***EvoNet*: A Phylogenomic and Systems Biology approach to identify genes underlying plant survival in marginal, low-Nitrogen soils**

Gil Eshel, [1*] (ge30@nyu.edu), Kranthi Varala,[2] Rodrigo Gutiérrez,[3] Dennis Stevenson,[4] Robert DeSalle,[5] W. Richard McCombie,[6] Jean-Michel Ané,[7] Heidi Kaeppler[7] and **Gloria Coruzzi**[1]

[1]New York University, New York; [2]Purdue University, West Lafayette; [3]Pontifica Universidad Católica de Chile, Santiago, Chile; [4]New York Botanical Garden, New York; [5]American Museum of Natural History, New York; [6] Cold Spring Harbor Laboratory, New York and [7]University of Wisconsin, Madison

http://evonet.org

Project Goals: This DOE BER sustainability project aims to identify the key genes and gene regulatory networks that enable "extreme survivor" plants to adapt and grow in marginal, extremely nitrogen (N) -poor soils in the hyperarid Atacama Desert in Chile. These "extreme survivor" species cover the main branches in flowering plants, and include 7 grass species of particular interest for biofuels. We focus on 28 "extreme survivor" Atacama species and compare their genomes to Californian "sister" species that live in a N-replete conditions in arid (27 species) or mesic (27 species) environments. Deep RNA-sequencing of these "triplet species" was used to fuel a novel phylogenomic analysis that helps identify individual genes that support the evolutionary divergence of the extreme survivors in Atacama Desert from their sister species in California. The genes thus identified will help to discover the mechanisms underlying physiological and developmental processes that allow plant survival in nitrogen-poor, dry soils. The genes and network modules so uncovered can potentially be translated to biofuel crops to greatly increase biomass and nitrogen use efficiency in marginal, low-fertility soils.

This collaborative project exploits the genomes of "extreme survivor" plants adapted to thrive in marginal, extremely Nitrogen (N) poor soils in the hyperarid Atacama Desert in the Chilean Andes. It uses a previously validated phylogenomic pipeline we developed called PhyloGeneious [1], which can identify genes that provide positive support to species divergence. By applying this phylogenomic pipeline to the gene sequences of these "triplet species", we can identify the genes that distinguish these "extreme survivors" in Atacama from their related "sister" species adapted to similarly dry regions in California (CA) not constrained by N and/or water availability and to mesic "sister" species growing in N and water (W) replete conditions. These "extreme survivor" species from the Atacama desert broadly cover the main branches in flowering plants, and therefore offer a wide range of genomic backgrounds within which the survival traits repeatedly arose i.e., multiple independent origins of trait.

Key to our phylogenomic approach is the "triplet species" sampling strategy. To maximize our ability to separate the trait-relevant signature from overall speciation events, our "triplet species" sampling will cover multiple independent origins of the low-N adaptive trait. In published studies, we showed that our phylogenomic pipeline could; i) identify genes that underlie convergent evolution of antioxidant synthesis in Rosids in a study of 150 plant genomes [1]; and ii) identify 100+ genes associated with the loss of Arbuscular Mycorrhizal (AM) symbiosis in the *Brassicaceae* [2]. We now extend this phylogenomic approach to the study of "extreme survivor strategies" as follows:

Aim 1. **Species collection and deep transcriptome sequencing**: (NYU, NYBG, Chile). ***Progress:*** We sequenced all 28 species collected in the Atacama Desert and 39 of their CA relative "sister" species (Table 1), with resulting average gene coverage of 87%, based on the BUSCO conserved single-copy orthologs existence assessment.

Aim 2. **Phylogenomic Analysis**: Perform phylogenomic analysis of 82 "triplet species" to identify genes that repeatedly support nodes that distinguish the extreme survivors in the Atacama Desert from their sister species in CA (AMNH, NYU). ***Progress:*** We performed phylogenomic analysis that includes each of the major plant lineages (*Poaceae* - 13 taxa; *Caryophyllales* – 9 taxa; *Lamiids* – 12 taxa; *Campanulids* – 25 taxa; *Fabaceae* – 11 taxa). This analysis identified hundreds of genes that provide phylogenetic support to the splits between the "extreme survivors" species that thrive in the Atacama Desert and their closest relatives from CA that grow in drought-adapted or mesic environments. We are further analyzing these gene sets to narrow down the best candidates for validation studies in a model grass.

Aim 3. **Network Analysis**: To combine phylogenomics (encoded protein sequence) and gene expression data to identify genes and network modules associated with adaptations to marginal, low-N soils (NYU, Chile). ***Progress:*** To exploit a comparative analysis of gene regulatory networks, we are currently developing a new module called **PhyloExpress** that extends the PhyloGeneious pipeline to include gene expression data. As a test data set, we are reconstructing, in parallel from sequence as well as gene expression data, the phylogeny of five diverse plant species (2 *Poaceae* - Maize, Rice, 2 *Rosids* - Soybean, Arabidopsis and 1 *Asterid* – Tomato).

Aim 4. **Functional Validation**: To functionally validate top-ranked candidate genes for low-N adaptation in Arabidopsis and Brachypodium (NYU, Chile, U Wisconsin). ***Progress:*** We have begun to transform Brachypodium with the most promising candidates from our preliminary analysis using our Atacama set and their closest sequence available sister species.

Table 1. Extreme survivor species in Atacama Desert, Chile, and their "sister" species in CA (Drought or Mesic). *Our project studies 28 triplet of species from Marginal (Dry +low-N, Atacama), Dry (Dry, California), and moist (Mesic, California) soils. All the Atacama species and 39/54 of the California species have already been sequenced (**RED**) while 15/54 California species are being collected (**BLACK**).*

Atacama Desert, Chile Extreme Survivor Species (Drought + low-N)	California "Sister" Species (Drought)	California "Sister" Species (Mesic)
Atriplex imbricata	*Atriplex lentiformis*	*Atriplex watsonii*
Mulinum crassifolium	*Sanicula crassicaulis*	*Conium maculatum*
Ambrosia artemisioides	*Ambrosia chamissonis*	*Ambrosia psilostachya*
Baccharis boliviensis and Baccharis tola	*Baccharis glutinosa*	*Baccharis salicifolia*
Trichocline caulescens	*Cirsium occidentale*	*Cirsium fontinale*
Chuquiraga atacamensis	*Coreopsis douglasii*	*Rudbeckia californica*
Parastrephia quadrangularis	*Grindelia hirsutula*	*Eriophyllum confertiflorum*
Senecio puchii	*Senecio californicus*	*Senecio mikanioides*
Tagetes multiflora	*Pectis papposa*	*Pluchea sericea*
Phacelia pinnatifida	*Eriodictyon tomentosum*	*Phacelia nemoralis*
Pycnophyllum bryoides	*Cerastium viride*	*Cerastium beeringianum*
Adesmia spinosissima	*Amorpha californica*	*Amorpha fruticosa*
Lupinus oreophilus	*Lupinus nana*	*Lupinus arboreus*
Lupinus subinflatus	*Lupinus hirsutissimus*	*Lupinus latifolius*
Aristida adscensionis	*Danthonia unispicata*	*Danthonia californica*
Bouteloua simplex	*Bouteloua curtipendula*	*Muhlenbergia filiformis*
Calamagrostis crispa	*Calamagrostis rubescens*	*Calamagrostis breweri*
Calamagrostis cabrerae	*Festuca californica*	*Festuca subuliflora*
Munroa decumbens	*Munroa squarrosa*	*Munroa utilis*
Nassella nardoides	*Nassella cernua*	*Nassella manicata*
Jarava frigida	*Stipa coronata*	*Stipa kingii*
Chorizanthe conmisuralis	*Chorizanthe palmeri*	*Rumex crispus*
Exodeconus integrifolius	*Lycium cooperi*	*Physalis lancifolia*
Fabiana denudata	*Nicotiana glauca*	*Petunia parviflora*
Solanum chilense	*Datura wrightii*	*Solanum douglasii*
Acantholippia deserticola	*Aloysia wrightii*	*Phyla nodiflora*
Fagonia chilensis	*Fagonia laevis*	*Tribulus terrestris*

References

1. Lee E et. al., A functional phylogenomics view of the seed plants. PLoS Genet 7(12):e1002411.
2. Delaux et. Al., Comparative phylogenomics uncovers the impact of symbiotic associations on host genome evolution. PloS Genet 10(7):e1004487.

The EvoNet grant is supported by the Genomics Science program within the Office of Biological and Environment Research in the DOE Office of Science.