Dimensions of Biodiversity: Parallel Evolutionary radiations in *Protea* and *Pelargonium* in the Greater Cape Floristic Region

**Intellectual Merit:** Current approaches to predicting changes in community structure and function are based largely on static assessments. They fail to consider relationships between species’ attributes that affect their function in a given environment (*functional traits*), genetic variation underlying those functional traits, and the dynamics of the communities of which they are part. Each of these relationships reflects the unique evolutionary history of the lineage in question, and lineages that have flourished through dramatic historical climate change are likely to harbor significant information about functional trait evolution that will shed new light on adaptation and evolutionary potential. We focus on *Protea* and *Pelargonium*, two South African plant genera that have undergone dramatic diversification since the Miocene during long-term aridification and dramatic shifts in climate. Southern Africa has higher plant diversity than most tropical regions, yet much of this diversity stems from radiations in just 30 major lineages, including *Protea* and *Pelargonium*. These genera span two biodiversity hot spots, the fynbos and succulent karoo biomes in the Greater Cape Floristic Region (GFCR). Our overall goal is to develop an understanding of functional trait variation in both genetic and evolutionary contexts in *Protea* and *Pelargonium*. This will allow us to predict the characteristics of communities in which they occur and the resilience of those communities to climate change. To achieve these goals, we will answer three sets of hierarchically nested questions:

I. How do functional traits vary within and between species? How are differences related to environmental variation (e.g., climate, soils, and fire)? What is the genetic basis of these differences?

II. What are the phylogenetic histories of functional trait change? How are the rates of trait evolution related to changes in niche parameters?

III. How are taxonomic and trait level diversity and abundance of focal taxa interrelated with community composition along environmental gradients? How are these patterns related to ecosystem resilience to fire?

To answer these questions, we will assess functional traits on nearly all species of *Protea* and *Pelargonium* found in the GFCR; for a subset of species we will examine detailed relationships between functional traits, photosynthesis, water use efficiency and leaf life span. We will identify the genetic basis of trait variation both through geographically extensive sampling within widely distributed species and through intensive population sampling in two hybrid species complexes using state-of-the-art next-generation sequencing technology to map variation in several thousand individual genomes. Functional trait evolution and its association with environmental variation will be assessed using highly resolved evolutionary histories inferred from high-coverage transcriptome sequences we will generate for nearly all species in both genera. We will evaluate rates of diversification of traits and parameters of species’ niches. Because trait evolution is best understood when viewed in the context of the community in which it evolved, we will also characterize the functional trait and taxonomic diversity of communities containing *Protea* and *Pelargonium* and use this information to infer community structure along environmental gradients. Ultimately, we will examine effects of taxonomic and trait diversity on ecosystem resilience in response to fire cycle interval, which strongly affects species and functional composition of communities in the region and which our ongoing research has found to be highly sensitive to projected climate changes.

**Broader Impacts:** We will build on our considerable past success as an integrated team with expertise in a wide range of ecological and botanical disciplines to expose post docs, grad students, undergrads and high school students to international research, including students actively recruited from underserved populations both in the US and South Africa. Through annual methods workshops for participants and hosting institutions, we will enhance scientific infrastructure via development Bayesian statistical approaches to large bio-spatial-temporal databases. We will actively enhance conservation infrastructure by working closely in all aspects of the project with the South African National Botanical Institute (SANBI) which has one of the best track records in the world in promoting biodiversity science, policy and conservation action.
RESULTS OF PRIOR SUPPORT:


PI-Hodges, co-PI Borevitz NSF EF-0412727: GEN-EN Genomics of Adaptation to the Biotic and Abiotic Environment in Aquilegia. 10/01/04-09/30/09. To develop genomic tools and genetic resources to establish an emerging model organism. Goals were largely met in the first 2 years of the proposal [1-3]. Microarrays for gene expression and SNP genotyping assays (Noutsos et al. in prep) have been developed with Nimblegen and Sequenom assays [4]. Virus Induced Gene Silencing has been developed [5] and genome sequencing is under way (JGI). Publications: 11 for grant; Lit Cited 1.

Co-PIs Schenk & Jones - NSF IOS-0641569: Collaborative Research: Resistance, Repair and Redundancy: Traits that protect shrubs against drought induced hydraulic failure 04/01/07-03/31/11 (1 yr no cost extension). Publications: 3 published (PNAS, American Journal of Botany, New Phytologist); 3 in prep; Lit Cited 1; Presentations: 3 at the Ecological Society of America 2008-2010, 2 at the Botanical Society of America (2010); Training: 1 technician at CS Fullerton; 2 grad students (1 female, 1 Hispanic male) at UConn; both locations more than 20 undergrads involved over the duration of the grant.


OVERVIEW:

Rapidly changing climate will drive dynamic responses at all levels of biodiversity from the traits of individuals to distributions of species and biomes. Current approaches to predicting such changes are based on static assessments [6]: e.g., what is a species’ current ecological envelope and where will that range of conditions occur in the future? But our ability to predict species and community responses will be enhanced by understanding their evolutionary past [7]. Clades that have evolved and flourished through dramatic historical climate change are likely to harbor significant information about the evolution of functional traits that can inform us about adaptation and evolutionary potential. Our project will center on two highly successful clades (the plant genera Protea and Pelargonium) that have diversified over the past 15 MY in response to (or in spite of) radical (at times rapid) changes in the climate of southern Africa. Our investigations will compare and contrast patterns of genetic and functional trait variation in these two genera at levels ranging from the intraspecific to the community.
The **taxonomic dimension** of our project will focus on “the comparison of the genetic and functional aspects of multiple taxa in a phylogenetic/evolutionary context” (from Dimensions FAQ) and will include phylogeographic analysis of selected species complexes and populations of interspecific hybrids. The **functional dimension** will examine the **associations between environmental parameters and functional plant traits**. The genetic dimension will use the burgeoning power of next-generation sequencing to identify genes ‘responsible for’ differences in traits within and among species. Ultimately, we will use our understanding of functional trait variation in *Protea* and *Pelargonium* to predict the characteristics of communities in which they occur and the resilience to climate change of their associated ecosystems.

**OUR FOCAL QUESTIONS:**

I. How do functional traits vary within and between species? How are differences related to environmental variation (e.g., climate, soils, and fire frequency)? What is the genetic basis of these differences?

II. What are the phylogenetic histories of functional trait change? Are the rates of trait evolution related to changes in niche parameters?

III. How are taxonomic and trait level diversity and abundance of focal taxa interrelated with community composition along environmental gradients? How are these patterns related to ecosystem resilience to fire?

**Why South Africa & the Cape Floristic Region?** We focus on the origin and maintenance of biodiversity in the Greater Cape Floristic Region (GCFR) of southwestern Africa, a region characterized by Mediterranean and arid climates with mostly winter rainfall [8] for several reasons.

(i) **The biological diversity of the GCFR rivals that of tropical rainforests** [9]; we have shown that the extraordinary level of β and γ diversity leads to plant biodiversity as high as anywhere in the world [10]. The GCFR also contains one of the highest proportions of species of global conservation concern [99]; extinction risk studies indicate that GCFR species are among the most vulnerable to climate change over the next 50 years [11].

(ii) **The GCFR is a compact, yet extremely environmentally diverse biogeographic region.** It contains two Biodiversity Hotspots, the Cape Floristic region (CFR) and the semi-arid Succulent Karoo (SK). Over 65% of GCFR plant species are endemic, an extraordinary level for a continental region. Wetter regions of the GCFR (40–320 cm mean annual precipitation [MAP]) include the fire-prone fynbos vegetation on nutrient-poor sandstone soils, characterized by sclerophyllous shrubs, and renosterveld on the more nutrient-rich shale soils, with fine-leaved, astereaceous shrubs and a multitude of geophytes. The arid SK (2–29 cm MAP) is dominated by low semi-succulent, leafy subshrubs (many in the Aizoaceae).

(iii) **Data on species distributions and biodiversity are extraordinarily rich:** the digital Protea Atlas Project (PAP: [12]); PRECIS data set [13]; presence/absence records for all 23,000+ native South African plant species (millions of herbarium sheets at ¼ degree grid scale; Acocks relevé data set: 50 years of community samples [14]; ~5000 geo-referenced relevé plots with full plant inventories. The ecology of the system is well known, with both habitat and community types fully digitally mapped [e.g., 15, 16].

(iv) **Background environmental data layers are extensive and detailed.** Climate data, available in grid form for the entire country [17], form the basis for understanding and explaining biodiversity patterns. Using these data and data from PAP, we developed a suite of hierarchical Bayesian models to predict the distribution of species and biodiversity based on data for presence-absence, abundance or presence-only [18, 19]. Silander and collaborators have assembled 50 years of weather station data for the region and are developing model-based protocols to yield an interpolated surface of daily climate data. Our collaborators at the University of Cape Town have produced a suite of 9 Global Climate Models for projected climate change, downscaled to weather stations in the region. We also have access to the complete record of fires in the protected areas (about 11,000km²) across the Western Cape Province.
Complementing climate and fire data are detailed digital data for edaphic traits, high resolution community types, endemicity, hydrography, satellite imagery and human land-use classes. In short, the GCFR offers the opportunity to answer questions about the generation and maintenance of regional biodiversity, not easily addressed in other hyper-diverse regions.

(v) The CFR has one of the world’s best developed conservation plans. The CFR “sets the global standard” for using scientific information in conservation planning/implementation [20], and there is good reason to believe that biodiversity science will be effectively implemented here for conservation. Members of our group, including senior personnel, are broadly linked with a number of GO and NGO conservation organizations across the region (e.g., South African National Biodiversity Institute [SANBI], CapeNature, CapeFlatsNature, SANParks, WWF, etc).

**PROTEA & PELARGONIUM:** Much of the overall floristic diversity of the GCFR has arisen through extensive radiations in about 30 clades [21]; at least 13 genera, including *Protea* and *Pelargonium*, have over 100 species [22]. These two genera have successfully radiated in response to historic climate change – in particular, the continuing aridification of southwestern Africa starting in the Miocene. From a community standpoint, these clades encompass much of the growth form diversity (excepting graminoids) of this region. In addition, both have local and international economic significance: Proteas - cut flowers; Pelargoniums - fragrances and pharmaceuticals; and both as germplasm for horticultural improvement.

**PROTEA** – Icon of the fynbos biome; Species diversity (~112; 62% Cape endemics); Molecular phylogeny [23]; Growth forms – evergreen shrubs or small trees, some prostrate; Significant prior work: reproduction and seed predation [24], functional traits [25, 26]; genetics [27]; community assembly [28].

**PELARGONIUM** – Species diversity (~280; 70% Cape endemics), some widespread species that defy simple circumscription, and many unidentified/undescribed species; Molecular phylogeny [29, 30]

**Growth forms:** diverse, from small shrubs & shrublets to stem succulents, geophytes and annuals – leaves deciduous or evergreen/persistent; Significant prior work: functional traits, rates of leaf trait evolution [30, 31], photosynthetic pathway [32], plasticity evolution [33], environmental niche modeling [34].

**RESEARCH PLAN**

I. VARIATION IN FUNCTIONAL TRAITS AND ITS GENETIC BASIS

How do functional traits vary within and between species? Are differences related to environmental variation (e.g., climate, soils, and fire frequency)? What is the genetic basis of such differences?

In the past decade, analyses of global databases for leaf traits have revealed universal trade-offs between rapid acquisition and conservation of resources. Additionally, a single trait, specific leaf area (SLA—leaf surface area supported by a given tissue investment) is strongly associated with a species' position along this continuum [35]. Because SLA is highly correlated with whole plant growth rates, and thus with performance, ecologists have focused on it and correlated traits as key "functional traits" [36, 37].

SLA and other leaf traits frequently covary with environmental variables such as mean annual precipitation (MAP), temperature (MAT) or soils at broad scales of communities and ecosystems [38, 39]. The specific shape and extent of that covariation depends on both the phylogenetic history [40] and plant growth form [41]. Understanding how environmental history has shaped the evolutionary history of traits in specific lineages and how current leaf trait variation predicts responses to future environmental change at the species level requires fine-scale (species level) mapping of the evolutionary link between environment and functional leaf traits, and between these traits and leaf physiology (photosynthesis, WUE). Moreover, because species differences in functional traits originate via divergence among populations, understanding the genetic basis of functional trait variation within and among populations and closely related species is essential to knowing the ultimate origin of current variation and the potential for the generation of future diversity.

**Background and Specific hypotheses:** *Protea* species are evergreen and individual leaves live multiple years – as such, we predict that functional leaf traits among proteas will have strong correlations with levels of environmental stress. We demonstrated such a relationship in white proteas for SLA in relation...
to cold stress [25] (Fig 1). In contrast, Pelargonium species range from evergreen to deciduous – thus we expect that both stress tolerance and stress avoidance strategies should be present, such that correlations between traits and environmental variables across the phylogeny will be weak or non-existent. Support for this prediction comes from an analysis of SLA for 100 species collected throughout the GFCR that shows no significant relationship with environmental variables (unpub data). Concomitantly, we predict that functional leaf traits in Pelargonium will be highly correlated with leaf longevity, reflecting a range of functional strategies modulated by growth form. Ultimately, we predict that deciduous species will be less affected by climate change than evergreen species when soil nutrients are sufficient to support the deciduous habit.

We will construct an axis of functional trait combinations (e.g., via PCA) corresponding to a range from drought tolerance to drought avoidance for both genera. We predict that Protea will be predominantly located at the drought tolerance end of the axis while Pelargonium will be distributed across the entire axis. Both extensive (for most species) and intensive trait sampling (a subset of species in each genus) will characterize a species’ location on this axis.

IA. Functional trait variation

Extensive trait study: Quickly measured leaf functional traits (TRAIT SET A): leaf size, length and width, fresh and dry mass, SLA, succulence, fluorescence, chlorophyll content, conductance, toughness, internode diameter at the most recently expanded leaf (a measure of relative stem succulence for Pelargonium), and plant height and growth form will be measured for 7-10 individuals in 60 Protea and 120 Pelargonium species; tissue samples for phylogenetic analyses will be collected simultaneously. Based on previous experience collecting Protea and Pelargonium in the GCFR, we believe the sample sizes for each clade to be realistic estimates over a two-year period. We will use the Protea Atlas Project and a Pelargonium database (9000 location records compiled from Acocks and PRECIS databases, published descriptions and our field data) to locate collection sites. Voucher specimens will be deposited in the UConn Herbarium (CONN) and Compton Herbarium (NBG).

Intensive trait study: To calibrate field measures of trait set A with detailed measures of leaf functional physiology across multiple species we will also measure TRAIT SET B (leaf longevity, $A_{max}$, instantaneous WUE [=net assimilation rate/stomatal conductance], integrated WUE [$\delta^{13}C$], % leaf C, % leaf N, stomata number and size), for 30 species/genus from sites near the center of each species’ distribution. For each species, detailed measures will be made midway through and near the end of the growing season to document performance under both peak and stress conditions. Leaf longevity in Protea can be assessed by counting the number of annual growth increments retaining leaves. Most Pelargonium species are seasonally deciduous so measuring leaf life span will require monthly field assessment (no prior data are available). To investigate trait interrelationships under controlled conditions, we will also measure TRAIT SETS A & B on 2 yr-old seedlings from these same species at the UConn greenhouses.

IB. Trait variation in relation to environmental variation.

To link functional traits with climate variables, we will use the extensive climate data layers available for South Africa [17] described above to approximate annual and seasonal climate parameters for each collection site. South African soils are broadly classified as derived from shale, sandstone or granite and are notoriously nutrient poor. Soil type has been suggested to drive some South African speciation [42, 43], and because leaf life span may be linked to soil fertility [44], we will also analyze soils for nutrient content and soil texture and measure soil depth from each collection site in the Extensive trait study. Fire history for each collection site will be based on our previous reconstruction of fire history for protected areas in the CFR; these show that fire return intervals have decreased 20% in the past 30 years [45].
Along with climate-induced changes in post-fire recovery rates \cite{46}, accelerating fire regimes are predicted to have profound effects on ecosystem processes and biodiversity patterns.

**Data Analyses:** We will use two approaches: (1) Hierarchical Bayesian and generalized linear models to investigate contemporary trait-environment correlations with environmental measurements as covariates, with populations (nested within species) and species as random effects, and with phylogenetic relationships among species being used to inform priors on species random effects \cite{47}. (2) Phylogenetically independent contrasts \cite{48} will be used to investigate the phylogenetic history of contemporary trait-environment correlations. (See IIC for a description of phylogenetic methods.)

**IC. Genetics of trait variation.**
Evaluating responses to either future or historical climatic change requires knowledge of the genetics basis of adaptive variation. In order for selection to produce adaptation to changing environmental conditions, genetic variation for functional traits must be available within species. We can estimate current levels and distribution of genetic variation and make inferences about past selection by examining populations of widespread species. **Historical selection** can be inferred at the landscape level from correspondences among phenotypic and genetic variation for traits and environmental gradients. The question of future adaptability depends to some extent on whether genetic variation is continuous throughout a population or species. If fragmentation is extensive and genetic variants isolated, then adaptation will be limited to new mutations. Genome wide patterns of mostly neutral genetic variation will reveal the extent of gene flow.

We will detect the genetic architecture of trait variation through mixed-model, kinship-corrected association mapping \cite{49, 50} that can account for residual population structure from non-random gene flow. This will allow us to detect functional loci and possibly the selection on them. In humans, such studies have identified single nucleotide polymorphisms (SNPs) associated with more than 500 traits (NHGRI, [http://www.genome.gov/26525384](http://www.genome.gov/26525384)).

**IC.1 Genetics of functional trait differences in widespread species** - To identify loci associated with functional trait differences, we will investigate the genetics of among-population functional trait differences in *Protea repens*, a widespread, abundant species found throughout the fynbos in the CFR, and in *Pelargonium scabrum*, a widespread species found in habitats from wet fynbos to Succulent Karoo (MAP range 2-200 cm/yr). We will collect seed from 40 plants in 20 populations of each species, establishing one offspring from each maternal line in experimental gardens at Kirstenbosch National Botanic Garden. We will use the PAP and *Pelargonium* databases to identify localities spanning as much of the variation in climate, soils, and fire frequency as possible. We will measure TRAIT SET A and collect genetic samples from ~1600 individuals total. This approach is suitable because shared ancestral alleles are likely to have sorted independently in isolated populations.

In genome-wide association studies of predominantly outcrossing organisms, alleles segregate within a single, well-mixed population – associations between alleles at a particular locus and the phenotype are likely to represent true QTL. In contrast, when allele frequencies differ among populations, associations among alleles at different loci may be due to purely statistical associations reflecting population structure \cite{51}. **Kin-GWAS** \cite{49, 50} is an effective approach available for eliminating false positives, allowing both fixed effects associated with population of origin and for (estimated) degrees of kinship within populations, as well as genome-wide and marker-specific genetic effects \cite{52, 53}. We will use [Kin-GWAS](http://www.genome.gov/26525384) (implemented in either **TASSEL** \cite{54} or **EMMAX** \cite{49}) to assess marker-trait associations. For genotyping, we will use state-of-the-art massively parallel sequencing to develop a large number of SNPs using RAD sequencing (see Genomics methods below).

**IC.2 Genetics of trait differences between closely related species**: Past selection can also be manifested genetically in the form of ‘legacy’ effects at many sites across the genome due to population structure, or as retention of particular functional loci in the face of current gene flow that reshuffles neutral genes. We will use standard genome-wide association analysis (**GWAS**) for traits segregating in hybrid zones (so that alleles and traits can segregate) using large within-population samples. This is the most powerful approach to date for distinguishing these alternatives.
In GWAS many associations involve SNPs in which the minor allele frequency is only 5-10%. We will sample up to 1000 individuals in each hybrid zone, giving us good chance to find an association with a trait with 10% heritability at 10% minor allele frequency. To assess the generality of the associations we discover, we will replicate the analyses for each species pair at two geographically distinct sites. We will measure traits on field-collected samples, but soils and climate will be relatively uniform within any collection site and we will perform association analyses within each site, minimizing the influence of phenotypic plasticity on trait measurements. In short, we will measure traits and determine the genotype of 6000 individuals in this part of the project. We will use Kin-GWAS (implemented in either TASSEL [54] or EMMAX [49]) to identify trait-marker associations. For genotyping, we will use state-of-the-art massively parallel sequencing to develop a large number of SNPs using RAD sequencing (see Genomics methods below).

**FOCAL TAXA:** We will investigate the genetics of trait differences for three examples of hybrid populations of closely related species: Protea venusta - P. punctata, Pelargonium ternatum - P. cordifolium and the Pelargonium carnosum complex.

Protea venusta is low-growing and prostrate, with small, oblanceolate leaves (25-60 mm long, 8-20 mm wide; length:width ratio 2.46±0.40; SLA 217 cm²/g). In contrast, Protea punctata is an erect shrub growing to as much as 4m tall, with large, ovate to elliptic leaves (35-80 mm L, 20-45 mm W; L:W ratio 2.99; SLA 158 cm²/g; [25, 55]). They also differ in the size, shape, and orientation of involucral bracts. Nonetheless, they hybridize freely when they come into contact at high elevations in the Swartberg mountains. In fact, a recent analysis using 10 microsatellite loci detected extensive hybridization at Blesberg even though samples consisted only of morphologically “pure” individuals [27]. Our association analyses will focus on growth morphology (erect vs. prostrate measured on an ordinal scale), leaf area, leaf shape (L:W ratio) and SLA.

Pelargonium carnosum s.l. is widely distributed across ~600 km in both Fynbos and Succulent Karoo. It has long been recognized as polymorphic, and Becker & Albers [56] have partitioned 4 taxa based on leaf and inflorescence characters. An AFLP phylogeny supported these clusters, but many specimens with mixed character states are found in the tree, indicative of hybridization among taxa. The 4 taxa occupy somewhat distinct ranges, but multiple ‘in-between’ locations harbor hybrids. The authors suggest that many trait differences “can be explained as adaptations to prevailing environmental conditions”, including differences in leaf pubescence and root structure. Our association analyses will focus on leaf traits (leaf area, SLA, pubescence).

In Fynbos, hybrids also occur between the morphologically distinct species Pelargonium ternatum (a shrublet to 0.6m) and P. cordifolium (a shrub to 1.5m).

Field measures of ‘parental’ and hybrid plants: ternatum - trifoliate leaves ~6cm², petiole 1cm, flowers single, petals 10-15mm, SLA=78 cm²/gm; cordifolium - cordate leaves ~56cm², petiole 9cm; umbel of 4-8 flowers, petals 20-30mm, SLA=114 cm²/gm; hybrid - lobed leaves 28cm²; petiole 9cm, SLA=103 cm²/gm. Association mapping will focus on stature and leaf traits.

**IC.3 Gene flow, mutation, and effective population size:** Data collected for genetic analysis of functional trait differences (section IC.1) will give us almost unprecedented insight into fundamental population genetic parameters for wild populations of a non-model organism. In addition to using the SNP data to estimate levels of within- and among-population diversity [57, 58], we will study the influence of recent land-use changes on patterns of gene flow by comparing estimates of recent and long-term migration rates [59]. Specifically, we will use existing methods to estimate recent rates of gene flow [60], to estimate the effect of environmental parameters on those rates [61], and to estimate long-term migration rates, mutation rates, and effective population size [62]. Analysis of these data with STRUCTURE [63-65] will also allow us to determine whether genetically defined clusters are congruent with current taxonomic boundaries in Protea repens and Pelargonium scabrum.

**IC.4 Genomics methods:** We will take advantage of massively parallel next-generation sequencing technology to develop extensive SNP markers for our association studies, landscape population genetics, and phylogenetics. Specifically, we will use a combination of sequenced RAD markers [66] to provide a
large number of SNPs for association mapping (RAD) and transcriptome sequencing (TS; see IIC) to provide a large number of haplotypes for phylogenetic analysis. In RAD sequencing, genomic DNA is digested with a restriction enzyme prior to ligation of an adapter and short-read Illumina sequencing. As a result, sequencing is concentrated on parts of the genome adjacent to the recognition site of the restriction enzyme regardless of whether or not that sequence is in a coding sequence. By concentrating on these reduced parts of the genome, we gain enhanced coverage, allowing more samples to be tested so allele frequency can be estimated (vs. misidentifying singleton sequencing errors as SNPs). Most SNPs will likely be anonymous nuclear markers, which will be very useful at population and hybrid-zone levels [67]. Although identified in short reads (~50bp), RAD loci as large as 500bp can be assembled from fragments sequenced at both ends. Such loci may also be useful for phylogenetic analysis.

II. Phylogenetic Signal

What are the phylogenetic histories of functional trait change? Are the rates of trait evolution related to changes in niche parameters?

Scoble and Lowe [7] have emphasized the need to identify species’ potential for adaptation to understand constraints on their distribution and their capacity to respond to environmental change. By coupling our phylogenetic inferences with detailed information on modern environments, we will evaluate the historical capacity of Protea and Pelargonium to respond to challenges in their environments. To do this we need to know not just that traits have evolved, but how quickly and whether changes of individual traits have been correlated. Although rates of species diversification have been high in both clades, we are interested in the tempo of change in traits related to adaptation. The issue of correlated change is important from the perspective of evolvability: Are there coordinated modules that change in concert through time (adaptive suites)? and Is there evidence for new patterns of coupling or decoupling of traits through time (a potential for breaking constraints)?

The age of the crown node in Protea has been estimated from ~15-18Mya. The crown node of Pelargonium dates to ~30Mya [23]; the winter rainfall clades, with > 80% of the species, have a Miocene origin (16-20Mya)[68]. South Africa in the Mid-Miocene experienced a period of rapid climate change with a progression from sub-tropical with palm and montane forest, to cool and dry habitats with drought-tolerant species [69]. Continued aridification has been accompanied by additional diversification in both genera: e.g., Pelargonium section Hoarea has produced 90+ extant species over the past 8My, and more than half of extant Protea taxa have arisen within the past 5My [23]. We will exploit the fact that both genera have replicate radiations in the fynbos (and in the SK for Pelargonium; e.g. clades A and C in Fig. 2), by making explicit comparisons of rates and patterns of trait change between clades within genera.

IIA. Phenotypic evolution: traits and niches

IIA.1 Evolutionary rates of functional traits: To understand both the scope and tempo of the radiations in these lineages, we will reconstruct the history of character state changes at nodes along the phylogenetic trees. This is straightforward using our explicit phylogenies (methods below) and standard comparative methods [48, 70, 71]. We will calculate the metric β from the reconstructions, where β is the variance of the standardized independent contrasts [72]. β can be estimated for any good sized clade and, since this particular measure of β is χ² distributed, differences in rates between clades or traits can be evaluated with F tests. We will also investigate what types of evolutionary processes may have produced the patterns of rate changes that we observe. We will evaluate three models of character change using GEIGER [73-75]: pure Brownian motion, evolutionary changes
focused at speciation events, or stabilizing selection (an Ornstein-Uhlenbeck process).

**IIA.2 Patterns of correlated trait evolution:** Tightly integrated functional traits should show similar patterns of character state reconstructions and similar rates of evolutionary change. Reduced integration among traits can result in mosaic evolution, with traits showing differing degrees of lability and varying rates of evolution across the tree. Ancestral state reconstructions of leaf morphology in *Pelargonium* found evidence for a mosaic pattern across the whole genus, with character state change ranging broadly from quite conservative to extremely labile. We also discovered an example of evolutionary increase in integration: clade A2 (Fig 2), consisting of species with succulent or underground stems, had significant correlations between leaf venation, leaf dissection and functional leaf sizes [30].

We will use the likelihood methods of Revell and Collar [76] to examine whether, and when, the evolutionary covariances between characters have changed in the history of a group. A particular advantage of this method is that it allows a test of whether there has been non-proportional change in the evolutionary variances/covariances among traits. We will focus on testing predictions related to the replicate radiations within biomes for each genus. For example, in *Protea*, do the white Proteas (sect. *Exsertae*) and rose and penduline Proteas (sects. *Pinifolia* and *Craterifolia*) show modifications of the same sets of leaf traits in response to increasing cold or drought? In *Pelargonium*, can we apply a single evolutionary rate to correlated trait evolution in clade A, or do clades A1 & A2 require separate rates?

**IIA.3 Reconstructing niche parameters.** By treating the parameters of the niche (MAP, MAT, soil type etc.) as characters, we can reconstruct the patterns of evolutionary change in realized niches as we did for traits in **IIA.1.** This will allow us to ask whether particular clades explore new niche space more readily or have been conservative (high vs. low $\beta$). We can also (as in **IIA.2**) examine the degree to which evolution can uncouple adaptation to niche dimensions. For example, both genera have members that occupy summer, year round or winter rainfall regions – is there a common pattern of functional trait evolution that accompanies the switch from one rainfall regime to another?

**IIB. Phylogenetic methods.**

We will have tissue for DNA analysis from ~70 species of *Protea* and ~120 species of *Pelargonium*. To ensure that we have sufficient markers suitable for phylogeny reconstruction, we will complement RAD sequencing with transcriptome sequencing (TS). As with RAD, TS concentrates our sequencing efforts on only part of the genome, enhancing coverage and reducing the possibility that sequencing errors are identified as SNPs. Both *Protea* and *Pelargonium* are distantly related to plants for which full genomes are known, thus our first step is to construct an EST library. We will then identify homologies with known or inferred genes in *Arabidopsis*, maize, and other plants for which complete genomes are available. For sequencing, we will isolate mRNA from actively photosynthesizing leaves and reverse transcribe the mRNA to cDNA before barcoding individual DNA and preparing multiplexed libraries for sequencing with an Illumina Genome Analyzer II. Bioinformatic processing of sequencing products will allow us to assemble large contigs from these short reads. We already have an EST library from *Protea mundii*, and the EST library for *Pelargonium* will be constructed in project year 1 from *P. cucullatum*.

Phylogenetic relationships within each genus will be estimated with methods analogous to those in analyses of the animal tree of life EST data [77, 78]. These researchers used BLAST searches to identify orthologs across the animal phylogeny for sequence alignment. Our contig assembly is based on orthology to the EST libraries of our focal genera, making assessment of sequence homology straightforward. Phylogenetic analysis of the animal data (a complete Bayesian analysis of all sequences) required several months on a computer cluster with 120 processors. Either hybridization or sorting of ancestral polymorphisms leads to different evolutionary histories for different genes, and both may contribute to discordant gene trees at lower levels of the taxonomic hierarchy in our analysis. As a result, we will focus our efforts on a less computationally intensive approach that allows for discordance among gene trees. This approach will also allow us to incorporate data from markers already available [23, 68]; any cryptic taxa will be apparent on the resulting phylogeny.

Much of the computational burden associated with a complete analysis arises from the large amount of missing data. Fewer than 50% of the cells in the taxon x character matrix were filled in Dunn et al.
[77], and likelihood or Bayesian algorithms “marginalize” across all possible character states in those OTUs for which data are missing. Because each of the loci in our analysis may have an independent evolutionary history, we will use standard Bayesian approaches to estimate phylogenetic relationships separately at each locus. Using well-established methods to assess the congruence of gene trees [79], we will combine the separate analyses of loci with congruent phylogenetic histories into a single phylogenetic supertree estimate. We will also perform a complete Bayesian analysis of all loci with congruent histories using a new 128-node cluster available to us at UC Davis.

III. COMMUNITY STRUCTURE AND ECOSYSTEM RESILIENCE
How are taxonomic and trait level diversity and abundance of focal taxa interrelated with community structure along environmental gradients? How are these patterns related to ecosystem resilience?

The products of evolution are best understood when viewed in the biotic and environmental contexts where they were formed. In designing this project we recognized a unique opportunity to interpret functional trait variation of our focal clades in the context of the communities in which they occur. *Protea* and *Pelargonium* are not just diverse South African clades, they also span two key axes of functional strategies within the GCFR flora, i.e. evergreen to seasonally deciduous, and overstory shrubs to subshrubs and geophytes (South Africa has the most diverse bulb communities on earth). They also traverse long environmental gradients in aridity, soil type, and fire frequency. By understanding how within and among-species trait variation along environmental gradients relates to community composition, we will refine predictions of community structure and ecosystem resilience to disturbance by fire under existing and future conditions (e.g., Fig. 3).

Fig. 3. Our preliminary analyses of climate model projections for temperature and precipitation in the GCFR suggest both will shift substantially – models generally agree (stippling) on higher temperatures throughout the region and on divergent trends for precipitation in the west (drier) and east (wetter).

III.A. Functional traits and community structure.
Species abundance patterns must be placed in a community context to make robust species or community level predictions, and species’ functional traits mediate their interaction with the surrounding community. We will develop models that link taxonomic and functional diversity with functional trait and phylogenetic data using novel modeling techniques including phylogenetically informed multiple linear regression [28, 47] and entropy maximization (EM) [80, 81]. We will use EM to reduce the dimensionality of the problem by grouping similar species into functional groups, reflecting a mechanistic understanding of species interactions [82].

III.A.1 Predicting abundance patterns along ecological gradients: Our working hypothesis is that functional composition of communities in the GCFR – and perhaps other diverse plant communities – is constrained relatively strongly by environment-trait interactions; but that species composition within functional groups is dominated by lottery-like dynamics as functionally similar species are selected by availability of propagules and stochastic population dynamics. To assess these hypotheses, we will simultaneously model abundance of functional groups using niche-based processes and abundance within functional groups using neutral processes. We will also determine the relative contribution of niche-based and neutral mechanisms to local trait and taxonomic diversity within communities. Within our focal taxa we will also determine the extent to which the degree of phylogenetic relatedness among focal taxa
affects the balance between niche-based and neutral processes.

In particular, we will identify functional traits in which local community trait distributions are unimodal and stable along replicated environmental gradients and determine how community-aggregated (i.e., abundance-weighted) average trait values (CATs) respond to environmental gradients. Although much trait variation exists within communities [83], maximum-entropy analysis using easily measured “soft” CATs predicts 50-75% of the variation in abundance of dominant species in a South African plant community [81]. In this project, we will expand this analysis along broader environmental gradients to cover a much larger number of communities and associated functional variation among communities (see III.C.3). In addition, we will investigate patterns of trait variance within communities as an indication of the strength of habitat filtering within the community. Using the detailed phylogenetic information developed in our research on Question II, we will investigate the strength of phylogenetic patterns (e.g. phylogenetic over- and under-dispersion and the relationship of these to functional traits) in community composition [84].

The CFR is well suited for examining the relative importance of niche and neutral mechanisms in community assembly [85, 86] because communities respond strongly to ecological gradients (fire return intervals - Fig. 6, below), annual precipitation (Fig. 4), and precipitation seasonality, while the apparently high levels of functional redundancy [87] and dispersal limitation [88, 89] suggest a role for lottery or neutral dynamics [10]. Sampling of relevé plots on the Cape Peninsula in 1966 and 1996 revealed strikingly high species turnover [90] (Fig. 5); because successional changes in species composition in these communities are minimal [91], high turnover observed suggests lottery-like patterns of replacement of species by functionally similar species [cf. 92]. Studies on some taxa have revealed patterns of phylogenetic over-dispersion in these communities [93], suggesting that competition may restrict the number of co-occurring functionally similar species. Our analyses have shown that the influence of phylogeny depends on the spatial scale at which processes are observed [28].

### III.A.2 Focal species as indicators of the community level distribution of functional traits

Community level patterns may inform species level patterns because many of our focal species appear to have functional equivalents with which they co-occur. A species may be absent from a sample simply because its functional equivalent replaces it, not because the environment is unsuitable. Since many of our focal species have narrow ranges, we can understand how traits of our focal species reflect environmental gradients only by placing their abundance patterns in a community context. We will examine these effects in models that relate average local trait values in our focal taxa both to environmental covariates and average community trait values.

Similarly, trait values in our focal clades may inform community level patterns: *Proteas* inform us about the dominant species, while *Pelargoniums* reflect the range of habitats through their diversity of growth forms (geophyte, annual, stem succulents and shrubs). *Proteas* are found throughout the CFR and represent the dominant growth form (overstory shrubs) in fynbos. In contrast, *Pelargoniums* span a range of functional strategies and are found throughout the GCFR. If we detect strong associations between traits in our focal species and community trait

---

**Fig 5.** Turnover is the rule rather than the exception. Site by species temporal turnover (3 *Protea*, 8 *Pelargonium*) in a survey of 80 50m relevé plots in the Cape Point Nature Reserve in 1966 and 1996.
distributions, our focal species can serve as indicator species for community traits. Such communities will be part of the “extended phenotype” [94] of our focal taxa. We will examine these relationships in regression models that relate community trait values to trait values in our local taxa and environmental covariates. Because of the marked differences in life-history and growth form between Protea and Pelargonium we will focus on models relating the distribution of local trait values in our focal taxa to the distribution of community trait values, rather than on simple regression models that focus on average trait values [cf. 83].

IIIA.3 Data collection: Community abundance and trait data (selected from TRAIT SET A) for all species will be collected at 15 sites containing Protea and/or Pelargonium that span the range of environmental variation (annual precipitation, precipitation seasonality, time since fire) found in the GCFR. Community composition at each site will be assessed with samples taken from 10-20 50m² relevé plots to account for within-site heterogeneity. While we will choose sites/communities that contain at least one focal clade; at the scale of relevés within sites, we expect turnover among functional equivalents such that not all relevés will contain focal clades, as shown in the historical relevé data. We will focus initially on the large set of sites (>10) for which relevé data are available through SANBI to generate a time series for abundance patterns. Remaining sites will be selected to maximize coverage along the environmental gradients. We will conduct two types of surveys; (1) detailed surveys for communities at the center and endpoints of the gradients with 10-20 trait samples per species, and (2) rapid surveys for the remaining sites with 3 trait samples per species.

IIIA.4 Modeling: We will develop two types of models to predict species and functional group abundance: phylogenetically informed multiple regression models (PIMR [27, 47] and entropy maximization algorithms (EM) [80, 81]. PIMR identifies relationships between species co-occurrence, traits, and environmental covariates, while controlling for unmeasured trait variation that is phylogenetically correlated. The method is a straightforward extension of standard linear or generalized linear mixed models in which the phylogenetic distance between each pair of species is used either to adjust the corresponding pairwise random-effect covariance (for analyses in which characteristics of individual species are the response variable) or as a covariate in the model (for analyses in which the characteristics of species pairs are the response variable).

EM finds abundance distributions whose CATs match empirical CATs. The CATs encompass niche relationships and act as constraints on community membership based on functional summaries of the communities. It can incorporate and test for dispersal limitation by comparing models with uniform priors on species abundances with models using informative prior distributions that reflect among-species differences in recruitment probability [81]. We will develop new approaches using EM that: (1) match the distribution of trait values within communities, rather than focusing only on community-average trait values, and (2) use niche-models to predict the abundance of functional groups and neutral models to predict the abundance of species within functional groups. This involves formulation of the EM model that builds on the concept of the degeneracy of states from physics and offers an alternative way to integrate niche and neutral dynamics - allowing niche constraints to operate on functional groups while neutral mechanisms dominate within functional groups.

IIIB. Community and ecosystem resilience
To examine the relationship between community composition and ecosystem resilience, we will examine effects of disturbance frequency on taxonomic and trait diversity within communities and of taxonomic and trait diversity on ecosystem resilience. Ultimately, we propose to predict both clade and community responses to climate change scenarios based on models that incorporate phylogenetic relationships, trait values, fire frequency, and biotic and abiotic niche parameters.
III.B.1 Determining the effect of disturbance frequency on taxonomic and trait diversity: Preliminary data from plots at Cape Point show that taxonomic diversity within our focal taxa is strongly related to fire-return interval (Fig 6). We will use the community data collected in IIIA.3 and data on fire return intervals from analysis of historical and satellite records to construct Bayesian regression models to examine the relationship between a suite of environmental covariates, including fire return interval, and two measures of taxonomic and trait diversity both for our focal taxa and for the communities in which they are embedded: (1) mean Shannon-Weaver diversity at each of the 15 sites, $s_i$, and (2) the proportion of Shannon-Weaver diversity accounted for by among-plot differences within each site, $f_i$ (see note on partitioning Shannon-Weaver diversity in the next paragraph). The two measures of diversity will be treated as a multivariate response:

$$\mathbf{y}_i \sim \text{MVN}(\mathbf{\mu}_i, \mathbf{\Sigma}) \quad \text{and} \quad \mathbf{\mu}_i = \mathbf{X}_i \mathbf{B}_i + \mathbf{e}_i,$$

where $\mathbf{y}_i$ is the two-dimensional vector $(s_i, f_i)$ of diversity measures at site $i$, $\mathbf{\mu}_i$ is the corresponding mean vector, $\mathbf{B}_i$ is the vector of regression coefficients for component $j$ of $\mathbf{\mu}_i$, and $\mathbf{X}_i$ is the vector of environmental covariates. We will place a vague inverse-Wishart prior on $\mathbf{\Sigma}$ in our analyses. In addition to separate analyses of the relationship between taxonomic and trait diversity, we will also consider a more general approach in which measures of taxonomic and trait diversity are simultaneously included in $\mathbf{y}_i$ allowing us to statistically control for correlation between measures of taxonomic and trait diversity.

Our measure of among-plot diversity is analogous to $F_{ST}$, the widely used measure of among-population diversity in population genetics [95, 96]. Specifically, if $p_{ijk}$ is the frequency of species $j$ in plot $k$ at site $i$, then the among-plot diversity, $f_i$, is

$$s_i = \frac{1}{K} \sum_k \sum_j p_{ijk} \ln(p_{ijk}) \quad \text{and} \quad f_i = \frac{1}{s_i} \sum_j p_{ij} \ln(p_{ij}),$$

where $p_{ij} = \frac{1}{K} \sum_k p_{ijk}$.

It is straightforward to generalize this univariate measure of diversity to a vector of trait values to provide estimates of trait diversity. We use an information-based measure of diversity in this analysis because it is also straightforward to generalize the discrete version of the measure just described to a continuous measure for traits, like SLA, that are not naturally discrete. As a result, we can easily combine variation in both discrete and continuous traits into a single measure of trait diversity.

III.B.2 Determining the effect of taxonomic and trait diversity on ecosystem resilience. Fire plays a central role in GCFR ecology, with fire return times establishing boundaries between biomes (fynbos shrubland vs forest; fynbos vs succulent karoo) and strongly affecting community composition within them. Fire return times are in turn influenced by climate and other environmental factors as well as community functional composition.

Preliminary results drawing from both direct biomass observations and remote-sensing reveal substantial variation in post-fire recovery times (Fig 7). A substantial amount of this variation is associated with differences in climate across the regions; our analyses find strong associations between post-fire regrowth rates in fynbos and local precipitation and temperature [46]. Differences in the composition of local communities are also likely to be important, since different species and functional types grow at different rates and tolerate different levels of drought. To explore the relationship between traits, community composition and resilience, we will construct Bayesian regression models in which the predicted rate of biomass accumulation at each of our community sample sites is associated both with environmental covariates and with measures of taxonomic and trait diversity (see III. B.1).

We will then use hierarchical Bayesian models we are currently developing to integrate the high-
quality fire history data for conservation areas (from land management agency CapeNature) with fire occurrence data derived from the more spatially extensive but temporally limited MODIS NDVI data. These models will project mean fire recurrence times for all populations sampled in fynbos shrubland for the recent past and, using climate projections, the near future. We will use a Bayesian approach in these analyses because it will allow us to accommodate spatial and temporal correlation in the data using random effects, and because it provides a full description of the statistical uncertainty associated with both the predicted rate of biomass accumulation and the diversity estimates.

**DATA MANAGEMENT**

All data, software, and protocols developed as part of this project will be made freely available on a project webserver to be housed at the University of Connecticut. General information about the project (participants, goals, project timeline, links to published papers and presentations, links to software used in or developed as part of the project, etc.) will be part of a project wiki to which all members of the project will contribute. All software developed as part of the project, whether WinBUGS code, R scripts or packages, or special purpose PHP, Perl, or C/C++ code will be disseminated under terms of the GNU Public License (v2.0 or later). EST libraries will be deposited in GenBank and short-read sequences from RAD sequencing and transcriptome sequencing will be deposited in NCBI’s Short Read Archive (SRA). In addition, all data collected as part of the project will be housed in a Metacat data repository (http://knb.ecoinformatics.org/software/metacat/), and we will use the Ecological Metadata Language (EML: http://knb.ecoinformatics.org/software/eml/) to provide the metadata associated with each data set. Individual records within each data set will be geo-referenced, linked to genomic data in SRA as appropriate, and linked the publications in which those data are used. Release of geo-referenced data will be coordinated with appropriate officials in South Africa to ensure that threatened taxa are not targeted by general collectors. We will make each data set publicly available as soon as the PIs are confident that the records within it are accurate and complete. In addition to using the project server to share complete data sets and protocols, we will take advantage of freely accessible groupware platforms (e.g., Google Wave or SAP Streamworks) to share preliminary documents, protocols, notes, and data freely among all labs.

**PROJECT MANAGEMENT**

Although the project involves work at three institutions in the United States in addition to field and experimental garden work in South Africa, the PIs, post-docs, graduate students, and technicians will work as an integrated team. We will hold annual project meetings (two in South Africa in conjunction with field work in years 1 and 3, one each at UConn, Chicago, and Davis in years 2, 4, and 5), and monthly videoconferences to plan and coordinate all project activities.

**PI responsibilities:** All PIs will be engaged in all aspects of the project, but each will have primary responsibility for particular parts appropriate to their expertise and experience. **Schlichting** will be responsible for overall project coordination and management. **Holsinger** will lead the team collecting and analyzing *Protea* trait data, and developing and maintaining the project server. **Jones & Schlichting** will lead the team collecting and analyzing *Pelargonium* trait data. **Silander** will lead the team responsible for collecting community-level trait data. **Borevitz** will take primary responsibility for generating and bioinformatic processing of RAD sequence data for GWAS and Kin-GWAS. **Latimer** will take primary

![Post-fire recovery time](image.png)

**Fig 7.** Post fire recovery time (years) across the fynbos biome as modeled using satellite (MODIS NDVI) observations and climate, topography, and incoming solar radiation (Wilson et. al. *in prep*).
responsibility for generating and bioinformatic processing of transcriptome data. *Holsinger, Jones,* and *Schlichting* will take joint responsibility for analyzing trait-environment correlations. *Borevitz, Holsinger,* and *Latimer* will take joint responsibility for trait-marker association analyses. *Holsinger, Latimer,* and *Silander* will take joint responsibility for community composition and ecosystem resilience analyses.

**SENIOR PERSONNEL:** Drs. Elizabeth Marais (emeritus, University of Stellenbosch) and Freek Baker will provide taxonomic, systematic and phylogenetic expertise on *Pelargonium.* Dr. Tony Rebelo (SANBI) will join field teams and verify *Protea* species identities.

**PERMITS:** We have received all necessary collecting and export permits from Cape Nature, the Northern Cape Province and SanParks for sites visited for background data collection; we anticipate positive responses to applications for the proposed work. We adhere to all collecting restrictions associated with permits, including those for Red-listed taxa.

### PROJECT TIMELINE

<table>
<thead>
<tr>
<th>Activity</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Focal clades</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trait set A*</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trait set B**</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Traits analysis</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Evolutionary rates</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><em>Community</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traits collection</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Traits analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resilience analysis</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><em>Widespread species</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Establish garden</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNP scoring†</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Hybrid zone</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample collection†</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNP scoring</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td><em>Phylogenetics</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transcriptome</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Analysis</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

*seeds/cuttings for garden and transcriptome samples will be collected in 2011
**seeds/cuttings for greenhouse will be collected in 2012
†traits will be scored at the same time

**BROADER IMPACTS**

*Integrating Research, Teaching, Mentoring and Learning:* This project comprises a diverse, complementary group of individuals, disciplines and institutions. In consequence it offers unique interdisciplinary and foreign education and research experience for students, post-docs and scientists. Among the key personnel are senior, mid-career and junior researchers, representing seven institutions across three countries: 6 universities - UConn, Chicago, UC Davis, Cape Town, Stellenbosch, and Wageningen - and the South African National Biodiversity Institute (SANBI, a parastatal NGO). The project blends a unique combination of experts in genetics, taxonomy and plant functional biology, as well as ecology (population to landscape), statistics and modeling, and conservation biology. We will take
advantage of this broad range of expertise to enhance student and post-doctoral training at all institutions, to broaden participation of underrepresented groups, to enhance conservation infrastructure in South Africa, and to disseminate results of our research to the general public.

**Enhancing student and post-doctoral training:** We will hold an annual methods workshop in conjunction with each annual project meeting (see **Project management**) for all grant participants. In addition, **workshops will be open to any member of the host institution.** Each workshop will last for 1-2 days and will focus on a different set of methodological or analytical tools used in the project, e.g., association mapping, comparative methods for estimating rates of trait evolution and for detecting correlated trait evolution, entropy maximization, hierarchical Bayesian modeling, next-generation sequencing, statistical phylogeography. In developing these workshops, we will build on a new UConn initiative with Silander & Holsinger among its leaders – “Cyberinfrastructure to Integrate Biodiversity and Environmental Research” (CIBER). CIBER currently includes 5 modules: 1) Cyberinfrastructure and GIS resources, 2) Modeling biodiversity patterns and processes, 3) Climate change: accessing and evaluating climate and weather station data, 4) Taxonomy, data collection, verification, imaging, and digitization standards. 5) Databases and serving bio-spatio-temporal data. Materials developed for our workshops will be added to the CIBER website, freely available under a Creative Commons license.

**Broadening participation of underrepresented groups:** We will build on the success of approaches we have used on current NSF grants in which we recruited minority graduate and undergraduate students (African American, Hispanic, Pacific Islander), a minority high school biology teacher, and a female, African American high school student (with RET and RAMHSS supplements), to participate in research in South Africa. Specifically, we will continue to participate in the UConn Summer Research Program for Minority Students, through which mentored undergraduates from minority institutions across the country as summer research interns in our labs. We will continue close collaboration with faculty and students from non-PhD-granting institutions (ROA supplements) and our outreach to minority high and middle schools in the region including GHAMS-CREC, Bulkley High School and Two Rivers Science Magnet School. We have also established linkages with South Africa underserved science schools in the Cape Town region, with the help of a former South African post-doc: LEAP school [www.leapschool.org.za/](http://www.leapschool.org.za/), Cape Academy [www.capecademy.co.za/](http://www.capecademy.co.za/), and COSAT (Centre of Science and Technology). We will use these linkages to help us recruit underserved students and assistants for fieldwork in South Africa. Finally, we will continue to participate in ESA-SEEDS, a minority recruiting program initiated in collaboration with the UNCF.

**Enhance conservation infrastructure:** A specific goal of this project is to foster international collaboration among diverse scientists and students, with established linkages among academic and NGO institutions in three countries. We will build on our existing relationships with South African and international conservation NGOs including TNC, CapeNature, WWF, CAPE [Cape Action for People and the Environment], Cape Flats Nature, and with South African GOs including CSIR and SANParks to ensure that the results of our work are easily accessible and readily interpretable to conservation leaders and land managers. SANBI’s mission is to promote biodiversity science, policy and conservation action, with a central focus on the GCFR; they support bioregional planning initiatives with coordinated conservation implementation, and a comprehensive threatened species action program. They bring together scientists, professionals, and members of the public to forge conservation policy and effect implementation. Their public outreach program includes local (indigenous) community empowerment through environmental education and awareness programs. SANBI’s participation helps to ensure that we directly link the science in this project with conservation application.

**Disseminating results:** In addition to the usual complement of scientific papers, books and presentations, we will continue to target publications for conservation applications [97] and to the general public [e.g., articles in Veld & Flora, CSSA Journal; 98]. In addition, the project web site will provide a variety of educational and general interest materials, including links to the specialized materials developed for project workshops. In developing these parts of the web site, we will draw on the lessons from the success of two public outreach web sites we have developed: Silander’s IPANE project: [http://www.IPANE.org](http://www.IPANE.org) and Rebelo’s Protea project: [http://protea.worldonline.co.za](http://protea.worldonline.co.za).


26. Prunier, R., K.E. Holsinger and J.E. Carlson. in prep. The role of local adaptation in an evolutionary radiation: root traits and whole plant integration in the white proteas (Protea section Exsertae). Evolution


POST-DOCTORAL RESEARCH ASSOCIATE MENTORING PLAN

Post-doctoral research associates associated with this project will be appointed at the University of Connecticut, at the University of California Davis, and at the University of Chicago. At UConn, one postDoc will be involved with each of three projects and primary mentoring will be as follows: genetics and taxonomy – Holsinger; functional trait analyses – Jones; community and ecosystem studies – Silander. All PIs @ UConn will be involved in the mentoring responsibilities via our weekly project meetings. Latimer will mentor the post-doc at UC Davis in conducting genomics work and its application via phylogenetic inference to community and ecosystem studies, including meeting each week with the post-doc, and Borevitz will mentor the post-doc at Chicago in conducting genomics work and its application to association mapping, including weekly meetings with the post-doc. All post-doctoral research associates will be encouraged to interact with PIs at all institutions, whether for scientific consultation associated with project activities, for professional development with respect to professional opportunities, or for advice on preparing papers for publications or presentations at professional meetings.

Orientation – At the time each research associate begins her/his appointment, orientation will include an in-depth conversation with the PIs at the host institution. The topics discussed will include: (a) expectations with respect to the degree of independence the research associate requires, (b) effective work habits and interaction with coworkers, (c) productivity including the importance of scientific publications and presentations, (d) documentation of research methodologies, and (e) an introduction to resources available at each campus (e.g., [http://postdocs.ucdavis.edu](http://postdocs.ucdavis.edu)).

Professional development – The research associates will be provided with a range of opportunities to develop the skills, knowledge, and experience needed to excel in their careers. They will receive training in scientific ethics and responsible conduct of research, both through informal instruction associated with research, data handling, data analysis, manuscript and presentation preparation, and dissemination of results and through formal workshops available at all home institutions. We will also encourage them to affiliate with one or more professional societies in a discipline relevant to their research interests, and we will encourage them to explore the challenges and opportunities of a range of career choices ranging from Ph.D.-granting institutions through liberal arts colleges, government agencies and NGOs. Past students and post-docs hold a variety positions including research scientists at botanical gardens, associate director of a bioinformatics facility, and scientists at NGOs or government agencies as well as traditional academic jobs at research universities and teaching colleges. To enhance their preparation for teaching, we will also encourage them to lead a graduate seminar on a topic closely related to their research and we will invite them to present guest lectures in graduate and undergraduate courses we teach.

Publications and presentations – We anticipate that each research associate will be lead author on at least two papers as part of the project, as well as being included as authors on several other publications. The research associates will have primary responsibility for field and lab studies, and statistical and phylogenetic analyses, reviewing relevant literature, and drafting manuscripts, but they will be guided by one or more of the PIs, and all of the PIs will be available for advice. We will also encourage the research associates to present results of their work at our annual project meetings and professional meetings (funding for this travel is included in our budgets).

Scientific interaction – In addition to participating in regular, weekly lab meetings of the host laboratories, the research associates will participate in regular, monthly teleconferences involving two or more PIs at different institutions. They will also help lead the annual project meetings that will be held in South Africa and at the three host institutions; these will be intensive collaborative work sessions that will cement strong contacts across the research team that will have lasting value for the postdocs as they build their careers. They will be encouraged to contact any of the project participants – PIs, post-doctoral research associate at the other institution, field technician/outreach assistant, graduate student, lab members – at any time for advice on experimental protocols, data analysis, interpretation, professional development, or any other aspect of their scientific work.

Assessment of the mentoring plan – The mentoring plan will be judged a success to the extent that the post-doctoral research associates are able to achieve the career goals they have set for themselves in the years immediately following their involvement with the project.