinks (Potter *et al.* 2016) used an exon capture approach to generate a SNP dataset (with almost no missing data) and used the program to detect likely range expansions for two lineages. It is likely the demographic history of *T. basedowii* is more complex than the single point source of expansion assumed by the model of Peter & Slatkin (2013), as populations may have diverged prior to acting as sources for others. The genetic structure in *T. basedowii* suggests a certain level of differentiation perhaps due to periods of minimal genetic exchange following an initial expansion (possibly during glacial maxima) or through ongoing isolation by distance. We interpret our results as indicating a westward expansion, but more complete sampling of populations and a more complete data set will improve the reliability of these inferences.

Our findings of a population expansion from eastern or central Australia also suggest that Pilbara populations of *T. basedowii* are relatively recent incursions into the sandy Fortescue River valley. These recent incursions may explain possible examples of hybridization with Pilbara species. One example of possible introgression involves *T. lanigera*, which based on genomic SNP data (Chapter 3) is part of a different clade and closely related to *T.* "shova" but which has an ITS sequence highly similar to that of *T. basedowii* (Anderson *et al.* 2016; Chapter 2). This possible ITS introgression in the absence of genomic mixing has previously been observed in a mixed population of *T. lanigera* and *T.* "shova" that shared ITS copies but were well differentiated across thousands of genomic SNPs (Chapter 3).

Historical biogeography of the T. basedowii *complex*

Our findings of early to mid-Pleistocene divergences in the *T. basedowii* complex and recent range expansion in *T. basedowii* allow us to propose, in conjunction with improved phylogenetic resolution (Chapter 3), a new hypothesis for the historical biogeography of the complex.

The two closest relatives of the *T. basedowii* complex (*T. plurinervata* N.T.Burb. and *T. concinna* N.T.Burb.) are currently distributed on the west coast and in the central interior of Western Australia (Anderson *et al.* 2016; Chapter 2), suggesting that ancestors of the *T. basedowii* complex probably occurred in central and western Western Australia (see Fig. 4.7). Recent work based on GBS (Chapter 3; summarised in Fig. 4.7b) suggests there are two main clades in the complex: a western group (*T.* sp. Pannawonica, *T.* sp. Peedamulla, *T. lanigera* and *T.* "shova") and an eastern group (*T. basedowii*, *T.* sp. Little Sandy Desert, *T.* sp. Shovelanna Hill, *T.* sp. Warrawagine, *T.* "nana"). These two groups are are also partly supported by chloroplast data (but note that *T.* sp. Shovelanna Hill has both chloroplast types and that the sample used in this study is part of the western group rather than the eastern group as predicted by GBS) (Anderson *et al.* 2016; Chapter 2). We suggest that early divergences in the complex included a split between the western and eastern groups.

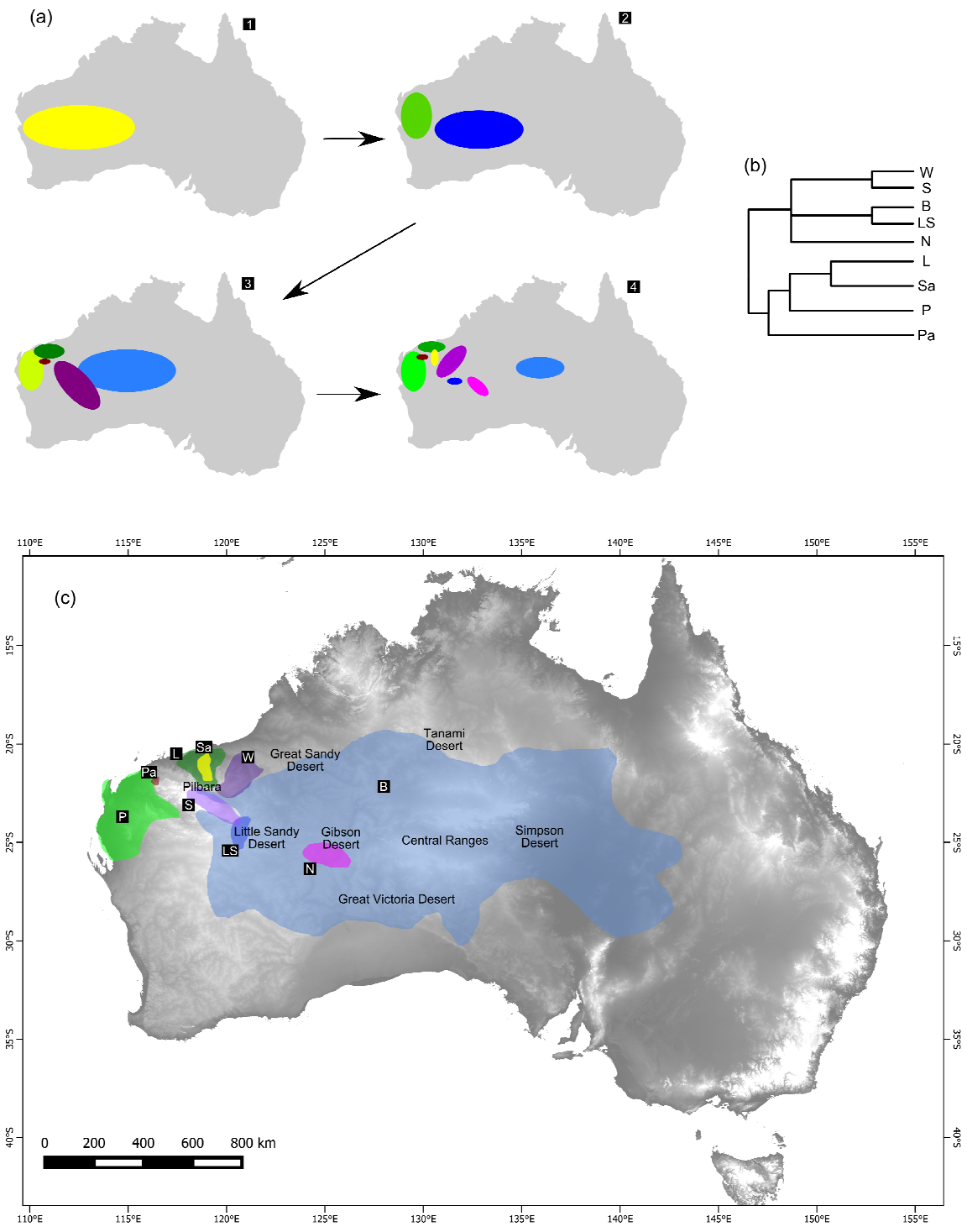

Figure 4.7. Hypothesized historical biogeography of the *Triodia basedowii* complex. (caption continued on next page)

Figure 4.7. (cont.) (**a**) Stages of lineage splitting referred to in the text, namely 1) ancestors of the complex in western and central Western Australia, 2) split between the eastern and western clades, 3) initial splitting of each group with the eastern group splitting into rocky and sandy lineages, and 4) range restrictions and lineage splitting of most of the current taxa prior to the expansion of *T. basedowii*. (**b**) Approximate topology of phylogenetic relationships for taxa based on GBS data (Chapter 3). (**c**) Current distribution of taxa in the complex. Taxa are denoted by letters, where P: *T.* sp. Peedamulla, Pa: *T.* sp. Pannawonica, L: *T. lanigera*, Sa: *T.* "shova", S: *T.* sp. Shovelanna Hill, W: *T.* sp. Warrawagine, LS: *T.* sp. Little Sandy Desert, N: *T.* "nana", and B: *T. basedowii*.

The western group probably diversified along the west coast and into the northern Pilbara. Current habitat preferences of *T.* sp. Peedamulla compared to *T.* "shova" (sandy compared to gravelly substrates) suggest that divergences may have been driven by ecological differentiation. Alternatively, species such as *T. lanigera* and *T.* "shova", which currently have overlapping distributions, may have diverged allopatrically if climatic conditions restricted their ranges. In the eastern group, there are three species that are largely restricted to rocky substrates (*T.* sp. Shovelanna Hill, *T.* sp. Warrawagine, and *T.* "nana") and two that are found on sandy substrates (*T. basedowii* and *T.* sp. Little Sandy Desert). The extent of sandy habitats prior to the formation of the dunefields *c.* 1 Ma is currently unknown, and it is possible that the ancestors of the sand specialists had available habitat in central Australia. We suggest that the eastern group split into a sandy lineage and a rocky lineage, potentially as a result of ecological differentiation. The rocky lineage diversified into three lineages, two of which are found in the southern and eastern parts of the Pilbara and the third in central Western Australia, possibly as a result of vicariant events caused by the extension of inhospitable sandy habitats during aridification. *Triodia basedowii* was likely restricted to central or eastern Australia before or during the formation of the dunefields *c.* 1 Ma (possibly splitting it from the lineage of *T.* sp. Little Sandy Desert at that point). More recently, *T. basedowii* has expanded westward across the new sandy habitat, with regional genetic differentiation possibly caused by glacial-cycle driven range restrictions or isolation by distance.

The higher lineage richness in the *T. basedowii* complex in the Pilbara compared to central Australia likely reflects a combination of maintenance of lineages and promotion of diversification. Our age estimates for the complex rule out a recent, young radiation (which would have made a lineage maintenance scenario unlikely), and so are unhelpful in distinguishing the two processes. Since both the east and west *T. basedowii* complex clades occur in the Pilbara and include taxa outside the Pilbara (e.g. *T.* "nana"), the Pilbara lineages are not monophyletic, so the higher lineage richness is likely not solely due to *in situ* diversification. Instead, the Pilbara has probably been colonised by both clades, with subsequent diversification (e.g. *T. lanigera* splitting from *T.* "shova" and *T.* sp. Shovelanna Hill splitting from *T.* sp. Warrawagine) possibly driven by substrate differences. It appears that the Pilbara

lineages have persisted there through climatic stresses during glacials. We can speculate that currently extinct lineages may have existed outside the Pilbara prior to the shift in glacial cycle amplitude *c*. 1 Ma.

Divergence dates across Poaceae

While the focus of our study is on the *T. basedowii* complex, we also present here the first molecular dating of divergences across all subfamilies of Poaceae using extensive alignments from whole chloroplast genomes. In addition, our phylogenetic analysis of alignments across Poaceae adds a new estimate of grass phylogeny to compare with existing estimates (e.g. Grass Phylogeny Working Group 2001; Bouchenak-Khelladi *et al*. 2008; Grass Phylogeny Working Group II 2012; Soreng *et al*. 2015). Our node ages for grass subfamilies (except Pooideae) fall between estimates from two recent studies (Prasad *et al.* 2011; Christin *et al.* 2014) that used select chloroplast regions and included the phytoliths (see Table 4.3). Our estimate for the age of crown Pooideae is *c*. 7 Ma younger than estimates in the above studies, and differences in branch lengths and node ages may be partly explained by our sparser sampling and potential omission of close relatives (see Heath *et al.* 2008). We included the controversial phytoliths at two positions to account for uncertainty in their phylogenetic placement. While Christin *et al.* (2014) suggested that dates obtained with the phytoliths should be reported cautiously, we think our placements account for the uncertainty and are preferable to excluding the phytoliths. The placements of the phytoliths did not substantially impact our estimates of subfamily crown ages except for Oryzoideae, suggesting that other fossils are constraining the tree, such as the Zucol *et al.* (2010) fossil, which has been included in Prasad *et al.* (2011) but not in other dating analyses (Christin *et al.* 2008, 2014; Vicentini *et al.* 2008; Bouchenak-Khelladi *et al.* 2009, 2010, 2014).

Conclusions

The *Triodia basedowii* species complex began to diversify prior to the mid-Pleistocene shift in glacial cycle amplitude and the earliest evidence for sandy dunefields in Australia (*c*. 1 Ma), and therefore its lineages have persisted (and possibly diversified further) through late Pleistocene climatic instability. Our results

suggest an evolutionary resilience in Australian arid zone plants to massive landscape changes and climate variability associated with increasing aridification since the Pliocene. They also illustrate variable responses of lineages to the effects of aridity, where, even within a species complex of close relatives, some are restricted to narrow ranges associated with specific substrates while others have been able to expand onto newer sandy dunefield habitats.

Acknowledgements

We thank S. van Leeuwen, N. Gibson, M. Langley, K. Brown, P. Jobson, and R. Davis for assistance with fieldwork, and E. Mayence for collecting a population of *T. plurinervata* from the Shark Bay area. We thank Melvin Duvall for providing an assembly of a previously sequenced *T.* sp. Shovelanna Hill chloroplast, and Nicolas Dierckxsens for providing a beta-test version of NOVOplasty. Michael Gardner and the staff of the Australian Genome Research Facility are thanked for running the shotgun sequencing. This research is partly supported by an Australian Research Council Linkage Project 120100350 (CIs Grierson, Krauss *et al.* in collaboration with Rio Tinto Iron Ore, Chevron, Botanic Gardens and Parks Authority and the Department of Parks and Wildlife). Collecting in eastern Western Australia and central Australia was supported through an ANZ Trustees Foundation - Holsworth Wildlife Research Endowment. B.M.A. was supported by an International Postgraduate Research Scholarship, an Australian Postgraduate Award and a UWA Top-up scholarship at The University of Western Australia.

Supplementary Information

Supplementary information can be found in Appendix C, along with some of the scripts written for data handling and analysis.

References

Anderson BM, Barrett MD, Krauss SL, Thiele K (2016) Untangling a species complex of arid zone grasses (*Triodia*) reveals patterns congruent with co-occurring animals. *Molecular Phylogenetics and Evolution*, **101**, 142–162.

Ariati SR, Murphy DJ, Udovicic F, Ladiges PY (2006) Molecular phylogeny of three

groups of acacias (*Acacia* subgenus *Phyllodineae*) in arid Australia based on the internal and external transcribed spacer regions of nrDNA. *Systematics and Biodiversity*, **4**, 417–426.

Baele G, Lemey P, Bedford T *et al.* (2012) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution*, **29**, 2157–2167.

Bankevich A, Nurk S, Antipov D *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, **19**, 455–477.

Beaulieu JM, O'Meara BC, Crane P, Donoghue MJ (2015) Heterogeneous rates of molecular evolution and diversification could explain the Triassic age estimate for angiosperms. *Systematic Biology*, **64**, 869–878.

Becker RA, Wilks AR, Brownrigg R, Minka TP, Deckmyn A (2016) maps: draw geographical maps. R package version 3.1.0.

Bouchenak-Khelladi Y, Salamin N, Savolainen V *et al.* (2008) Large multi-gene phylogenetic trees of the grasses (Poaceae): progress towards complete tribal and generic level sampling. *Molecular Phylogenetics and Evolution*, **47**, 488–505.

Bouchenak-Khelladi Y, Slingsby JA, Verboom GA, Bond WJ (2014) Diversification of C_4 grasses (Poaceae) does not coincide with their ecological dominance. *American Journal of Botany*, **101**, 300–307.

Bouchenak-Khelladi Y, Verboom GA, Hodkinson TR *et al.* (2009) The origins and diversification of C4 grasses and savanna-adapted ungulates. *Global Change Biology*, **15**, 2397–2417.

Bouchenak-Khelladi Y, Verboom GA, Savolainen V, Hodkinson TR (2010) Biogeography of the grasses (Poaceae): a phylogenetic approach to reveal evolutionary history in geographical space and geological time. *Botanical Journal of the Linnean Society*, **162**, 543–557.

Bouckaert R (2015) bModelTest: Bayesian site model selection for nucleotide data. *bioRxiv*. DOI: 10.1101/020792.

Bouckaert R, Heled J, Kühnert D *et al.* (2014) BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLOS Computational Biology*, **10**, 1–6.

Byrne M, Yeates DK, Joseph L *et al.* (2008) Birth of a biome: insights into the assembly and maintenance of the Australian arid zone biota. *Molecular Ecology*, **17**, 4398–4417.

Christin P-A, Besnard G, Samaritani E *et al.* (2008) Oligocene CO2 decline promoted C4 photosynthesis in grasses. *Current Biology*, **18**, 37–43.

Christin P-A, Spriggs E, Osborne CP *et al.* (2014) Molecular dating, evolutionary rates, and the age of the grasses. *Systematic Biology*, **63**, 153–165.

Clark P, Alley R, Pollard D (1999) Northern Hemisphere ice-sheet influences on global climate change. *Science*, **286**, 1104–1111.

Crisp M, Cook L (2013) How was the Australian flora assembled over the last 65 million years? A molecular phylogenetic perspective. *Annual Review of Ecology, Evolution, and Systematics*, **44**, 303–324.

Crisp MD, Hardy NB, Cook LG (2014) Clock model makes a large difference to age estimates of long-stemmed clades with no internal calibration: a test using Australian grasstrees. *BMC Evolutionary Biology*, **14**, 263.

Davey JW, Hohenlohe PA, Etter PD *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews. Genetics*, **12**, 499–510.

Department of the Environment (2006) *MVG 20 - Hummock Grasslands*. Australian Government, Canberra, ACT.

Doyle JA (1992) Revised palynological correlations of the lower Potomac group (USA) and the cocobeach sequence of Gabon (Barremian-Aptian). *Cretaceous Research*, **13**, 337–349.

Doyle JJ, Dickson EE (1987) Preservation of plant samples for DNA restriction endonuclease analysis. *Taxon*, **36**, 715–722.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLOS Biology*, **4**, e88.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Drummond AJ, Suchard MA (2010) Bayesian random local clocks, or one rate to rule them all. *BMC Biology*, **8**, 114.

Duvall MR, Fisher AE, Columbus JT *et al.* (2016) Phylogenomics and plastome evolution of the chloridoid grasses (Chloridoideae: Poaceae). *International Journal of Plant Sciences*, **177**, 235–246.

Eaton DAR (2014) PyRAD: assembly of *de novo* RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1844–1849.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Fu Y-B (2014) Genetic diversity analysis of highly incomplete SNP genotype data with imputations: an empirical assessment. *G3: Genes, Genomes, Genetics*, **4**, 891–900.

Fujioka T, Chappell J, Fifield LK, Rhodes EJ (2009) Australian desert dune fields initiated with Pliocene-Pleistocene global climatic shift. *Geology*, **37**, 51–54.

Gardner MG, Fitch AJ, Bertozzi T, Lowe AJ (2011) Rise of the machines – recommendations for ecologists when using next generation sequencing for microsatellite development. *Molecular Ecology Resources*, **11**, 1093–1101.

Grass Phylogeny Working Group (2001) Phylogeny and subfamilial classification of the grasses (Poaceae). *Annals of the Missouri Botanical Garden*, **88**, 373–457.

Grass Phylogeny Working Group II (2012) New grass phylogeny resolves deep

evolutionary relationships and discovers C_4 origins. *New Phytologist*, **193**, 304–312.

Gustafsson F, Gunnarsson F (2003) Positioning using time-difference of arrival measurements. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03)* , pp. 553–556.

Heath TA, Hedtke SM, Hillis DM (2008) Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution*, **46**, 239–257.

Hesse PP (2010) The Australian desert dunefields: formation and evolution in an old, flat, dry continent. In: *Australian Landscapes* (eds Bishop P, Pillans B), pp. 141–164. The Geological Society of London.

Hesse PP, Magee JW, van der Kaars S (2004) Late Quaternary climates of the Australian arid zone: a review. *Quaternary International*, **118-119**, 87–102.

Hijmans RJ (2015) geosphere: spherical trigonometry. R package version 1.5-1.

Jabaily RS, Shepherd KA, Gardner AG *et al.* (2014) Historical biogeography of the predominantly Australian plant family Goodeniaceae. *Journal of Biogeography*, **41**, 2057–2067.

Kadereit G, Freitag H (2011) Molecular phylogeny of Camphorosmeae (Camphorosmoideae, Chenopodiaceae): implications for biogeography, evolution of C_4-photosynthesis and taxonomy. *Taxon*, **60**, 51–78.

Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, **30**, 772–780.

Kearse M, Moir R, Wilson A *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.

Kurtz S, Phillippy A, Delcher AL *et al.* (2004) Versatile and open software for comparing large genomes. *Genome biology*, **5**, R12.

Lanfear R, Calcott B, Ho SYW, Guindon S (2012) Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**, 1695–1701.

Lazarides M (1997) A revision of *Triodia* including *Plectrachne* (Poaceae, Eragrostideae, Triodiinae). *Australian Systematic Botany*, **10**, 381–489.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Linder HP, Hardy CR, Rutschmann F (2005) Taxon sampling effects in molecular clock dating: An example from the African Restionaceae. *Molecular Phylogenetics and Evolution*, **35**, 569–582.

MacGinitie HD (1953) Fossil plants of the Florissant beds, Colorado. *Publications of the Carnegie Institution of Washington*, **599**.

Manchester SR (2001) Update on the megafossil flora of Florissant, Colorado. *Proceedings of the Denver Museum of Nature & Science*, **4**, 137–161.

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10–12.

McIlroy D, Brownrigg R, Minka TP, Bivand R (2015) mapproj: map projections. R package version 1.2-4.

McLoughlin S (2001) The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. *Australian Journal of Botany*, **49**, 271–300.

Mummenhoff K, Linder P, Friesen N *et al.* (2004) Molecular evidence for bicontinental hybridogenous genomic constitution in *Lepidium sensu stricto* (Brassicaceae) species from Australia and New Zealand. *American Journal of Botany*, **91**, 254–261.

Pebesma EJ, Bivand RS (2005) Classes and methods for spatial data in R. *R News*, **5**, 9–13.

Peter BM, Slatkin M (2013) Detecting range expansions from genetic data. *Evolution*, **67**, 3274–3289.

Peter BM, Slatkin M (2015) The effective founder effect in a spatially expanding population. *Evolution*, **69**, 721–734.

Peterson PM, Romaschenko K, Johnson G (2010) A classification of the Chloridoideae (Poaceae) based on multi-gene phylogenetic trees. *Molecular Phylogenetics and Evolution*, **55**, 580–598.

Piperno DR, Pearsall DM (1998) The silica bodies of tropical American grasses: morphology, taxonomy, and implications for grass systematics and fossil phytolith identification. *Smithsonian Contributions to Botany*, **85**, 1–40.

Potter S, Bragg JG, Peter BM, Bi K, Moritz C (2016) Phylogenomics at the tips: inferring lineages and their demographic history in a tropical lizard, *Carlia amax*. *Molecular Ecology*, **25**, 1367–1380.

Prasad V, Strömberg CAE, Alimohammadian H, Sahni A (2005) Dinosaur coprolites and the early evolution of grasses and grazers. *Science*, **310**, 1177–1180.

Prasad V, Strömberg CAE, Leaché AD *et al.* (2011) Late Cretaceous origin of the rice tribe provides evidence for early diversification in Poaceae. *Nature Communications*, **2**, 480.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Python Software Foundation (2016) Python Language Reference, version 2.7. https://docs.python.org/2.7/.

R Development Core Team (2015) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rambaut A (2014) FigTree. http://tree.bio.ed.ac.uk/software/figtree/.

Rambaut A, Suchard M, Xie D, Drummond A (2014) Tracer. http://beast.bio.ed.ac.uk/Tracer.

Ronquist F, Teslenko M, van der Mark P *et al.* (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space.

Systematic Biology, **61**, 539–542.

Sakaguchi S, Bowman DMJS, Prior LD *et al.* (2013) Climate, not Aboriginal landscape burning, controlled the historical demography and distribution of fire-sensitive conifer populations across Australia. *Proceedings of the Royal Society B. Biological Sciences*, **280**, 20132182.

Shepherd K, Waycott M, Calladine A (2004) Radiation of the Australian Salicornioideae (Chenopodiaceae)—based on evidence from nuclear and chloroplast DNA sequences. *American Journal of Botany*, **91**, 1387–1397.

Soreng RJ, Peterson PM, Romaschenko K *et al.* (2015) A worldwide phylogenetic classification of the Poaceae (Gramineae). *Journal of Systematics and Evolution*, **53**, 117–137.

South A (2011) rworldmap: a new R package for mapping global data. *The R Journal*, **3**, 35–43.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology*, **57**, 758–771.

Strömberg CAE (2005) Decoupled taxonomic radiation and ecological expansion of open-habitat grasses in the Cenozoic of North America. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 11980–11984.

Toon A, Crisp MD, Gamage H *et al.* (2015) Key innovation or adaptive change? A test of leaf traits using Triodiinae in Australia. *Scientific Reports*, **5**, 12398.

Urban MA, Nelson DM, Jiménez-Moreno G *et al.* (2010) Isotopic evidence of C_4 grasses in southwestern Europe during the Early Oligocene-Middle Miocene. *Geology*, **38**, 1091–1094.

Veevers JJ, Powell CM, Roots SR (1991) Review of seafloor spreading around Australia. I. Synthesis of the patterns of spreading. *Australian Journal of Earth Sciences*, **38**, 373–389.

Vicentini A, Barber JC, Aliscioni SS, Giussani LM, Kellogg EA (2008) The age of the grasses and clusters of origins of C4 photosynthesis. *Global Change Biology*, **14**, 2963–2977.

Walker BJ, Abeel T, Shea T *et al.* (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS ONE*, **9**, e112963.

Walker J, Geissman J, Bowring S, Babcock L (2012) *Geological Time Scale v. 4.0*. Geological Society of America.

Wysocki WP, Clark LG, Kelchner SA *et al.* (2014) A multi-step comparison of short-read full plastome sequence assembly methods in grasses. *Taxon*, **63**, 899–910.

Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly

using de Bruijn graphs. *Genome Research*, **18**, 821–829.

Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, **30**, 614–620.

Zucol AF, Brea M, Bellosi ES (2010) Phytolith studies in Gran Barranca (central Patagonia, Argentina): the middle-late Eocene. In: *The Paleontology of Gran Barranca: Evolution and Environmental Change through the Middle Cenozoic of Patagonia* (eds Madden RH, Carlini AA, Vucetich MG, Kay RF), pp. 317–340. Cambridge University Press.