'Next generation' biogeography: towards understanding the drivers of species diversification and persistence

Christian Lexer[1]*, Sofia Mangili[1], Eligio Bossolini[1], Felix Forest[2], Kai N. Stölting[1], Peter B. Pearman[3], Niklaus E. Zimmermann[3] and Nicolas Salamin[4,5]

[1]Unit of Ecology & Evolution, Department of Biology, University of Fribourg, CH-1700, Fribourg, Switzerland, [2]Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3DS, UK, [3]Landscape Dynamics Unit, Swiss Federal Research Institute WSL, CH-8903, Birmensdorf, Switzerland, [4]Department of Ecology & Evolution, University of Lausanne, Biophore, CH-1015, Lausanne, Switzerland, [5]Swiss Institute of Bioinformatics, Quartier Sorge, CH-1015, Lausanne, Switzerland

*Correspondence: Christian Lexer, Unit of Ecology & Evolution, Department of Biology, University of Fribourg, Chemin du Musée 10, CH-1700, Fribourg, Switzerland.
E-mail: christian.lexer@unifr.ch

ABSTRACT

The drivers of species diversification and persistence are of great interest to current biogeography, especially in those global biodiversity 'hotspots' harbouring most of Earth's animal and plant life. Classical multispecies biogeographical work has yielded fascinating insights into broad-scale patterns of diversification, and DNA-based intraspecific phylogeographical studies have started to complement this picture at much finer temporal and spatial scales. The advent of novel next-generation sequencing (NGS) technologies provides the opportunity to greatly scale up the numbers of individuals, populations and species sampled, potentially merging intraspecific and interspecific approaches to biogeographical inference. Here, we outline these prospects and issues by using the example of an undisputed hotspot, the Cape of southern Africa. We outline the current state of knowledge on the biogeography of species diversification within the Cape, review the literature for phylogeographical evidence of its likely drivers and mechanisms, and suggest possible ways forward based on NGS approaches. We demonstrate the potential of these methods and current bioinformatic issues with the help of restriction-site-associated DNA (RAD) sequencing data for three highly divergent species of the Restionaceae, an important plant radiation in the Cape. A thorough understanding of the mechanisms that facilitate species diversification and persistence in spatially structured, species-rich environments will require the adoption of novel genomic and bioinformatic tools in biogeographical studies.

Keywords

Biodiversity hotspots, Cape Floristic Region, diversification, gene flow, next generation sequencing, phylogeography, RAD, radiation, speciation, species cohesion.

INTRODUCTION

Understanding the drivers and mechanisms of species diversification and persistence is of central interest to biogeography, evolutionary biology and conservation genetics (Frankham *et al.*, 2004; Futuyma, 2009; Höglund, 2009; Ladle & Whittaker, 2011). The need to understand these issues is particularly pressing in the world's biodiversity hotspots, i.e. those regions of the planet where both the challenges and payoffs of conservation measures are expected to be greatest (Myers *et al.*, 2000; Mittermeier *et al.*, 2005). The origin and maintenance of biological diversity in species-rich environments have been addressed using the concepts of historical biogeography, phylogenetics and comparative biology (Wiens & Donoghue, 2004; Emerson & Gillespie, 2008; Jablonski, 2008;

Linder, 2008; Antonelli *et al.*, 2009; Cavender-Bares *et al.*, 2009; Salamin *et al.*, 2010). These approaches have greatly advanced our understanding of key issues such as: what was the speed of a particular radiation, or of entire suites of radiations making up a biodiversity hotspot? Has diversification involved explosive bursts of speciation (= adaptive radiation), or has it proceeded gradually? Can environmental aspects such as area, climate or topography explain spatial and temporal patterns of diversity?

While large-scale comparative approaches have been remarkably successful in addressing these questions on the origin and accumulation of biodiversity (e.g. Wiens & Donoghue, 2004; Linder, 2008; Cavender-Bares *et al.*, 2009; Verboom *et al.*, 2009; Salamin *et al.*, 2010), *within*-species phylogeographical and genetic data are required to address

the *mechanisms* responsible for species diversification and persistence in strongly structured, species-rich environments, and to predict the likely future responses to environmental (e.g. climate) change. Mechanisms mediating species persistence may comprise intrinsic barriers to gene exchange maintaining the integrity of populations, or ecological mechanisms affecting entire genomes via population demography and specific genome regions via natural selection (Conner & Hartl, 2004; Stinchcombe & Hoekstra, 2008). Unfortunately, classical multispecies biogeographical studies tend to treat species as homogeneous units and employ just one (or at best a few) individual(s) per species (Salamin *et al.*, 2010), which precludes an understanding of genetic variation at exactly the level that matters most to the evolutionary process – the pool of standing variation available to individual-level, natural selection and drift (Futuyma, 2009). Within-species phylogeographical studies, on the other hand, often employed hundreds or even thousands of individuals but have thus far been limited to only one or a few related species per study. This lack of overlap and synthesis between multispecies biogeographical studies and intraspecific phylogeographical work makes it virtually impossible to judge the commonalities and mismatches between processes operating at macroevolutionary versus microevolutionary time-scales (Jablonski, 2008).

Molecular and analytical tools exist to track phylogeographical lineages *within* species or groups of closely related taxa (Arbogast & Kenagy, 2001; Excoffier, 2004; Avise, 2009; and see e.g. Dawson, 2001; Cannon & Manos, 2003; Costa, 2003; Swart *et al.*, 2009; Pinheiro *et al.*, 2011, for recent examples published in this journal). Nevertheless, these studies are thinly and unevenly spread across the globe. Southern Hemisphere biomes, for example, find little representation in phylogeographical studies (Beheregaray, 2008). Also, most phylogeographical studies to date are based on 'neutral' genetic markers, which do not permit tracking of the adaptive genetic variation that allows populations to respond directly to environmental change (for discussion of this issue see e.g. Davis & Shaw, 2001; Stinchcombe & Hoekstra, 2008; de Carvalho *et al.*, 2010; Hohenlohe *et al.*, 2010). The striking lack of intraspecific genetic data for geographical regions of great scientific and conservation interest is most readily exemplified by an undisputed biodiversity hotspot, the Cape of southern Africa.

DRIVERS OF SPECIES PERSISTENCE: THE CAPE OF SOUTHERN AFRICA AS AN EXAMPLE

The Cape Floristic Region (CFR) of southern Africa is a global biodiversity hotspot (Myers *et al.*, 2000; Linder, 2005; Mittermeier *et al.*, 2005) covering < 90,000 km^2 and harbouring more than 9000 vascular plant species, roughly 70% of which are endemic to this region (Goldblatt & Manning, 2002; note the broader delimitation as Greater CFR in recent studies: Linder *et al.*, 2010). The hypotheses regarding origin and maintenance of this diversity revolve around

climatic stability *and* change. Climatic stability over millions of years is now thought to be the chief factor responsible for the accumulation of species richness over geological times. This becomes apparent from molecular phylogenetic studies indicating an early onset of important plant radiations in the Cape (reaching back into the Oligocene, > 25 Ma; Verboom *et al.*, 2009; Schnitzler *et al.*, 2011) and relatively constant and moderate (by global standards) rates of net diversification ever since (Linder, 2008; Valente *et al.*, 2010; Schnitzler *et al.*, 2011).

In contrast, the key mechanisms currently generating and maintaining species richness in the Cape appear to be tightly coupled to climatic oscillations. Range modelling of Cape biomes (mediterranean fynbos shrubland and succulent karoo semi-desert) and individual species based on fossil pollen indicates rapid range dynamics within the last few hundreds of thousands of years (i.e. during the Pleistocene; Midgley *et al.*, 2003, 2005). Given the mountainous topography and east–west climatic differentiation of the area, this suggests the presence of 'species pumps', with topographic diversity continuously shuffling populations along elevational and longitudinal (moisture) gradients and across a varied mosaic of soil types and microclimates (Midgley *et al.*, 2003, 2005; Schnitzler *et al.*, 2011). Thus, the key to understanding species richness in the Cape biodiversity hotspot appears to lie in understanding how so many species have evolved and *persisted* as cohesive units in so little space (Valente *et al.*, 2010), despite the frequent opportunity for contact and genetic interactions that tend to accompany climatic oscillations (Hewitt, 2000). This calls for DNA-based phylogeographical studies that explicitly address the population-level mechanisms of divergence and persistence, i.e. the actual 'engine' of the evolutionary process. Strikingly few phylogeographical studies are currently available, considering the intensity of the debate surrounding the determinants of species richness in the Cape. We shall use the Cape as an example to sketch the current state-of-the-art, existing limitations, and potential future solutions for DNA-based biogeographical work in spatially structured, species-rich environments.

PHYLOGEOGRAPHICAL STUDIES IN THE CAPE: FEW AND FAR BETWEEN

An exhaustive literature search for intraspecific phylogeographical studies in the South African Cape using Web of Science (search terms: various combinations of 'Cape' and 'Africa' with 'phylogeography', 'DNA', 'population', and 'marker') double-checked by Google Scholar reveals 17 studies, including 12 on animals and only 5 on plants (Table 1). We mined these 17 studies for answers to two basic questions: (1) Is there phylogeographical structure in the studied taxa? (2) Are there indications of recent range dynamics (e.g. demographic/range expansions) in the phylogeographical data? The analytical tools used by these studies allow separate tests for *structure* and *range dynamics,* thus these two features were reported separately (Table 1). We then cross-tabulated

Table 1 Literature survey of within-species phylogeographical and population genetic work carried out for plants and animals in an exemplary world biodiversity hotspot, the Cape of southern Africa, including reference, target species, taxonomic group, mating system (outcrossing/mixed/unknown), generation time [plants, annual (ann) vs. perennial (per); animals, iteroparous (itero) vs. semelparous (semel)], genomic compartment sampled (cp, chloroplast; mt, mitochondrial; nuc, nuclear), and indications of phylogeographical structure and recent range dynamics (demographic/range expansions) as indicated by the molecular genetic data.

References	Species	Group	Mating system	Generation time	Genome	Structure*	Dynamics†
Tolley *et al.* (2010)	*Strongylopus grayii*	Amphibia	Outcrossing	per/itero	mt	y	n
Bergh *et al.* (2007)	*Elytropappus rhinocerotis*	Angiosperma	Outcrossing	per/itero	nuc	n	(y)‡
Prunier & Holsinger (2010)	*Protea* spp.	Angiosperma	Mixed	per/itero	nuc	y	y
Rymer *et al.* (2010)	*Gladiolus* spp.	Angiosperma	Mixed	per/itero	cp/nuc	n	n
Segarra-Moragues & Ojeda (2010)	*Erica coccinea*	Angiosperma	Unknown	per/itero	nuc	y	n
Ramdhani *et al.* (2010)	*Schotia* spp.	Angiosperma	Unknown	per/itero	cp/nuc	(n)‡	y
McDonald & Daniels (2012)	*Peripatopsis capensis*	Euonychophora	Outcrossing	per/itero	mt/nuc	y	(y)‡
Price *et al.* (2007)	*Platypleura stridula*	Insecta	Outcrossing	ann/semel	mt/nuc	y	(y)‡
Downie & Williams (2009)	*Porthetes hispidus*	Insecta	Outcrossing	unknown	mt/nuc	n	y
Smit *et al.* (2007)	*Elephantulus edwardii*	Mammalia	Outcrossing	unknown	mt	y	y
Van Hooft *et al.* (2002)§	*Syncerus caffer*	Mammalia	Outcrossing	per/itero	mt/nuc	y	y
Daniels *et al.* (2007)	*Chersina angulata*	Sauropsida	Outcrossing	per/itero	mt	y	y
Swart *et al.* (2009)	*Agama atra*	Sauropsida	Outcrossing	per/itero	mt/nuc	y	y
Tolley *et al.* (2006)	*Bradypodion* spp.	Sauropsida	Outcrossing	unknown	mt	y	y
Portik *et al.* (2011)	*Trachylepis sulcata*	Sauropsida	Outcrossing	per/itero	mt/nuc	y	y
Willows-Munro & Matthee (2011)	*Myosorex varius*	Soricomorpha	Outcrossing	per/itero	mt/nuc	y	y
Heideman *et al.* (2011)	*Scelotes* spp.	Squamata	Outcrossing	unknown	mt/nuc	y	not tested

*The presence (y, yes) or absence (n, no) of phylogeographical structure in the reviewed studies was inferred from phylogenetic trees or networks including commonly used support statistics (Posada & Crandall, 2001) or from analysis of molecular variance (AMOVA; Excoffier & Lischer, 2010) or related spatially explicit types of *F*-statistics.

†The presence (y, yes) or absence (n, no) of demographic/range expansions in the reviewed studies was inferred from mismatch analysis (Rogers & Harpending, 1992), Bayesian skyline plots of effective population size (N_e) through time (Heled & Drummond, 2008), properties of allele frequency spectra, or spatial trends in genetic diversity.

‡Parentheses indicate results regarded as tentative.

§The geographical range covered by this study on Cape buffalo greatly exceeds the Greater Cape Floristic Region.

the results with four variables that may help to interpret the answers to these questions: taxonomic group, mating system, generation time and genome sampled (Table 1). These explanatory variables were chosen because of their known impact on genetic structure in animals and plants (Hamrick & Godt, 1996; Morjan & Rieseberg, 2004).

Despite the limited number of available studies (arguably a surprising result in its own right), two patterns become readily apparent from the data. First, there are clear indications for both phylogeographical structure *and* recent range dynamics, and these results hold regardless whether genealogies were sampled from nuclear or from cytoplasmic genomes (Table 1). Second, there is more variation in the results for plants than for animals, despite the small number of studies of the former. The limited amount of phylogeographical information currently available points to open questions and emerging issues requiring the attention of biogeographers and phylogeographers with interests in hotspots, and we shall outline these below.

At first sight, widespread phylogeographical structure is consistent with the idea of species divergence and persistence due to the pronounced topographic and ecological heterogeneity frequently encountered in the Cape (Goldblatt & Manning, 2002; Linder, 2003; Schnitzler *et al.*, 2011). Thus, phylogeographical patterns *within* species are reminiscent of patterns of diversity seen at the community level: high species turnover between habitats (beta diversity) is known to be a crucial determinant of species richness in the Cape (Simmons & Cowling, 1996; Linder, 2003). More data on species diversity and phylogeographical structure in different organismal groups are needed to test whether both levels of diversity are affected by the same drivers. Note that patchy and mosaic-like habitats are not exclusive to plants, but are also encountered by species of animals living in the Cape (e.g. reptiles; Tolley *et al.*, 2006; Swart *et al.*, 2009; Portik *et al.*, 2011). The frequently found signature of demographic/range expansions, on the other hand (Table 1), is also consistent with the impact of Pleistocene climatic changes on the range dynamics of extant species, as hypothesized by Midgley *et al.* (2003, 2005). Given the peculiar history and nature of the Cape, what then are the actual *drivers* of current species divergence and persistence, especially for the extraordinary diversity of plants inhabiting this hotspot?

From the viewpoint of population genetics, the drivers of divergence and persistence may be either neutral or associated with natural selection. In a *neutral* scenario, mutation and drift during periods of geographical isolation facilitate species divergence and reproductive isolation, whereas neutral gene flow keeps populations of the same species together as cohesive units (Conner & Hartl, 2004; Morjan & Rieseberg, 2004). Alternatively, these processes may be driven by *natural selection*. Differential adaptation can trigger divergence and contribute to isolating mechanisms that maintain newly arisen, distinct forms in sympatry or parapatry (Wu & Ting, 2004; Gavrilets & Vose, 2005; Smadja & Butlin, 2011), while selective sweeps across habitat mosaics can maintain species as cohesive units despite restrictions in dispersal (Morjan & Rieseberg, 2004; Kane & Rieseberg, 2007). Distinguishing between these different mechanisms (neutral versus selective) matters greatly, because their outcomes differ strongly in terms of the speed of divergence and chance of persistence of divergent forms (Morjan & Rieseberg, 2004; Smadja & Butlin, 2011).

It helps to keep in mind that the mechanisms mediating cohesion (and thus persistence) of species are closely related to those maintaining reproductive isolation between them (Rieseberg & Burke, 2001). This is the case because the opportunity for interspecific gene exchange depends crucially on levels of gene flow within species (Petit & Excoffier, 2009), which provides yet another motivation for merging within- and between-species approaches to biogeographical inference. Importantly, the footprint of selection between conspecific populations will often be limited to genomic regions surrounding very specific sets of functionally important genes or their regulatory elements (Schlötterer, 2003; Kane & Rieseberg, 2007; Stinchcombe & Hoekstra, 2008), whereas differentiated genome regions appear to grow quite quickly in speciating populations (Feder & Nosil, 2010). Thus, phylogeographers interested in the mechanisms of diversification and persistence will need to greatly expand their *genomic coverage* to discover and delimit the genome regions important in the process, in addition to extending taxonomic coverage. Fortunately, the genomic tools to enable this are now becoming available.

STRETCHING THE LIMITS OF PHYLOGEOGRAPHY: THE POWER AND PROMISE OF NEXT-GENERATION SEQUENCING

So-called next-generation sequencing (NGS) approaches to the ultra high throughput sequencing of DNA are currently transforming the ways in which phylogeographers are able to track the dynamics of genetic diversity in space and time. The power of these approaches lies in their ability to yield hundreds of millions of short (at present typically 50–200 DNA bases each) sequence reads per run (Metzker, 2010), in contrast to conventional Sanger sequencing that typically yields only a few hundreds of reads. The unprecedented throughput of NGS approaches allows phylogeographers to discover tens to hundreds of thousands of molecular genetic (DNA) markers in the genomes of non-model species and simultaneously type them in wild populations by direct re-sequencing. Such population level re-sequencing may be done at the level of entire genomes (Rubin *et al.*, 2010; Turner *et al.*, 2010) or, more affordably for larger sample sizes, in partial genomic scans revealing 'only' a few tens of thousands of genetic markers in each individual animal or plant. Many variations to this theme exist (Parchman *et al.*, 2012), but arguably the most relevant of these to phylogeographers at the present time is restriction-site-associated DNA sequencing (RAD-seq; Baird *et al.*, 2008; Emerson *et al.*, 2010; Hohenlohe *et al.*, 2010).

The nature and work flow of RAD-seq have been described in detail elsewhere (e.g. Baird *et al.*, 2008; Hohenlohe *et al.*, 2010; Amores *et al.*, 2011; Davey & Blaxter, 2010; Etter *et al.*, 2011) and need not be reiterated here. The main issues in applying NGS technology to phylogeographical research are: (1) How can the massive amounts of data from population-level NGS, such as RAD-seq, be used to address the evolutionary drivers of species diversification? (2) How can these approaches be used across multiple species, effectively merging *between-* and *within-*species approaches to biogeographical study, particularly in non-model taxa without completely sequenced genomes? We shall use the Cape hotspot example and NGS data for Cape taxa to illustrate these issues.

PUTTING NGS TO WORK IN BIOGEOGRAPHICAL STUDIES OF SPECIES-RICH ENVIRONMENTS

We suspect that perhaps the greatest impact of NGS in biogeography and phylogeography will lie in its ability to help identify and pinpoint the evolutionary drivers of species divergence, expansion, and persistence, arguably the 'holy grail' of current biogeography and evolutionary genetics (Morjan & Rieseberg, 2004; Gavrilets & Vose, 2005; Klopfstein *et al.*, 2006; Excoffier *et al.*, 2009; Smadja & Butlin, 2011). NGS approaches hold immense potential for this purpose, because the 'genomic footprints' of neutral population divergence and demographic changes (e.g. population expansions and contractions) differ from those caused by natural selection affecting very particular sets of functionally important genes and their surrounding DNA (Schlötterer, 2003; Klopfstein *et al.*, 2006; Stinchcombe & Hoekstra, 2008; Excoffier *et al.*, 2009; see above for *why* the distinction between neutral and selective forces matters). In the case of the Cape hotspot example, particularly worthwhile goals would be: (1) to test whether and how often divergence of populations and species within major radiations is driven by slow neutral processes in geographical isolation versus rapid ecological speciation due to strong divergent selection, and (2) to test whether secondary contact and interspecific gene flow in sympatry or parapatry triggered by range dynamics (Table 1) constrains or facilitates adaptive evolution and speciation (for a rationale see Hewitt, 2000; de Carvalho *et al.*, 2010).

The use of NGS to address these important issues is likely to bring novel analytical opportunities and challenges. Phylogeographical reconstruction has traditionally relied on *neutral* genetic markers (Avise, 2009), but NGS approaches offer considerable power to uncover regions of the genome that respond to natural (e.g. divergent ecological) selection (Hohenlohe *et al.*, 2010; Rubin *et al.*, 2010; Turner *et al.*, 2010). Of course, non-neutral 'outlier loci' may also be removed from the data set to improve inferences about population history (Stinchcombe & Hoekstra, 2008), or the very large number of markers generated may be assumed to approximate a neutral distribution (Emerson *et al.*, 2010). Nevertheless, population geneticists have long suspected that marker loci affected by selection may in fact be efficient tools for estimating gene flow between locally adapted populations (Lenormand *et al.*, 1998; Guichoux *et al.*, 2013), analogous to the estimation of migration and selection along clines (Barton & Hewitt, 1985). Likewise, markers undergoing range-wide selective sweeps are predicted to be highly useful for delimiting related species, because their rapid spread *within* species strengthens their resistance to introgression from related taxa (Petit & Excoffier, 2009). Thus, we anticipate that NGS approaches will transform the way in which biogeographers make use of DNA genealogies to address questions on the origin and maintenance of biological diversity.

It would be premature to forecast the myriad creative ways biogeographers will find to put NGS approaches to work for tackling these important issues, once provided with the necessary bioinformatic infrastructure and 'know-how'. One particular challenge faced by biogeographers and phylogeographers at the present time is the need to analyse sets of *multiple* species, usually in the *absence* of fully assembled reference genomes. This issue is particularly pronounced in biodiversity hotspots such as the Cape, which is dominated by unusual organismal families (Linder *et al.*, 2010), because these groups of taxa have not been 'on the radar' of genomics research in the past.

DE NOVO APPROACHES: NGS WITHOUT THE SAFETY NET OF A COMPLETELY SEQUENCED GENOME

Perhaps two of the greatest challenges in current NGS-based biogeography and phylogeography are the application of NGS approaches (e.g. RAD-seq) to taxa without existing reference genomes and their use in comparative studies involving multiple species. A recurrent theme across all current NGS approaches, including RAD-seq, is that these two issues are tightly connected because of the way polymorphisms are called (recognized and identified) in populations (for details of the bioinformatic work flow see Hohenlohe *et al.*, 2010; Amores *et al.*, 2011; Catchen *et al.*, 2011; Etter *et al.*, 2011).

Typically, single nucleotide polymorphisms (SNPs) in genotyping-by-sequencing approaches are identified either after reference mapping of short DNA sequence reads against a reference genome (e.g. Hohenlohe *et al.*, 2010) or, when no reference genome is available, against *de novo* assembled clusters or 'stacks' (Emerson *et al.*, 2010; Catchen *et al.*, 2011) of short DNA sequence reads from many individuals. Either approach requires definition of a sequence mismatch threshold (e.g. 95% sequence identity) during sequence alignment, to make sure that similar DNA fragments recovered from populations are indeed allelic (= homologous variants of the same genetic locus). *De novo* assembly of sequence clusters is the method of choice in non-model species with no available reference genome, and the robustness of the approach has been thoroughly evaluated by computer simulations and genetic mapping (Amores *et al.*, 2011; Catchen *et al.*, 2011). The possibility to build dense genetic linkage maps of RAD-seq markers in non-model species (e.g. Amores *et al.*, 2011) is an important aspect for biogeographical studies, because linkage mapping yields recombination distances that are highly useful for interpreting selection signatures and for parameterizing historical demographic models. However, one practical issue of the *de novo* approach requires special attention here.

The process of cluster building easily results in the 'loss' of many polymorphisms, because a stringent maximum mismatch threshold is required during clustering of sequence reads to avoid the inclusion of paralogous (non-allelic variants) variants, especially when working with taxa with complex, recently duplicated genomes (Hohenlohe *et al.*, 2011). This issue is greatly aggravated when *multiple divergent species* are present in the data set. Sequence reads derived from different species will often be too divergent to be 'recognized' as members of the same cluster (= genetic locus) in spite of shared ancestry, ultimately resulting in the loss of detected polymorphic sites or in the reconstruction of separate loci ('allele splitting'), depending on the filtering criteria used. Genetic divergence between species may also manifest itself in the loss of endonuclease restriction sites (which play a crucial role in genotyping-by-resequencing approaches such as RAD-seq; Baird *et al.*, 2008) or in the loss of entire stretches of DNA. A bioinformatic analysis of pilot data from RAD-seq of an important plant radiation at the Cape may serve to illustrate these issues.

REMOVING THE NET: TOWARDS *DE NOVO* SNP CALLING IN MULTISPECIES DATA SETS

To illustrate our point, we generated and analysed RAD-seq data for three southern African species of the Restionaceae family (restiads): *Restio capensis* (L.) H. P. Linder & C. R. Hardy, *Restio triticeus* Rottb., and *Hypodiscus aristatus* (Thunb.) C. Krauss (see Appendix S1 in Supporting Information). These species originate from highly divergent clades within the Restionaceae family (Hardy *et al.*, 2008; Linder & Hardy, 2010), thus providing a conservative setting for testing the applicability of RAD-seq in multispecies studies. We followed essentially the same laboratory procedures recently employed for a pair of highly divergent, hybridizing

Northern Hemisphere tree species by our group (Stölting *et al.*, 2012). RAD sequence reads were assembled by *de novo* clustering because no reference genome is currently available for Restionaceae. The widely used Stacks software (Catchen *et al.*, 2011) was used to build sequence clusters in each of the three restiad species using commonly used criteria (e.g. a maximum of three mismatches; see legend of Fig. 1). For the purpose of the present contribution, we used clusters shared by two individuals sampled from each species and examined sharing of these clusters between taxa, using custom perl scripts and open source software (Chen & Boutros, 2012) (Fig. 1). The individuals were picked randomly from 300 RAD-sequenced plants currently used for population genomic studies in our laboratory.

The Venn diagram indicates the great loss of sequence clusters when combining information from these three divergent taxa (Fig. 1). Nevertheless, more than 5600 unique sequence clusters were shared across all three species, thus providing a rich source of DNA sequences that can be mined for SNP diversity (quantifying that diversity requires the analysis of larger samples and is beyond the scope of the present contribution). The diagram is also suggestive of a relationship between levels of cluster sharing and evolutionary divergence, with the congeneric taxa *Restio capensis* and *R. triticeus* sharing a much greater number of unique clusters than *Hypodiscus aristatus* with either of these. Note that the split between the genus *Restio* and the group containing *Hypodiscus* is the most basal one in a comprehensive molecular phylogeny of the Restionaceae family (Hardy *et al.*, 2008), the estimated divergence time of the entire radiation being > 35 Ma (Linder, 2005). Our results illustrate the
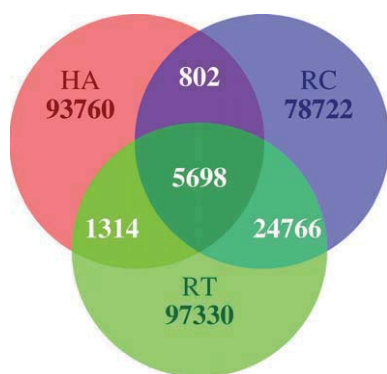

Figure 1 Venn diagram illustrating the potentials and limits of studying clusters of restriction-site-associated DNA (RAD) sequence reads across three highly divergent species of the South African Restionaceae: *Restio capensis* (RC), *Restio triticeus* (RT) and *Hypodiscus aristatus* (HA). The black numbers are the numbers of unique RAD clusters shared by two individuals examined for each species. The white numbers are the numbers of unique RAD clusters shared by different combinations of species. Clustering ('stack building') was carried out with the Stacks software (Catchen *et al.*, 2011), allowing for a maximum of three mismatches between reads of the same cluster and a minimum coverage of two copies per cluster (see text for details).

potentials and limits of NGS approaches for multispecies biogeographical studies of large radiations such as that of Restionaceae in the Cape, in particular, the amount of putatively homologous sequence information (cluster sharing) uncovered across different evolutionary time-scales (Fig. 1). There is great potential for improvement of all these issues, as existing cluster building software advances (e.g. Catchen *et al.*, 2011) and the read length of NGS approaches increases.

Another currently open issue of multispecies NGS approaches in biogeography is that divergence will vary not only between *different pairs of species* within radiations of animals and plants, but also *along the genomes* of any given set of species. This is the case because the genomes of plants and animals remain 'porous' for up to millions of years following the onset of speciation (Wu & Ting, 2004), and the combined action of selection and drift creates considerable variation in the 'genomic landscape of divergence' until whole-genome isolation is complete (Feder & Nosil, 2010). This implies that the efficiency of *de novo* clustering will vary greatly along genomes, with far worse performance of *de novo* SNP calling methods in hyperdivergent regions of the genome. Here, many sequence clusters will simply drop out from the analysis because they exceed the a priori defined mismatch threshold (i.e. homology criterion) used during cluster building, potentially leading to highly uneven coverage of genomes in multispecies studies. In a related contribution (Lexer & Stölting, 2012), we touch on these open issues based on exemplary RAD-seq data for two highly divergent (several million years) hybridizing plant species with a completely sequenced reference genome. We suspect that tackling these issues in multispecies biogeographical studies will require analytical approaches that account for variation in the depth of genealogies across the genome (which may be estimated from multispecies data) and experimental manipulation of sequence alignment (i.e. mismatch) and sequencing depth (i.e. coverage) thresholds in each single case, accompanied by extensive simulation studies of the relevant parameter space. Recent work on divergent adaptation, gene flow and hybrid speciation in Lake Victoria cichlid fishes (Keller *et al.*, 2012) demonstrates how much can be learned by applying RAD sequencing (or similar NGS approaches) to multispecies studies of non-model taxa without sequenced reference genomes, and similar examples from plants are forthcoming.

CONCLUSIONS AND OUTLOOK

Understanding the mechanisms and drivers of species divergence, expansion, and persistence in spatially structured, species-rich environments of high conservation value (i.e. biodiversity hotspots) will require concerted efforts to resolve the phylogeography and genetics of diverging populations in a much greater number of taxa and with much greater genomic coverage than previously achieved. We have demonstrated this need using the Cape of southern Africa as an

example, but our own experience shows that the same applies to other Southern Hemisphere biodiversity hotspots (Palma-Silva *et al.*, 2009, 2011). The NGS technologies to facilitate scaling-up this process are now available and increasingly affordable. Moreover, suitable bioinformatic approaches are in place to handle the enormous amounts of sequence data gathered in each single species. A challenge ahead will be the development of bioinformatic tools able to efficiently handle NGS data involving multiple genetically divergent species. This will be particularly important for research on plants. The notorious complexity and evolutionary 'fluidity' of plant genomes challenge the reliable assignment of orthologous (= allelic) genetic variants, necessary to trace genealogies within and among diverging populations of radiating species.

REFERENCES

Amores, A., Catchen, J., Ferrara, A., Fontenot, Q. & Postlethwait, J.H. (2011) Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics*, **188**, 799–808.

Antonelli, A., Nylander, J.A.A., Persson, C. & Sanmartín, I. (2009) Tracing the impact of the Andean uplift on Neotropical plant evolution. *Proceedings of the National Academy of Sciences USA*, **106**, 9749–9754.

Arbogast, B.S. & Kenagy, G.J. (2001) Comparative phylogeography as an integrative approach to historical biogeography. *Journal of Biogeography*, **28**, 819–825.

Avise, J.C. (2009) Phylogeography: retrospect and prospect. *Journal of Biogeography*, **36**, 3–15.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. & Johnson, E.A. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376.

Barton, N.H. & Hewitt, G.M. (1985) Analysis of hybrid zones. *Annual Review of Ecology, Evolution, and Systematics*, **16**, 113–148.

Beheregaray, L.B. (2008) Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology*, **17**, 3754–3774.

Bergh, N.G., Hedderson, T.A., Linder, H.P. & Bond, W.J. (2007) Palaeoclimate-induced range shifts may explain current patterns of spatial genetic variation in renosterbos (*Elytropappus rhinocerotis*, Asteraceae). *Taxon*, **56**, 393–408.

Cannon, C.H. & Manos, P.S. (2003) Phylogeography of the Southeast Asian stone oaks (*Lithocarpus*). *Journal of Biogeography*, **30**, 211–226.

de Carvalho, D., Ingvarsson, P.K., Joseph, J., Suter, L., Sedivy, C., Macaya-Sanz, D., Cottrell, J., Heinze, B., Schanzer, S. & Lexer, C. (2010) Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree. *Molecular Ecology*, **19**, 1638–1650.

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

Cavender-Bares, J., Kozak, K.H., Fine, P.V.A. & Kembel, S.W. (2009) The merging of community ecology and phylogenetic biology. *Ecology Letters*, **12**, 693–715.

Chen, H. & Boutros, P.C. (2012) VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*, **12**, 35.

Conner, J.K. & Hartl, D.L. (2004) *A primer of ecological genetics*. Sinauer Associates, Sunderland, MA.

Costa, L.P. (2003) The historical bridge between the Amazon and the Atlantic Forest of Brazil: a study of molecular phylogeography with small mammals. *Journal of Biogeography*, **30**, 71–86.

Daniels, S.R., Hofmeyr, M.D., Henen, B.T. & Crandall, K.A. (2007) Living with the genetic signature of Miocene induced change: evidence from the phylogeographic structure of the endemic angulate tortoise *Chersina angulata*. *Molecular Phylogenetics and Evolution*, **45**, 915–926.

Davey, J.L. & Blaxter, M.W. (2010) RADSeq: next-generation population genetics. *Briefings in Functional Genomics*, **9**, 416–423.

Davis, M.B. & Shaw, R.G. (2001) Range shifts and adaptive responses to Quaternary climate change. *Science*, **292**, 673–679.

Dawson, M.N. (2001) Phylogeography in coastal marine animals: a solution from California? *Journal of Biogeography*, **28**, 723–736.

Downie, D.A. & Williams, J.G. (2009) Population structure of *Porthetes hispidus* (Coleoptera: Curculionidae), a pollinator of the African cycad *Encephalartos friderici-guilielmi*. *Annals of the Entomological Society of America*, **102**, 1126–1134.

Emerson, B.C. & Gillespie, R.G. (2008) Phylogenetic analysis of community assembly and structure over space and time. *Trends in Ecology and Evolution*, **23**, 619–630.

Emerson, K.J., Merz, C.R., Catchen, J.M., Hohenlohe, P.A., Cresko, W.A., Bradshaw, W.E. & Holzapfel, C.M. (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences USA*, **107**, 16196–16200.

Etter, P.D., Preston, J.L., Bassham, S., Cresko, W.A. & Johnson, E.A. (2011) Local *de novo* assembly of RAD paired-end contigs using short sequencing reads. *PLoS ONE*, **6**, e18561.

Excoffier, L. (2004) Special Issue: Analytical methods in phylogeography and genetic structure. *Molecular Ecology*, **13**, 727.

Excoffier, L. & Lischer, H.E.L. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Excoffier, L., Foll, M. & Petit, R.J. (2009) Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 481–501.

Feder, J.L. & Nosil, P. (2010) The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*, **64**, 1729–1747.

Frankham, R., Ballou, J.D. & Briscoe, D.A. (2004) *A primer of conservation genetics*. Cambridge University Press, Cambridge.

Futuyma, D.J. (2009) *Evolution*. Sinauer, Sunderland, MA.

Gavrilets, S. & Vose, A. (2005) Dynamic patterns of adaptive radiation. *Proceedings of the National Academy of Sciences USA*, **102**, 18040–18045.

Goldblatt, P. & Manning, J.C. (2002) Plant diversity of the Cape region of southern Africa. *Annals of the Missouri Botanical Gardens*, **89**, 281–302.

Guichoux, E., Garnier-Géré, P., Lagache, L., Lang, T., Boury, C. & Petit, R.J. (2013) Outlier loci highlight the direction of introgression in oaks. *Molecular Ecology*, **22**, 450–462.

Hamrick, J.L. & Godt, M.J.W. (1996) Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **351**, 1291–1298.

Hardy, C.R., Moline, P. & Linder, H.P. (2008) A phylogeny for the African Restionaceae and new perspectives on morphology's role in generating complete species phylogenies for large clades. *International Journal of Plant Sciences*, **169**, 377–390.

Heideman, N.J.L., Mulcahy, D.G., Sites, J.W., Jr, Hendricks, M.G.J. & Daniels, S.R. (2011) Cryptic diversity and morphological convergence in threatened species of fossorial skinks in the genus *Scelotes* (Squamata: Scincidae) from the Western Cape Coast of South Africa: implications for species boundaries, digit reduction and conservation. *Molecular Phylogenetics and Evolution*, **61**, 823–833.

Heled, J. & Drummond, A.J. (2008) Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology*, **8**, 289.

Hewitt, G. (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Höglund, J. (2009) *Evolutionary conservation genetics*. Oxford University Press, Oxford.

Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A. & Cresko, W.A. (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.

Hohenlohe, P.A., Amish, S.J., Catchen, J.M., Allendorf, F.W. & Luikart, G. (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, **11**, 117–122.

Jablonski, D. (2008) Biotic interactions and macroevolution: extensions and mismatches across scales and levels. *Evolution*, **62**, 715–739.

Kane, N.C. & Rieseberg, L.H. (2007) Selective sweeps reveal candidate genes for adaptation to drought and salt tolerance in common sunflower, *Helianthus annuus*. *Genetics*, **175**, 1823–1834.

Keller, I., Wagner, C.E., Greuter, L., Mwaiko, S., Selz, O.M., Sivasundar, A., Wittwer, S. & Seehausen, O. (2012) Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology*, doi:10.1111/mec.12083.

Klopfstein, S., Currat, M. & Excoffier, L. (2006) The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, **23**, 482–490.

Ladle, R.J. & Whittaker, R.J. (eds) (2011) *Conservation biogeography*. Wiley-Blackwell, Chichester, UK.

Lenormand, T., Guillemaud, T., Bourguet, D. & Raymond, M. (1998) Evaluating gene flow using selected markers: a case study. *Genetics*, **149**, 1383–1392.

Lexer, C. & Stölting, K. (2012) Whole genome sequencing (WGS) meets biogeography and shows that genomic selection in forest trees is feasible. *New Phytologist*, **196**, 652–654.

Linder, H.P. (2003) The radiation of the Cape flora, southern Africa. *Biological Reviews*, **78**, 597–638.

Linder, H.P. (2005) Evolution of diversity: the Cape flora. *Trends in Plant Science*, **11**, 536–541.

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

Linder, K.P. & Hardy, C.R. (2010) A generic classification of the Restioneae (Restionaceae), southern Africa. *Bothalia*, **40**, 1–35.

Linder, H.P., Johnson, S.D., Kuhlmann, M., Matthee, C.A., Nyffeler, R. & Swartz, E.R. (2010) Biotic diversity in the Southern African winter-rainfall region. *Current Opinion in Environmental Sustainability*, **2**, 109–116.

McDonald, D.E. & Daniels, S.R. (2012) Phylogeography of the Cape velvet worm (Onychophora: *Peripatopsis capensis*) reveals the impact of Pliocene/Pleistocene climatic oscillations on Afromontane forest in the Western Cape, South Africa. *Journal of Evolutionary Biology*, **25**, 824–835.

Metzker, M.L. (2010) Sequencing technologies – the next generation. *Nature Reviews Genetics*, **11**, 31–46.

Midgley, G.F., Hannah, L., Millar, D., Thuiller, W. & Booth, A. (2003) Developing regional and species-level assessments of climate change impacts on biodiversity in the Cape Floristic Region. *Biological Conservation*, **112**, 87–97.

Midgley, G.F., Reeves, G. & Klak, C. (2005) Late Tertiary and Quaternary climate change and centres of endemism in the southern African flora. *Phylogeny and conservation*

(ed. by A. Purvis, J.L. Gittleman and T. Brooks), pp. 230–242. Cambridge University Press, Cambridge.

Mittermeier, R.A., Gil, P.R., Hoffman, M., Pilgrim, J., Brooks, T., Mittermeier, C.G., Lamoreux, J. & da Fonseca, G.A.B. (2005) *Hotspots revisited: Earth's biologically richest and most endangered terrestrial ecoregions*. University of Chicago Press, Chicago.

Morjan, C.L. & Rieseberg, L.H. (2004) How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology*, **13**, 1341–1356.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

Palma-Silva, C., Lexer, C., Paggi, G.M., Barbara, T., Bered, F. & Bodanese-Zanettini, M.H. (2009) Range-wide patterns of nuclear and chloroplast DNA diversity in *Vriesea gigantea* (Bromeliaceae), a neotropical forest species. *Heredity*, **103**, 503–512.

Palma-Silva, C., Wendt, T., Pinheiro, F., Barbara, T., Fay, M.F., Cozzolino, S. & Lexer, C. (2011) Sympatric bromeliad species (*Pitcairnia* spp.) facilitate tests of mechanisms involved in species cohesion and reproductive isolation in Neotropical inselbergs. *Molecular Ecology*, **20**, 3185–3201.

Parchman, T.L., Gompert, Z., Mudge, J., Schilkey, F.D., Benkman, C.W. & Buerkle, C.A. (2012) Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, **21**, 2991–3005.

Petit, R.J. & Excoffier, L. (2009) Gene flow and species delimitation. *Trends in Ecology and Evolution*, **24**, 386–393.

Pinheiro, F., de Barros, F., Palma-Silva, C., Fay, M.F., Lexer, C. & Cozzolino, S. (2011) Phylogeography and genetic differentiation along the distributional range of the orchid *Epidendrum fulgens*: a Neotropical coastal species not restricted to glacial refugia. *Journal of Biogeography*, **38**, 1923–1935.

Portik, D.M., Bauer, A.M. & Jackman, T.R. (2011) Bridging the gap: western rock skinks (*Trachylepis sulcata*) have a short history in South Africa. *Molecular Ecology*, **20**, 1744–1758.

Posada, D. & Crandall, K.A. (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology and Evolution*, **16**, 37–45.

Price, B.W., Barker, N.P. & Villet, M.H. (2007) Patterns and processes underlying evolutionary significant units in the *Platypleura stridula* L. species complex (Hemiptera: Cicadidae) in the Cape Floristic Region, South Africa. *Molecular Ecology*, **16**, 2574–2588.

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

Ramdhani, S., Cowling, R.M. & Barker, N.P. (2010) Phylogeography of *Schotia* (Fabaceae): recent evolutionary processes in an ancient thicket biome lineage. *International Journal of Plant Sciences*, **171**, 626–640.

Rieseberg, L.H. & Burke, J.M. (2001) A genic view of species integration. *Journal of Evolutionary Biology*, **14**, 883–886.

Rogers, A.R. & Harpending, H. (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.

Rubin, C.J., Zody, M.C., Eriksson, J., Meadows, J.R.S., Sherwood, E., Webster, M.T., Jiang, L., Ingman, M., Sharpe, T., Ka, S., Hallbook, F., Besnier, F., Carlborg, O., Bed'hom, B., Tixier-Boichard, M., Jensen, P., Siegel, P., Lindblad-Toh, K. & Andersson, L. (2010) Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature*, **464**, 587–591.

Rymer, P.D., Manning, J.C., Goldblatt, P., Powell, M.P. & Savolainen, V. (2010) Evidence of recent and continuous speciation in a biodiversity hotspot: a population genetic approach in southern African gladioli (*Gladiolus*; Iridaceae). *Molecular Ecology*, **19**, 4765–4782.

Salamin, N., Wueest, R.O., Lavergne, S., Thuiller, W. & Pearman, P.B. (2010) Assessing rapid evolution in a changing environment. *Trends in Ecology and Evolution*, **25**, 692–698.

Schlötterer, C. (2003) Hitchhiking mapping – functional genomics from the population genetics perspective. *Trends in Genetics*, **19**, 32–38.

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

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

Simmons, M.T. & Cowling, R.M. (1996) Why is the Cape Peninsula so rich in plant species? An analysis of the independent diversity components. *Biodiversity and Conservation*, **5**, 551–573.

Smadja, C.M. & Butlin, R.K. (2011) A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology*, **20**, 5123–40.

Smit, H.A., Robinson, T.J. & Van Vuuren, B.J. (2007) Coalescence methods reveal the impact of vicariance on the spatial genetic structure of *Elephantulus edwardii* (Afrotheria, Macroscelidea). *Molecular Ecology*, **16**, 2680–2692.

Stinchcombe, J.R. & Hoekstra, H.E. (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*, **100**, 158–170.

Stölting, K.N., Nipper, R., Lindtke, D., Caseys, C., Waeber, S., Castiglione, S. & Lexer, C. (2012) Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Molecular Ecology*, doi:10.1111/mec.12011.

Swart, B.L., Tolley, K.A. & Matthee, C.A. (2009) Climate change drives speciation in the southern rock agama (*Agama atra*) in the Cape Floristic Region, South Africa. *Journal of Biogeography*, **36**, 78–87.

Tolley, K.A., Burger, M., Turner, A.A. & Matthee, C.A. (2006) Biogeographical patterns and phylogeography of

dwarf chameleons (*Bradypodion*) in an African biodiversity hotspot. *Molecular Ecology*, **15**, 781–793.

Tolley, K.A., Braae, A. & Cunningham, M. (2010) Phylogeography of the Clicking Stream Frog *Strongylopus grayii* (Anura, Pyxicephalidae) reveals cryptic divergence across climatic zones in an abundant and widespread taxon. *African Journal of Herpetology*, **59**, 17–32.

Turner, T.L., Bourne, E.C., Von Wettberg, E.J., Hu, T.T. & Nuzhdin, S.V. (2010) Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature Genetics*, **42**, 260–263.

Valente, L.M., Reeves, G., Schnitzler, J., Mason, I.P., Fay, M.F., Rebelo, T.G., Chase, M.W. & Barraclough, T.G. (2010) Diversification of the African genus *Protea* (Proteaceae) in the Cape biodiversity hotspot and beyond: equal rates in different biomes. *Evolution*, **64**, 745–759.

Van Hooft, W.F., Groen, A.F. & Prins, H.T. (2002) Phylogeography of the African buffalo based on mitochondrial and Y-chromosomal loci: Pleistocene origin and population expansion of the Cape buffalo subspecies. *Molecular Ecology*, **11**, 267–279.

Verboom, G.A., Archibald, J.K., Bakker, F.T., Bellstedt, D.U., Conrad, F., Dreyer, L.L., Forest, F., Galley, C., Goldblatt, P., Henning, J.F., Mummenhoff, K., Linder, H.P., Muasya, A.M., Oberlander, K.C., Savolainen, V., Snijman, D.A., van der Niet, T. & Nowell, T.L. (2009) Origin and diversification of the Greater Cape flora: ancient species repository, hot-bed of recent radiation, or both? *Molecular Phylogenetics and Evolution*, **51**, 44–53.

Wiens, J.J. & Donoghue, M.J. (2004) Historical biogeography, ecology and species richness. *Trends in Ecology and Evolution*, **19**, 639–644.

Willows-Munro, S. & Matthee, C.A. (2011) Linking lineage diversification to climate and habitat heterogeneity: phylogeography of the southern African shrew *Myosorex varius*. *Journal of Biogeography*, **38**, 1976–1991.

Wu, C.I. & Ting, C.T. (2004) Genes and speciation. *Nature Reviews Genetics*, **5**, 114–122.

SUPPORTING INFORMATION

Appendix S1 RAD sequencing results for South African Restionaceae (restiad) species discussed in the main text.