Title of Proposal:

Cross-species genomic analysis of Photosystem II: Building connections from molecular structure to phenotype

Lead PI (Name, Title, Affiliation(s), email)

Carmela Rosaria Guadagno (Associate Research Scientist, University of Wyoming (**UW**) – Botany Department, cguadagn@uwyo.edu)

Co-PI (Name, Title, Affiliation(s), email)

Marilyn Gunner (Professor of Physics, City College of New York (**CCNY**), mgunner@ccny.cuny.edu)

Grant Administrator:

Farrell Rapp - Office of Research Economic Development (University of Wyoming)
307-766-2047
fgraf@uwyo.edu, research@uwyo.edu

Keywords: comparative genomics, plants, water dynamics, photosynthesis, crystal structures, molecular modeling

Abstract: The use of scale-invariant properties can improve our understanding of genome to phenome associations. This project uses the first principles of biophysics to develop cross-scale correlations between photosystem II related genes and drought phenotypes for agricultural species. We will perform a comparative genomic study searching for changes in the sequences to be imported into protein crystal structures for molecular modeling. Modeled water affinity across species will be correlated to existing phenotypic information to build associations for the ability of plants to grow and strive under water limitations.

Significance: Improving our cross-scale understanding of how plants respond to external water limitations (*i.e.,* drought) is essential to enhancing agricultural productivity as water uptake, transport, and utilization are the basis of yield formation. Water dynamics in and around photosystem II (PSII), a key protein in the photosynthetic electron transport chain, can be considered a scale-invariant property that follows the first principles of biophysics across space and time [1]. Water is a substrate of the protein, where it is oxidized to oxygen [2]. Characterizing PSII water dynamics can provide a long-needed connection between plant hydraulics, photosynthesis, and biomass production, necessary to implement current crop models for growth and productivity. Current advanced computer tools to investigate water bound in the protein, such as Multi Conformational Continuum Electrostatic (MCCE) method, allow for fast simulations of drying, but they require reliable data to see how individual proteins might connect to the resultant phenotypes [3,4]. We propose to build a cross-species comparative dataset to study PSII protein stability, from plant and chloroplast genomic databases. From the genes' sequences, changes in the peripheral amino acid chains will be built into existing structures *in silico*, and water binding affinity determined in MCCE. Sequence changes will be also correlated to leaf and whole plant phenotypes. In addition, given the advances of these models and their scalability to leaf and whole-plant physiology, similar approaches are transferable to novel drought resistant phenotypes in future crop breeding.

Aim 1: Use existing databases to correlate changes in the PSII domain that map onto different drought phenotypes. Plant water dynamics — the changes in plant water balance in time (diel and seasonal) and space (across different organs) — have remained recalcitrant to study due to the difficulty in connecting observed drought resistance phenotypes across wide genetic variation to the underlying molecular processes [5]. Following species-specific changes in water dynamics down to the sequence of key proteins and up to leaf and whole-plant levels can provide a far more integrated, process-informed, multi-level understanding of modeled yield prediction. Despite the development of rich genomic resources to study the changes responsible for ecological adaptations, reliable comparative data on PSII-related genes for desired phenotypes that would allow analysis at the protein level do not exist [6,7]. Numerous species-specific genome databases have been published but they are too often unstable and comparative work is still delayed by the different levels of functionalities of such platforms. Aim 1 of this proposal provides a well-curated, labeled dataset of cross-species sequences from plant and chloroplast databases in the PSII-related domains. Leveraging existing rich genomic resources [8-10] and comparative platforms such as CoGe [11], we will compare genomic structures from different genotypes of *Brassica*, building upon literature references and previous work from PI Guadagno and collaborators who characterized members of this species as heterogeneous in their water use efficiency [12,13]. We will generate and release a large, labeled dataset of PSII-related genes of diverse *Brassica* genotypes that will

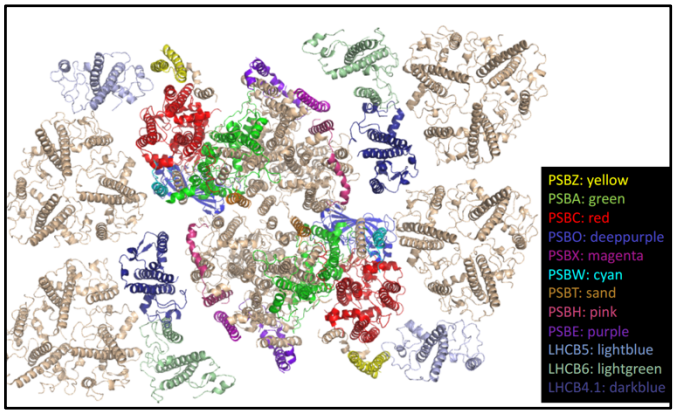

Figure 1: Cryo-EM structure of the PSII super-complex from A. thaliana highlighting the position of peptidic subunits (color-coded in the legend) with different likelihood of efficiency changes under for homology modeling and MCCE work

highlight informative changes in the amino acid chains. It is important to note that there might be RNA editing of some chloroplast-encoded proteins, which will affect the amino acid sequence that is predicted from the DNA sequence creating false positives for our follow-up molecular analyses. The *Brassica* comparative dataset will provide a crucial training dataset and a method to assess the performance of mutated amino acid chains and molecular models developed and trained to perform PSII water dynamics (Aim 2) for other crops. For this comparative analysis, we will then use sequences from other crops of interest, among others tomato, rice, soybean, spinach, and cotton. Species of interest are characterized by different drought resistance [14], and they are all heavily cultivated as annual crops, with the use of peculiar water strategies which are tuned to improve flavor, yield, or fiber quality, highly impacting current agriculture, and global food production. Species showing changes in proteins within, upstream or downstream of PSII on the photosynthetic pathway will be selected for further molecular structural analyses. The position of the PSII peptides will be mapped to the actual position of the protein in the PSII super-complex structures (Fig. 1).

Aim 2: MCCE simulations of PSII water dynamics and whole plant phenotypic correlation
Plants must adjust to survive current conditions and anticipate future ones, but contributors and associations required for the multiple levels of control needed for the cross-scale mechanisms that leads to a healthy plant is still absent. A process connection between plant hydraulics, photosynthesis, and biomass production is now necessary to implement current crop models for growth and productivity. We will follow the response to changes in water availability of PSII, a key protein for the photosynthetic electron transport, which requires the presence of both internal and external waters to be efficient and to channel the protons released during the water splitting reaction of photosynthesis out of the protein and protons into the quinone acceptor side [1,2]. This information will implement current simulations of PSII water dynamics with far-reach impacts on modeled predictions of crop productivity. We will test the hypothesis that drought phenotypes reflect changes in processes within PSII rather than upstream (protein synthesis) or downstream (function of proteins further down the photosynthetic electron transfer, chain). Co-PI Gunner and her students have used computer simulations to address PSII function starting from available 3-dimensional crystal structures using the MCCE program developed in the Gunner lab (Fig. 2). Preliminary simulations show water molecules released from PSII with changes in water chemical potential and the levels of water release support the use of osmotic potential calculations to define standard water potential values for PSII [15]. MCCE will access the role of water limitations on PSII function. The affinity of an individual water molecule to each binding site and the connection of pathways for proton entry (at the quinone) and exit (at the oxygen evolving complex - OEC) will be determined. We will work to connect how the calculations of a drying PSII by increasing pressure (*i.e.,* drier microenvironment) against the stress phenotypic data (*i.e*., drier macroenvironment) from PI Guadagno [16,17] during progressive soil drought and literature reports. The goal is to see if changes in redox potential or proton transfer pathways that controls PSII efficiency correlate with the changes in the water chemical potential.

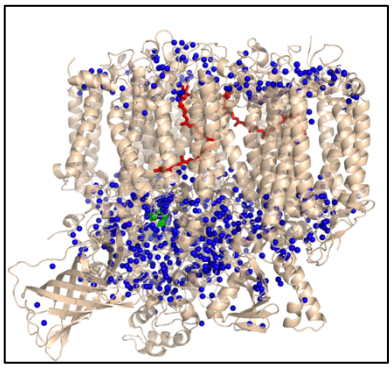

Figure 2: There are ~300.000 water molecules in and around a PSII super complex with unknown function (purple). In red quinones and in green the OEC.

There are over 500 plant and chloroplast genomes available across different platforms, providing us with rich sequence information. We will determine changes in amino acid sequences across crops spanning a large spectrum of drought resistance. The sequence information will be connected to the protein structure and through computer simulations of PSII function to plant phenotype by incorporating amino acid changes found in the genomic analysis into the PSII protein *in silico* and then calculate the water affinity, cofactor redox potential, and intra-protein proton pathways in the PSII homologs in MCCE. The amino acids, which cause interesting changes, will become suggested changes in future cross-species phenotypic experiments under drought stress. PSII represents an excellent choice for initial study as its activity and efficiency can be easily monitored by fluorescence across species and stress [1,2,18]. PSII activity can also be lowered by blockage of downstream reactions along the electron transport chain (*e.g.,* cytochrome b_6f and PSI) and we will use the simulation techniques worked out in this preliminary work on these proteins to connect nano-scale simulations to leaf water potential. In the big picture of the G2P research, our analysis will begin to bridge genotypic information to testable drought phenotypes using first principles of biophysics (*i.e.,* water dynamics and electron transport) *via* microscopic PSII function.

Furthering the aims of the AG2PI: A current challenge in G2P is to find an appropriate framework that advances our understanding of the hierarchy of life—breaking down the current silo mentality and allowing tools, data, and models to become cross-disciplinary. Our work will start building a common language based on a scale-invariant property (*i.e.,* water dynamics) that can describe the characteristic features of a living system at all spatiotemporal scales. By understanding the biological connections across different scales, we will be able to clarify and model crops' response to a changing environment. **Metrics of success:** We will publish two peer-reviewed papers; one publication will share our results on PSII genomics and structural correlations, while a second publication will emphasize the relevance of trans-disciplinary work to address complex biological questions and report on our framework to analyze cross-scale plant-environment interactions. We will also measure secondary metrics such as Github pull requests, access to the datasets, and the number of participants present at the hosted workshop.

Expected outcomes & deliverables: Our project will foster the first steps towards the development of a trans-disciplinary and cross-scale mentality for the new generation of plant scientists in AG2PI. Aim 1: We will provide a large, annotated comparative dataset of sequence data for photosynthetic organisms, with a particular focus on crop species, using currently available platforms (*e.g.,* ensembl, plantGDB) with varying precision for assessing the stability of PSII-related amino acid chains and evaluation by the plant science community. Species showing changes in proteins within, upstream or downstream of PSII on the photosynthetic pathway will be selected for further molecular structural analyses. Aim 2: We will use the sequence databanks to identify the variation of amino acid sequences in PSII and determine if they correlate to phenotypic differences across species, with particular attention to the species-specific response to water limitations in crops of high economical interest. The changes in sequence will be added to PSII structures *in silico* and changes in computed function will be determined using MCCE. We will provide parameters and modified structures *via* GitHub.

- Novel molecular models of water dynamics in and around PSII for photosynthetic organisms from algae to angiosperms, with focus on crops of interest
- A comprehensive annotated comparative dataset for PSII core and related proteins
- Improved understanding of photosynthesis and water dynamics influencing this vital process for crops under a changing environment

- Cross-scale engagement in photosynthesis research for integrated optimization of crop phenotyping and ultimately yield
- Hosted AG2PI workshop to engage with G2P community to create a common vernacular that emphasizes the realities of cross-scale biological connections

Qualifications of the project team: The PIs involved in this project span a range of skills and expertise making them uniquely positioned to carry out the proposed trans-disciplinary work and promote its implementation in G2P research. PI Guadagno is an early-career scientist whose specific area of research is focused on plant physiology and phenotyping methodologies with solid expertise in genome to phenome associations. Guadagno is a leading expert in the field of plant stress physiology with substantial experience in addressing how plants cope with water limitations, from mild to extreme drought and mortality. Her latest work leverages phenomic technologies in conjunction with genomics to study abiotic stress tolerance in crop plants. Co-PI Gunner is a leading expert in computational molecular biophysics who has studied the electron and proton transfer reactions in PSII and in bacterial reaction centers. Her work focuses on the thermodynamics of electron and proton transfer reactions. The team will carry out a comparative analysis of PSII sequences in various photosynthetic organisms to find differences that will be used to investigate changes in molecular structures significantly correlated to the resistance of different species to drought.

Proposal timeline:

Tasks	Dec 2021	Jan 2022	Feb 2022	March	April 2022	May 2022	June 2022	July 2022	Aug 2022	Sept 2022	Oct 2022	Nov 2022
Aim 1												
Build comparative cross-species sequencing dataset (UW & CCNY)												
Subset genes of interest coding for PSII-related proteins and photosynthetic pathways (UW)												
Aim 2												
Select amino acids to mutate in the crystal structures (CCNY)												
Molecular models and simulations (CCNY)												
Engagement with SRAP community (UW)												
Cross-scale correlation with phenotyping data in crops (UW)												
Data release and publication preparation												
Host joint online AG2PI workshop												

Engaging AG2P scientific communities & underrepresented groups: Future cross-disciplinary research will require having a workforce skilled in the analysis and interpretation of multiple modalities of data. To build engagement in the undergraduate population and advance their preparation for the jobs of tomorrow, we will invite students from both institutions to an online, AG2PI workshop to share the developed simulations and training data to the community and solicit feedback and ideas on models that can be applied to different environmental problems. We see this as a fantastic opportunity to build on the existing NSF-EPSCoR funded Summer Research Apprentice Program (SRAP) at UW [19]. This program was developed to support underrepresented minorities (ethnic, racial, gender minority, and first-generation in college) in STEM fields. CCNY is a public college with a diverse, low-income population, providing a real gateway to the middle class [20].

Bibliography/References cited

1) Butler, W.L. 1978. Annual Review of Plant Physiology, 29 (1): 345-378
2) Govindjee, G. 2004. Chlorophyll a fluorescence: a bit of basics and history. In: chlorophyll a fluorescence: a signature of photosynthesis. 1-42. Springer, Dordrecht, Netherlands
3) Song, Y., *et al.* 2009. Journal of Computational Chemistry, 30: 2231
4) Gunner, M., *et al.* 2011. Proteins, 79 (12): 3306-3319
5) Tardieu, F., *et al*., 2011. Current opinion in plant biology, 14 (3): 283-289
6) Chen F., *et al*., 2018. Frontiers in Plant Science, 9: 418
7) Song X., *et al*., 2021. Frontiers in Genetics, 12: 618
8) CpGDB - *http://www.gndu.ac.in/CpGDB/*
9) Plant GDB - *http://www.plantgdb.org/*
10) *https://uswest.ensembl.org/index.html*
11) *https://genomevolution.org/coge/OrganismView.pl*
12) Wang, D.R., *et al.,* 2019. Journal of Experimental Botany, 70 (9): 2561-2574
13) Greenham, K., *et al*., 2017. Elife, 6: e29655
14) Khanna-Chopra R., and Singh K., 2015. Drought Resistance in Crops: Physiological and Genetic Basis of Traits for Crop Productivity. In: Tripathi B., Müller M. (eds) Stress Responses in Plants. Springer, Cham.
15) *https://gunnerlab.github.io/Stable-MCCE/*
16) Guadagno, C.R., *et al*., 2017. Plant Physiology, 175: 223-234
17) Beverly, D.P. , *et al*., 2020. Frontiers in Forest and Global Change, 3: 589493
18) Guadagno, C.R., *et al*., 2021. Photosynthetica, 59 (SI): 49-61
19) *https://www.uwyo.edu/uw/news/2018/07/uw-apprentice-program-takes-high-school-students-to-lab-and-field.html*
20) Mwamba, J., 2020 *https://www.ccny.cuny.edu/news/brookings-study-affirms-ccny-economic-mobility-engine*