

# Project Description

## I. Introduction

The focus of our research is on the evolutionary history of the *Linum* (flax), an economically important genus that also displays an interesting biogeographic distribution with disjunct species groups in Eurasia, Africa, and the Americas, as well as several island endemic species in the Galapagos Islands and Madagascar. Botanists have been intrigued by widely disjunct plant taxa for over a 150 years (Gray, 1846; Hooker, 1860). Detailed examinations of the relationships of extant and fossil disjuncts, combined with expanding knowledge of the Earth's geological and climatological history, have resulted in numerous hypotheses that seek to explain historic and current patterns of biodiversity by invoking the processes of migration, vicariance, dispersal, speciation, and extinction (Raven, 1963; Thorne, 1972; Raven and Axelrod, 1974; Axelrod, 1975; Tiffney, 1985). More recently, molecular phylogenetic methods, which provide estimates of evolutionary relationships that are repeatable and independent of taxonomic knowledge, have made it possible to test these biogeographic hypotheses explicitly (Lavin and Luckow, 1993; Lavin, 2000). We can now identify true disjunctions as opposed to apparent disjunctions resulting from convergent evolution, and in some cases it is also possible to rule out or favor certain historical biogeographic hypotheses based on the compatibility of their predictions with molecular phylogenies (Donoghue et al. 2001; Davis, 2002).

Disjunct distributions are the product of a combination of external factors, such as dispersal agents and continental drift, and biological characteristics such as the production of highly dispersible propagules and the occurrence of mating systems that facilitate colonization of new areas. Baker's Law states that self-compatible species that are capable of autonomous self-pollination are more likely to be successful colonists than obligate outbreeding species (Baker 1955, Stebbins 1957). This is because the co-dispersal of an additional individual of a compatible mating type, which would be required for self-incompatible taxa, is not necessary for the establishment of new populations in such species. Though this "law" is borne out by numerous observations of widespread, disjunct, or invasive species (Baker 1967; Rambuda and Johnson, 2004), it has only rarely been tested in a phylogenetic context (Morrell et al. 2000).

We propose to investigate the interplay between biogeography and mating system in a disjunct lineage of *Linum* using a phylogenetic context provided by molecular data. By resolving the phylogeny of this lineage using complete taxon sampling and molecular phylogenetic techniques, we will be able to reconstruct its biogeographic history and determine the importance of dispersal, migration, and vicariance in the diversification of the flaxes. Concurrently, estimation of the breeding systems of these species will allow us to examine the veracity of Baker's Rule by looking for correlation between the occurrence of self-compatibility and dispersal within this clade.

## II. Background to Proposed Research

### 1. The genus *Linum*

*Linum* is the largest genus in the angiosperm family Linaceae, with approximately 180 recognized species distributed throughout the temperate and subtropical regions of the world. *L. usitatissimum* has been cultivated for nearly ten thousand years as a source of fibers that are woven into linen, as well as seed oils that are important components of paints, varnishes, and other products (Muir and Westcott 2003). Several other species of flax are cultivated, either for their seed oils, as fodder for animals, or as ornamentals (Hooker, 1872; Heywood, 1993). Seeds and leaves from many species have found use in traditional natural pharmacopias as laxatives, anti-inflammatories, and treatments for ailments such as sore throat, burns, and angina. Indeed, modern research has shown that flax seed oils contain quantities of  $\Omega$ -3 fatty acids, which are known to have numerous health benefits (Muir and Westcott 2003) and flax seed has become a popular 'health food' for human consumption. Potential anti-cancer compounds have been isolated from the seeds of several species (Weiss et al. 1975; Belma, 1998). In spite of the

agronomic and potential medicinal importance of this genus, the evolutionary relationships among flax species are largely unknown.

Flaxes are typically herbaceous perennials inhabiting open, temperate habitats throughout the world. The genus is traditionally divided into five sections, *Linum*, *Dasylinum*, *Syllinum*, *Cathartolinum*, and *Linopsis*, based on morphological characters such as flower color, the presence/absence of stipular glands, trichome distribution, presence/absence of staminodia, and leaf arrangement. A preliminary study of the phylogeny of *Linum* using DNA sequence data from 40 representative species identified two major lineages in the genus (McDill and Repplinger, in prep). The first comprises representatives of the blue-flowered flaxes: sections *Linum* (which includes *L. usitatissimum*) and *Dasylinum*. These sections are essentially confined to the Old World (only 5 species of blue flaxes are native to areas outside of Eurasia).

Sister to the blue flaxes is a yellow/white-flowered lineage that includes the remainder of the genus *Linum* (sections *Linopsis*, *Syllinum*, and *Cathartolinum*), as well as the members of four taxa that have traditionally been recognized as segregate genera (*Hesperolinon*, *Sclerolinon*, *Radiola*, and *Cliococca*) but are shown to be nested within *Linum*. This “yellow-flowered” lineage, the subject of the research proposed here, exhibits an interesting disjunct biogeographic distribution, having diversified extensively in the Mediterranean region (25 spp.), South Africa (14 spp.), South America (20 spp.), and North America (56 spp.), and with outlying species in the Galapagos islands (2 spp.), Madagascar (3 spp.), East Africa (2 spp.), and India (1 spp.) (Rogers, 1982). With its high species diversity on four continents apparently derived from a single ancestral species, this group of flaxes must have experienced several intercontinental exchanges in its history.

## 2. Biogeography

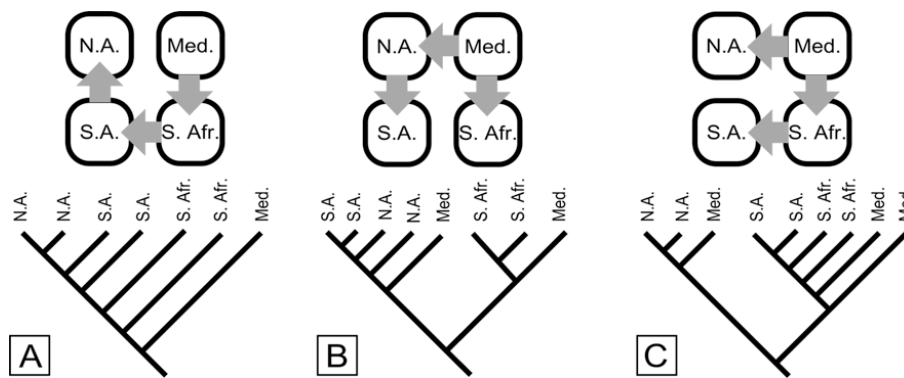
Most species of the yellow-flowered lineage are native to the temperate regions of Europe, Africa, and the Americas. Several biogeographic hypotheses have been proposed that seek to explain such warm-temperate disjunct distributions in plants. In the Southern Hemisphere, transatlantic disjunctions between Africa and South America are generally explained in terms of vicariance due to the breakup of Gondwana (Raven and Axelrod 1974), or as the result of long-distance dispersal by animals, wind, or on floating rafts of debris (Thorne, 1972; Houle, 1998, Renner, 2004). In the Northern Hemisphere, transatlantic disjunctions imply long-distance dispersal (Raven, 1963), Gondwanan vicariance followed by dispersal into the Northern Hemisphere (Davis et al., 2002), or migration across ancient North Atlantic or Beringian land-bridge connections followed by isolation when the bridges became discontinuous (Thorne, 1972; Raven and Axelrod, 1974). The Boreotropics Hypothesis (Tiffney, 1985; Wolfe, 1975) postulates the movement of tropical and warm temperate taxa across a North Atlantic land bridge during the Eocene and Oligocene (55 to 25 mya). This idea has been used to explain modern ampho-Atlantic disjunctions of taxa whose age or phylogenetic relationships are incompatible with Gondwanan vicariance (Lavin and Luckow, 1993; Davis et al., 2004).

Amphi-tropical disjunctions, in which closely-related species are found in temperate areas of the Northern and Southern hemispheres but are missing from the intervening tropical regions, are observed in 56 genera (including *Linum*) in both the Old and New World (Thorne, 1972). These are generally considered the products of long-distance seed dispersal, with migrating shorebirds frequently implicated as the agents of transportation (Raven 1963). This distribution could also be the product of a historically continuous distribution of species which became disjunct due to extinction in the tropics.

The quadripartite distribution of yellow-flowered *Linum* implies that several trans-Atlantic and trans-tropical events have taken place in the course of its diversification, but the application of any explicit hypotheses as explanations for the distribution has been purely conjectural. Rogers (1982) proposed that this lineage initially diversified in the Mediterranean region and that New World species were derived from South African ancestry. This scenario was based on patterns of morphological similarities among species, and no mechanism for the disjunctions were posited. Raven and Axelrod (1974) stated that the trans-Atlantic distribution of Linaceae was influenced by the breakup of Gondwana, but this conclusion was based on the age of pollen fossils of *Ctenolophon* (Gemeraad et al. 1968), a genus once placed in Linaceae but that is now known to be only distantly related to it (Savolainen et al., 2000).

The lack of any fossil record for *Linum* prevents us from using the timing of geological events, such as the continuity of the North Atlantic Eocene land bridge or the breakup of Gondwana, as indicators of which geological processes may have played a role in the diversification of the genus. More recent long-distance dispersals could provide a ready explanation for the distribution of linums. However, the fruits of all flaxes are dehiscent capsules and the seeds are large (2-5 mm in length) and smooth, with no obvious adaptations for dispersal. The applicability of such hypotheses can be tested by comparing the phylogenetic expectations that they imply to an independent phylogeny derived from molecular data.

Any historical biogeographic scenario has certain implications for the phylogeny of the taxa involved. While many alternative scenarios exist that could explain the combined amphi-Atlantic and amphi-tropical disjunctions in *Linum*, three simplified scenarios and their concomitant ‘expected’ phylogenetic resolutions are presented in Figure 1. The expectation of Rogers’ hypothesis is shown in Figure 1A, with North American (N.A.) and South American (S.A.) taxa comprising a monophyletic group that is nested among South African (S. Afr.) taxa. The basal position of Mediterranean species would support the hypothesis that this lineage initially diversified in that region. Figure 1B represents a scenario involving a northern hemisphere transatlantic disjunction which would be consistent with the movement of *Linum* across a North Atlantic land bridge or dispersal from Europe to North America. In this scenario, amphi-tropical disjunctions are the product of independent colonizations of South America and South Africa from North America and the Mediterranean, respectively. Figure 1C represents a more complex scenario, with independent Northern and Southern trans-Atlantic events.



**Figure 1:** Hypothesized dispersal patterns and compatible phylogenetic resolutions. **A.** A single origin for New World *Linum* from South African ancestry, with dispersal first into South America and from there into North America. **B.** A single origin for New World *Linum*, with North America colonized first, perhaps via a Boreotropical land bridge. **C.** Independent origins of North and South American *Linum*, from the Mediterranean and South Africa respectively.

### 3. Breeding System Evolution and Biogeography

The advantages of outcrossing in plants have been recognized for over 200 years (Knight, 1799, Darwin 1868, 1876), and inbreeding species have often been considered to be evolutionary dead-ends (Stebbins 1970). In spite of the advantages of more vigorous offspring and increased adaptability in outcrossing species, the shift from outcrossing to self-pollinating inbreeding systems is extremely common in plants (Stebbins, 1970; Wyatt, 1983; 1986, 1988). A significant factor expected to contribute to this trend is selection for reproductive assurance: the need for self-pollination in situations where outcrossing would be limited due to mate and/or pollinator scarcity (Jain, 1976; Schoen et al., 1996). Mating system is linked to biogeography through Baker’s Law, which states that self-pollinating taxa are more likely to be successful colonizers due to their independence from the need for mates and pollinators which may be absent in newly-colonized areas (Baker 1967). A prediction based on this law is that in a biogeographically dispersed group of plants such as *Linum*, self-compatibility will have evolved in the same lineages that have colonized new continents or islands, particularly if the colonization is the result

of long-distance dispersal, which would preclude the concerted movement of pollinators or populations of compatible mating types.

Heterostyly, a dimorphic self-incompatibility system that promotes outcrossing, is known to occur in 15 species of yellow-flowered flaxes, all of which are confined to the Old World, and mostly to the Mediterranean region (Ockendon, 1968). Two South African taxa are heterostylous (Rogers, 1981). Though the heterostylous species of *Linum* are well-studied (Dulberger, 1974; Murray, 1986; Wolfe, 2001; Nicholls 1985a, 1985b), mating systems in the remaining species are essentially unknown. The distribution of heterostyly exclusively in the Old World may be a result of Baker's Law: only homostylous (and presumably at least partially self-compatible) lineages have colonized, and diversified in, the New World. But testing of this hypothesis awaits information on breeding systems of the remaining homostylous species found on all four continents.

### III. Research Goals

1. Expand our current taxon sampling to include all members of the yellow-flowered lineage of *Linum* that has diversified in the Mediterranean region, Africa, and the Americas.
2. Resolve the phylogeny of this lineage using molecular phylogenetic methods and DNA sequence data from multiple chloroplast markers and the nuclear ITS.
3. Reconstruct the biogeographic history of this lineage by using parametric bootstrapping and other methods to determine the ancestral area and test alternative biogeographic hypotheses.
4. Investigate the relationship (or correlation) between breeding system and intercontinental disjunctions within this lineage by estimating the breeding system of each species.

### IV. Experimental Methodology

#### 1. Taxon Sampling

Our goal is to include all species of yellow-flowered linums in our datasets, as well as representatives of the segregate genera (*Hesperolinon*, *Sclerolinon*, *Radiola*, and *Cliococca*) that are nested within *Linum*. Such complete sampling is necessary because accurate phylogenetic reconstruction is highly sensitive to taxon sampling (Rannala et al. 1998). Biogeographic analyses also require exhaustive sampling, since any unrepresented species could potentially represent unaccounted-for dispersal or vicariance events, depending on their phylogenetic placement. Due to the large number of species to be sampled, and their far-flung distributions, this research relies heavily on sampling DNA from herbarium collections; collection of fresh material for each species would require many years. To date, we have succeeded in isolating useable DNA (as evidenced by successful amplification of selected chloroplast markers and/or ITS) from existing specimens representing approximately 75% of our target sample. In order to attain our goal of sampling all 120 yellow-flowered species of *Linum*, funds requested here are directed primarily toward collecting species for which recent specimens suitable for DNA extraction are not otherwise available. These include 11 South American, 7 South African, and 11 North American species which are rarely collected and not well represented in herbaria. Leaf material for DNA extraction will be collected in silica gel desiccant for preservation until DNA extractions are performed. Voucher specimens of all new collections will be deposited at major herbaria in their home countries, with duplicates placed at the University of Texas at Austin (TEX). Adequate samples for all of the yellow-flowered species from the Mediterranean region are available from existing collections.

To root the phylogeny of this lineage, outgroups representing 13 species from the blue-flowered clade of *Linum* as well as representatives of three other temperate genera (*Anisadenia*, *Reinwardtia* and *Tirpitzia*) and one tropical genus (*Hugonia*) of Linaceae will be sequenced and included in analyses. Tropical members of Linaceae have been previously determined to be the basal members of the family in phylogenetic analyses of the angiosperm phylogeny gene *rbcL* (McDill, in prep.).

#### 2. Laboratory methods

DNA isolation from silica-dried and herbarium specimens follows a small-scale extraction protocol modified from the Doyle and Doyle method (Loockerman and Jansen, 1996). Regions of the chloroplast and nuclear genomes are amplified by PCR and sequenced bi-directionally via dye-terminator cycle sequencing using BigDye. All sequence data are collected using an MJ BaseStation fragment analyzer, and sequence contigs are assembled and edited in Sequencher 4.0 (Gene Codes Corporation, Ann Arbor MI).

Six DNA sequence datasets are currently in development, including the nuclear ITS and five regions of the chloroplast genome which we have found to be of phylogenetic utility in *Linum* (*trnL* intron, *trnL-F* intergenic spacer, *trnK* 3' intron, *psbA-F* spacer, and the *ndhF* gene). Additional markers may be sequenced in order to increase phylogenetic resolution.

Paralogous copies of ITS have not so far been detected from species of *Linum* (M. Reppinger, pers. comm.), and phylogenies based on direct-sequenced ITS data are congruent with those of chloroplast markers. However, the possibility exists that non-homologous copies of the ITS could be amplified as the dataset is expanded to completion. This is a significant possibility particularly in the case of tetraploid and hexaploid species of *Linum* reported in North America, South America, and South Africa. For this reason, all ITS PCR products will be cloned and at least five clones from each species will be sequenced to assess infraspecific variation.

### 3. Phylogenetic Methods

Sequence alignments are performed by automated alignment algorithms in Clustal X (Thompson et al. 1997), followed by manual adjustment in MacClade (Maddison and Maddison, 2000). Phylogenetic analyses utilize the software program MrBayes 3.0 (Huelsenbeck and Ronquist, 2001) for Bayesian inference of phylogeny, and PAUP\* 4.0 (Swofford, 2003) for inference using parsimony and maximum likelihood criteria. For likelihood-based analyses in Mr. Bayes and PAUP\*, an appropriate model of molecular evolution, selected using the likelihood ratio test as implemented in ModelTest (Posada and Crandall, 1998), will be applied to each marker. Data sets for each marker will be analyzed independently, and will be tested for combineability using the ILD test (Farris et al. 1994, implemented in PAUP\*) and the BIONJ method (Zelwer and Daubin 2004). Non-parametric bootstrapping (Felsenstein 1985), decay analysis (Bremer 1988), and Bayesian posterior probabilities from MrBayes will be taken as measures of support for clades.

### 4. Biogeographic analyses

Once phylogenetic sampling is complete, the biogeographic history of the yellow-flowered lineage of *Linum* will be investigated using a variety of methods for determining its ancestral area and for testing the applicability of different biogeographic scenarios. Biogeographic analyses will be performed on each marker phylogeny separately, as well as on phylogenies resulting from analysis of appropriately combined datasets.

We are primarily interested in the amphi-Atlantic and amphi-tropical events in the history of these flaxes. Since no species of *Linum* occurs in more than one of the areas involved, the distribution of each species will be defined simply and unambiguously as the major geographic area to which it is endemic. Most species in the yellow-flowered lineage can thus be scored as North American, South American, Mediterranean, or South African. The endemic species of Madagascar, the Galapagos Islands, India, and East Africa will also be scored as having distinct distributions.

To determine the number of amphi-Atlantic and amphi-tropical events required to explain the current distribution of *Linum*, species distributions will first be treated as characters, mapped onto the phylogeny, and optimized using parsimony criteria in MacClade 4. The ancestral area calculation methods of Ronquist (1994) and Hausdorf (1998) will be utilized to test the long-standing assumption that this lineage of *Linum* initially diversified in the Old World Northern Hemisphere (Rogers 1982). Dispersal-vicariance analysis (Ronquist 1997a) as implemented in the computer program DIVA (Ronquist 1997b) will also be used to optimize dispersal and vicariance events on the phylogeny. All of

these methods require fully-resolved phylogenies for optimization of distributional characters. In the event of polytomies due to lacking or conflicting phylogenetic signal, analyses will be performed on a number of arbitrarily resolved phylogenies to determine the effect of alternate resolutions on the number of estimated dispersal and vicariance events.

Finally, a parametric bootstrapping approach known as the SOWH-test (Goldman et al., 2000), which does not require fully-resolved phylogenies, will be used to test how well different *a priori* biogeographic scenarios fit our data. This method utilizes constraint trees – hypothetical phylogenies that represent testable expectations for each scenario. For instance, Rogers hypothesis that New World *Linum* share a single origin from South African taxa implies that New World *Linum* will be monophyletic. To test whether our data support this hypothesis, a constraint tree holding New World species monophyletic will be imposed on parsimony searches in PAUP – this will impose the expectation of New World monophyly on our data. The parsimony score of our best estimate tree (found without the constraint enforced) will be subtracted from the score of the constrained tree to generate a test statistic. The significance of this difference is determined from a distribution of constrained-unconstrained tree length differences derived from analyses of 100 simulated datasets, which are generated according to the parameters of the substitution models calculated for our real data, using the computer program Seq-Gen 1.2.5 (Rambaut and Grassly, 2001). If the test statistic is greater than 95% of the distribution, the hypothesis is rejected and we conclude that our data is in fact estimating a different topology than that derived under the constraint.

## 5. Mating System Evolution

Species' mating systems, if unknown from previous research, will be estimated via pollen/ovule ratios (Cruden 1977). It must be noted that pollen/ovule ratios are only surrogates for definitive experimental mating-system determinations, but this is the only practical method for estimating mating systems throughout this diverse lineage of plants. To reduce destructive sampling from individual herbarium specimens, pollen will be counted from two anthers per flower, from five flowers (each from a different individual when possible) per species. The average number of pollen grains per anther will be multiplied by the number of anthers per flower and the product divided by the number of ovules (10 in all species of *Linum*) to derive the pollen/ovule ratio. The putative mating system for each species will be designated according to Cruden (1977). Mating systems will also be estimated for representatives of the segregate genera nested among the yellow-flowered linums, and for the outgroup taxa. Because factors other than mating system can influence the number of pollen grains produced (Kearns and Inouye 1993), estimates from P/O ratios will be supplemented with experimental breeding system determinations performed in the greenhouse on species for which seeds are available. The collecting trips proposed here will also allow collection of seeds from additional species, to augment the greenhouse sample. Cross and self-pollination manipulations for determination of breeding systems in greenhouse-grown populations will follow Kearns and Inouye (1993). Correlation of dispersal events reconstructed by DIVA with putative self-pollinating lineages will be tested using the Correlated Changes Test implemented in MacClade 4.0.

## V. Preliminary Results

Figure 2 presents the results of a Bayesian analysis of 3 chloroplast markers and the nuclear ITS combined, and includes data from 53 members (slightly less than 50% of our target sample) of the yellow-flowered lineage (the clade marked with the filled star). The analysis is rooted with members of the blue-flowered sections of *Linum* as well as representatives of other temperate and tropical genera of Linaceae. Although the phylogeny lacks resolution at certain critical nodes, some preliminary conclusions are possible. Both Hausdorf's and Ronquist's methods of ancestral area calculation indicate that Eurasia is the area in which *Linum* and its yellow-flowered lineage initially diversified, no matter the resolution of the indicated polytomies. Within the yellow-flowered group, a "Eurasian" clade is sister to a clade containing New World and South African species. Lack of resolution and support within that New

World/South African clade precludes definitive conclusions concerning the relationships of those areas. However, it appears that the initial biogeographic event in this lineage was colonization of North America by Eurasian ancestors, either by dispersal or perhaps movement across a North Atlantic land bridge. The SOWH-test rejects the monophyly of North American and North+South American species, falsifying Rogers' hypothesis that New World *Linum* share an African origin. Rather, these data indicate the South American and South African species probably originated from North America. Parsimony and DIVA reconstructions indicate a single colonization of South Africa, but we cannot yet determine whether the origin was from North or South America due to lack of resolution and missing taxa. Parsimony and DIVA reconstruct at least one and as many as four colonizations of South America by North American ancestors, and one possible movement of a South American lineage back to North America (represented by *L. kingii*).

While preliminary data have illuminated the evolutionary and biogeographic history of *Linum* from the remaining species of this lineage, and information from additional markers, will be necessary to resolve these questions. We are currently expanding sampling and sequencing additional markers to address this. Sampling of the missing South American and South African taxa is of particular importance, and expeditions are planned to collect material for these species pending funding.

To assess the suitability of pollen/ovule ratios as a proxy for breeding system, we determined the ratio in *L. berlandieri*, a North American species, and also determined its mating system via experimental manipulation. Both methods indicate that this species is autogamous and sets seeds primarily through self pollination ( $P/O=28.5\pm 2.1$ ; full seed set from unmanipulated bagged flowers). The distribution of heterostyly in the preliminary phylogeny indicates the evolutionary lability of this characteristic, with numerous independent losses or gains of heterostyly occurring in the Eurasian yellow-flowered clade, and an independent derivation of heterostyly in two South African taxa. Mating systems are yet to be determined from the remaining taxa.

## **VI. Significance and Broader Impacts**

The research proposed here is part of a larger project aimed at resolving the phylogeny of the entire genus *Linum*. The phylogeny of *Linum* will be an important contribution to two centuries of work on flax taxonomy, evolution, and agronomics. The phylogeny will provide a framework for revisionary taxonomic treatments and will guide future research in areas such as crop improvement and the discovery of new medicinal compounds. The phylogeny of the yellow-flowered lineage specifically will allow us to test historical biogeographic hypotheses and increase knowledge of the floristic relationships among continents. This project will contribute to the Simpson lab group's ongoing research program on South American biogeography.

The results of this research will be disseminated in major journals and at evolutionary and botanical conferences. We will also seek to disseminate our results to a more general audience in the form of popular articles and speaking engagements at venues such as the Ladybird Johnson Wildflower Center in Austin, TX. Through collaboration and experience, this research will strengthen Ph.D. candidate Joshua McDill's present skills with field and laboratory methods, phylogenetic and biogeographic analysis, and hypothesis testing. McDill and Simpson will also train several undergraduate students in laboratory and analytical methods in the course of this research. Finally, this work will promote international collaboration with local botanists in South America and South Africa, and all data, duplicate specimens of collections, and subsequent publications will be made available to herbaria and universities in each region.

**Figure 2.** Majority rule consensus tree resulting from Bayesian analysis of combined *trnL* intron, *trnL-F*, *trnK* 3' intron, and ITS (totalling approximately 2000 bp). Numbers above branches are bayesian posterior probabilities representing sampling of 4,000,000 MCMC generations after stationarity. ★ denotes the origin of the yellow-flowered lineage that is the focus of our research. ☆ is the common origin of all the species groups that have diversified in South Africa and the Americas. ● indicates heterostylous species. ○ indicates the single verified autonomous self-pollinating yellow-flowered taxon, *L. berlandieri*.

