Botanical Geospatial Diversity Cyberinfrastructure

Group Leader: *Brian J. Enquist (benquist@email.arizona.edu)*

Community group members: *Brad Boyle, Jeannine Cavender-Barres Rick Condit, Bob Peet, Mark Schildhauer, Nate Swenson (BIEN group) and David Ackerly and Walter Jetz (iPlant)*

1. The biological challenge the seed CI is aimed at addressing.

The plant sciences are currently witnessing an explosion of data. Addressing the most pressing questions in the plant sciences and global change biology are linked to three impediments to accessing these data: (*i*) taxonomic standardization, (*ii*) interaction and networking among data sources, and (*iii*) the ability to easily access, discover, and visualize patterns and processes within confederated databases and sources that range over differing spatial and temporal scales. The first two impediments are currently being addressed by iPlant via the BIEN working group [1] and the iPlant Tree of Life project, iPToL [2]. Here we address the third impediment – data discovery and visualization.

We propose to generate a novel CyberInfrastructure (CI) tool that will allow the botanical community and any interested user to access, explore, and visualize the rich spatial patterns within a network of core databases and data networks. Specifically, we aim to develop a 'geospatial discovery tool' or GSDT for the plant sciences. The GSDT will enable users to easily discover, explore, and visualize geographic distributions, phylogenetic and trait diversity.

2. The societal significance of the challenge.

Perhaps the most pressing challenge for the plant sciences, ecology, and global change biology is to assess ***how plant distribution, functioning, and diversity vary across space (and time) and ultimately how this diversity will respond to global change***. Our ability to predict how plant traits, functions, diversity, and distributions will change, however, remains limited [3, 4]. Much research that underlies ecology, evolution, comparative plant biology, and global change biology relies on spatial and comparative data [5-8]. For example, the geographic distributions of plant species reflect physiological tolerances [9, 10], evolutionary and climatic history, and offer insights into the traits that underlie adaptation [11-16] and the mechanisms involved in population divergence [17]. In order for biologists to predict how individual taxa and entire communities will respond to a changing world requires a tool for understanding how the physical environment influences where certain plant taxa grow, the mechanisms underlying local adaptation, and the controls on genetic and trait diversity, and ultimately what limits their geographic ranges.

3. A detailed description of the functionalities of the seed CI.

The secret to a successful CI tool will be ease of use, reliability, and practicality. We believe that a tool that enables quick viewing of spatial aspects of diversity would be enormously valuable and useful to applied and basic research as well as for educational outreach and public interest. We envision the GSDT to be a web service offered via iPlant. The font end will be a clickable and zoomable map – ideally based on Google Earth. The map will summarize but also offer a portal to a rich tapestry of botanical data. More importantly, the GSDT will enable, for the first time, discovery of how botanical diversity is distributed across the face of the earth. The GSDT will be developed in partnership with iPlant, iPToL, and the National Center for Ecological Analysis and Synthesis (NCEAS[1]). The GSDT will enable the user to easily discover, from a 'clickable' map, botanical trait, taxonomic, and phylogenetic diversity. It will utilizing an integrated database from the BIEN working group as well as phylogenetic output from the iPToL group. This database already consists of 25 million observations and consists of a generalized database schema that

[1] http://www.nceas.ucsb.edu/

can now integrate any additional additions of data. Underlying the GSDT will be five core datasets/streams. These include (*i-iii*) herbarium records, ecological plots, and plant functional traits (see Ref 1) that are currently accessible via a common database housed at NCEAS; (*iv*) phylogenetic information, tools, and output from the iPlant Tree of Life group; and (*v*) high resolution global climate data[2] as well as the next generation of state-of-the-art set of environmental layers that incorporate well-known but rarely used measures that have direct links to physiological processes like frost, water stress, growing season, soil properties, drainage properties that are currently being formulated at NCEAS[3]. The user will have the option of enabling summarized raster data and vector layers from the BIEN database. The GSDT will be accessed by two types of users – a 'basic user' who is only interested in summarized data while an 'expert user' user is interested in accessing the original data as well as performing additional analyses.

(1) ***Discovering Geographic Diversity –*** Here the user will have the ability to point to a geographic location or area. Clicking on a bounding box or user delineated area will return the following output

a. Output (Basic User) – The basic user is interested in the following questions "*what grows here?*" and/or "*how many species (or other taxa) grow here*?", or "*what traits characterize species that grow here*?". The output will return the following: (*i*) A list of all taxa that have been recorded; (*ii*) A taxa list with overlapping geographic ranges; (*iii*) A picture of the phylogenetic tree for all taxa that are found in that area; (*iv*) A list of traits (summary of trait means, variance etc.).

b. Output (Expert User) will return all output for the 'Basic User' plus the option to: (*i*) Download all original data; (*ii*) Analysis tools for inference; (*iii*) Climate data summary for that area, location; (*iv*) Tools to assess measures of diversity (taxa, phylogenetic, traits) but based on differing assumptions, experts etc. compare diversity between different points on the map.

(2) ***Phylogenies in Space*** – Here the user is primarily interested in putting clades on maps and discovering the associated clade traits and interactions with the physical environment. Here the user will point to a clade in a larger phylogeny generated by iPToL by clicking on a node. The user request would then query in the BIEN database to assess the coordinates of all taxa (and their associated trait and climate attributes) contained within that clade. The output would consist of,

a. Output (Basic User): (*i*) A map showing where that clade occurs - including clade diversity maps calculated from where range boundaries overlap; and summarized graphs and statistics of (ii) the climate and physical environmental attributes associated with that clade and (iii) trait values associated with that clade.

b. Output (Expert User) will return all output for the 'Basic User' plus the option to download all original data ase well as perform additional analyses as listed above. Here we are also interested in discussing the ability of the GSDT to be a downloadable tool that will access user specified datasets/databases.

4. Design, development and implementation time line.

We request seed CI support for 1 staff programmer/graduate student with expertise in GIS, database programming. Currently, we already have two excellent candidates in mind who could easy 'hit the ground running' with this project. We will be able to start on this project immediately. The current BIEN database 2.0 is already up and running and accessible at NCEAS. The creation of an early prototype of tool #1 could be available by spring 2011. Tool #2 will require close interactions between the BIEN group and the iPToL group but we envision that a prototype of this tool could be available within 6-8 months. Design and development of the GSDT will be done in collaboration with the BIEN and iPToL working groups as well as CI staff at iPlant.

---

[2] http://www.worldclim.org/methods

[3] NCEAS Project 12504 . Brian McGill et al. *Choosing (and making available) the right environmental layers for modeling how the environment controls the distribution and abundance of organisms* http://www.nceas.ucsb.edu/projects/12504

5. Management plan.

The technical staff person employed by this seed CI funding will primarily interact with the members of this proposal. The overall direction of this project will also be guided by a smaller primary advisory committee made up from the BIEN, iPToL, and plant adaptation working groups within iPlant as well as the CI development staff at iPlant. The role of this smaller core group will be to provide outside advice on the technical development and to facilitate 'cross pollination' of larger CI needs within iPlant and the botanical community. The smaller advisory group will be responsible for obtaining input from computational experts in spatial programming and GIS as well as from the user community. This will be done via personal contacts at our respective institutions, via the website hosting the GSDT, and from presentations at national meetings including The Ecological Society of America meetings, the Botanical Society of America meetings, and the Evolution meetings.

6. Bibliography

1. Enquist, B.J., et al. *The Botanical Information and Ecology Network (BIEN): Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity.* iPant White Paper [The iPlant Collaborative] 2009 [cited 2009; Available from: www.iplantcollaborative.org/sites/default/files/BIEN_White_Paper.pdf.

2. Sanderson, M., et al. *Assembling the Tree of Life for the Plant Sciences (iPTOL)* [The iPlant Collaborative] 2009 [cited 2009; Available from: http://www.iplantcollaborative.org/sites/default/files/iPTOL_White_Paper.pdf.

3. Williams, J.W. and S.T. Jackson, *Novel Climates, No-Analog Plant Communities, and Ecological Surprises: Past and Future.* Frontiers in Ecology and Evolution 2007(5): p. 475-482. .

4. Loiselle, B.A., et al., *Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes?* 2008. **35**: p. 105-116.

5. Brown, J.H., *On the Relationship Between Abundance and Distribution of Species.* American Naturalist, 1984. **124**(#2): p. 255-279.

6. Lomolino, M.V., B.R. Riddle, and J.H. Brown, *Biogeography*. 2006: Sinauer.

7. McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D. , Anderson, M.J., Benecha, H.K., Dornela, M., Enquist, B.J. , Green, J.L., He, F., Hurlbert, A.H., Magurran, A.E., Marquet, P.A., Maurer, B.A., Ostling , A., Soykan, C.U., Ugland, K.I. and E. P. White, *Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework.* Ecology Letters, 2007: p. 995-1015.

8. Woodward, F.I., *Climate and plant distribution*. 1987, Cambridge: Cambridge University Press.

9. Nobel, P.S. and E.G. Bobich, *Plant frequency, stem and root characteristics, and CO_2 uptake for Opuntia acanthocarpa: elevational correlates in the northwestern Sonoran Desert.* Oecologia, 2002. **130**: p. 165-172.

10. Osmond, C.B., et al., *Stress physiology and the distribution of plants.* Bioscience, 1987. **37**: p. 38-48.

11. Feild, T.S., T. Brodribb, and M. Holbrook, *Hardly a relict: Freezing and the evolution of vesselless wood in winteraceae.* Evolution, 2002. **56**(3): p. 464-478.

12. Swenson, N.G. and B.J. Enquist, *Ecological and evolutionary determinants of a key plant functional trait: Wood density and its community-wide variation across latitude and elevation.* American Journal of Botany, 2007. **94**(3): p. 451-459.

13. Sebastiaan, V., et al., *Life-history traits are correlated with geographical distribution patterns of western European forest herb species.* Journal of Biogeography. **34**: p. 1723-1735.

14. Condit, R., S.P. Hubbell, and R.B. Foster, *Changes in tree species abundance in a Neotropical forest: Impact of climate change.* Journal of Tropical Ecology ;, 1996. **12**(pt.2)): p. 231-256.

15. Eckhart, V.M., M.A. Geber, and C.M. McGuire, *Experimental studies of adaptation in Clarkia zantiana I. Sources of variation across a subspecies border. .* Evolution, 2003. **58**: p. 59-70.
16. Wilczek, A.M., et al., *Effects of Genetic Perturbation on Seasonal Life History Plasticity.* Science, 2009. **DOI: 10.1126/science.1165826**.
17. Pauwelsa, N.M., et al., *When population genetics serves genomics: putting adaptation back in a spatial and historical context.* Current Opinion in Plant Biology, 2008(2): p. 129-134.