

AG2PI SEED GRANT - PROJECT FINAL REPORT

PROJECT NAME	Community engagement to improve standards and integration for genotype, phenotype, and environmental data for model and non-model plants
--------------	---

PROJECT PRINCIPAL INVESTIGATOR	TODAY'S DATE	PROJECT START DATE	DATE OF COMPLETION
Irene Cobo Simón, PhD	March 8, 2023	November 29, 2021	November 29, 2022
TEAM MEMBERS (co-PI, co-I, personnel)	COLLABORATORS		
Margaret Staton, PhD (Co-PI) Jill Wegrzyn, PhD (Co-PI) Victoria Burton (undergraduate biocurator - personnel) Meghan Myles (undergraduate biocurator - personnel)			

ACCOMPLISHMENTS

Please provide a short summary of the conclusions (both successes and failures) made from your project. Include a description of how this project will provide benefits to the agricultural genome to phenome community and, possibly, to a broader audience.

This project was a tremendous success in light of enhancing genotype to phenotype connectivity in plants. We refined and updated the FAIR TPPS (TriPal Plant PopGen Submit) module, updated the TreeSnap mobile application to collect FAIR trait data for landscape-based studies, and developed community resources that train the scientific community at all levels, from biocurators to TriPal database administrators.

Furthermore, some of our planned activities exceeded the objectives of this project (participation in the Genotype to Phenotype Working Group and collaboration with the Standards for Genetic Variation Working Group of AgBioData to expand CartograPlant and TPPS standards to make them interoperable with crop and germplasm databases).

The outcomes of this proposal are benefiting the international scientific community, not only to the agricultural genome to phenome community, but also to all the scientific community working with genotypic, phenotypic and environmental data from any plant species (e.g. CartograPlant supports landscape genomics as it applies to forest health, invasive species, and response to climate change). Thus, CartograPlant is intended to serve as a community resource for plant molecular ecologists, evolutionary biologists, quantitative geneticists, and plant breeders.

Having a centralized and up-to-date platform to integrate, visualize and analyze high-throughput biological data, such as CartograPlant, is key in the current big data era in plant biology. Our work focused on centralizing data, which is often decentralized across independent databases.

The georeferenced plants and metadata collected in CartograPlant, using ontologies and standards enforced through the TPPS collection, also enables meta-analysis. As a direct result of this funding, we are able to support advanced meta-analysis analysis for end users including population structure, association mapping, and landscape genomics analysis.

Products

Please list any products from this project. This may include (but not limited to) publication, concept/white paper, workshop, conference presentation, website, publicly available data or pipelines, etc. Reminder: you are required to make your products available to the broader stakeholder community using standard USDA practices, open source, FAIR, or other models.

ACTIVITY / PRODUCT	DESCRIPTION (include URL, if applicable)	OUTCOME / METRICS
(A) Expand TPPS adoption of MIAPPE standards to improve interoperability across a wider range of experimental designs and systems	TriPal Plant Popgen Pipeline (TPPS) has now adopted MIAPPE (Minimal Information About a Plant Phenotyping Experiment) standards to improve its interoperability across a wider range of experimental designs and systems. This new release of TPPS and the training materials can be found in this link: https://treegenesdb.org/tpps	We successfully updated MIAPPE v1.1 to v2.0. The documentation including the new features of the latest release of TPPS is openly available here: https://tpps.readthedocs.io/_/downloads/en/latest/p
(B) Develop ontology-based traits for landscape-based data collection WildType mobile application that can be directly ingested by TPPS	This objective was successfully completed with TreeSnap (the partner application) rather than WildType. The application has been updated and is publicly available here for Apple and Android platforms: https://treesnap.org/	63,610 phenotypes (of which 49 unique) from 9,102 georeferenced trees belonging to 130 species and 51 genera are currently available in CartograPlant from TreeSnap, ready for data visualization and analysis (https://cartograplant.org/)
(C) Develop training materials, workshops, and host data submission competitions	<p>This objective has been also successfully accomplished since a total of two workshops have been performed:</p> <ol style="list-style-type: none"> <i>1. Introduction to CartograPlant (GxPxE). 22/08/2022. 2.5 hours (online). Link with openly available training materials: https://docs.google.com/document/d/1Sa81D69GKj_MCoWk2MDdmpPOKWd7GAf9P2nyEBzYyF0/edit</i> <i>2. CartograPlant Workshop: Hands-On Training With an Analysis Framework Integrating Genotype, Phenotype, and Environmental Data at the PAG 30 International conference (San Diego, California, https://plan.core-apps.com/pag_2023/event/3cee484866e6d00c5e492240c4d5e8c5)</i> 3. Two data submission competitions, one announced via Twitter (@TreeGenes) (deadline August 22, 2022) and another performed during the 1st North American Forest Genetics Society Meeting (Asilomar, California) (June 14-16, 2022). 	<ul style="list-style-type: none"> - New features have been added to CartograPlant and disseminated to the workshop participants, including data quality filtering and visualization tools (e.g. histograms, Venn diagrams for phenotypes and genotypes, population structure analysis, multicollinearity calculation and visualization) - The workshops had a total of 34 and 31 participants, respectively - The data submission competitions led to a total of 1324 new plants submitted to CartograPlant

<p>(D) Participation of Dr. Irene Cobo Simón as Chair of the Genotype-Phenotype Working Group of AgBioData</p>	<ul style="list-style-type: none"> - Dr. Cobo as Chair of the Genotype-Phenotype WG (https://www.agbiodata.org/past-working-groups) - White paper entitled <i>Agricultural sciences in the big data era: omics data standardization, utilization and integration in crops</i> (in preparation) - Workshop at PAG 30 International conference entitled <i>The AgBioData Consortium: Challenges and Recommendations for FAIR Genetic, Genomic and Breeding Data</i>. Session entitled: <i>Challenges and Opportunities in Connecting Genotype to Phenotype Data</i> (https://plan.core-apps.com/pag_2023/event/3cee484866e6d00c5e492240c4d27482) - AgBioData monthly webinar presentation (March 1, 2023): https://www.youtube.com/watch?v=63bheER6eX4 	<p>The Genotype-Phenotype Working Group has a total of 12 active participants.</p> <p>The workshop had a total of 50 participants</p> <p>The monthly webinar had a total of 30 attendees</p>
<p>(E) Recruit and train an undergraduate student in biocuration, underrepresented in STEM, to continue our efforts in including important large-scale population studies into CartograPlant</p>	<p>The undergraduate students Victoria Burton and Meghan Myles were successfully trained in biocuration and contributed to upload genotypic, phenotypic and environmental data, and their associated meta-data, from a total of 29,089 and 44,319 plants, respectively, in CartograPlant (https://cartograplant.org/). Meghan now serves as the lead for a team of four undergraduate biocurators and has developed training material for curation best practices located here: https://gitlab.com/TreeGenes/data-curation.</p> <p>Both students are presenting their work on biocuration of <i>genotype to phenotype data</i> at UConn's Undergraduate Research Symposium as graduating seniors this coming April 2023.</p>	<p>The success of their training can be measured by the fact that Victoria Burton and Meghan Myles have contributed to include genotypic, phenotypic and/or environmental data, and their associated meta-data, from a total of 29,089 and 44,319 plants, respectively in CartograPlant (https://cartograplant.org/).</p>
<p>(F) Dissemination of the project results in International Conferences</p>	<p>Dissemination of the project results in four national (USA) and international conferences:</p> <p>1. Plant and Animal Genome Conference PAG 30, January 13-18, 2023. San Diego (California, USA)</p> <ul style="list-style-type: none"> - Talk: Emily Grau, Irene Cobo-Simón, Risharde Ramnath, Jill Wegrzyn. Cartograplant: Leveraging Ontologies for Landscape and Association Genetics. 	

2. Botany 2022, Anchorage (Alaska, USA), July 24-27, 2022

- **Talk:** Irene Cobo-Simón, Risharde Ramnath, Sean Buehler, Peter Richter, Charles Demurjian, Barnaly Pande, Emily Strickland, Victoria Burton, Nicole Butch, Alicia Abrams, Alex Lowe, Emily Grau, Abdullah Almsaeed, Nic Herndon, Margaret Staton, Jill Wegrzyn. CartograPlant: Cyberinfrastructure to improve plant health and productivity in the context of a changing climate

3. 1st North American Forest Genetics Society Meeting (Asilomar, California) (June 14-16, 2022):

- **Talk:** Emily Grau. *TreeGenes and CartograPlant: Resources for Forest Tree Genetics*

4. Plant Genomes Onlile (April 27-28, 2022):

- **Talk:** Irene Cobo-Simón, Risharde Ramnath, Sean Buehler, Peter Richter, Charles Demurjian, Barnaly Pande, Emily Strickland, Victoria Burton, Nicole Butch, Alicia Abrams, Alex Lowe, Emily Grau, Abdullah Almsaeed, Nic Herndon, Margaret Staton, Jill Wegrzyn. CartograPlant: Cyberinfrastructure to Improve Plant Health and Productivity in the Context of a Changing Climate.

5. Plant and Animal Genome International Conference (PAG) XXIX. Jan. 9-11, 2022. Virtual

- **Talk:** Irene Cobo-Simón, Nic Herndon, Margaret Staton, Emily Grau, Sean Buehler, Peter Richter, Risharde Ramnath, Charlie Demurjian, Abdullah Almsaeed, Jill Wegrzyn. Integrating, visualizing and analyzing tree environment, phenotype and genotype using CartograPlant, WildType and Tripal Galaxy. *Digital Tools and Resources, Session 3.*

- **Talk:** Irene Cobo-Simón, Risharde Ramnath, Sean Buehler, Peter Richter, Charlie Demurjian, Barnaly Pande, Emily Strickland, Alicia Abrams, Alex Lowe, Emily Grau, Abdullah Almsaeed, Nic Herndon, Margaret Staton, Jill Wegrzyn. CartograPlant: Cyberinfrastructure to Improve Forest Health and Productivity in the Context of a Changing Climate. *Forest Tree Session.*

- **Poster:** Irene Cobo-Simón, Nic Herndon, Margaret Staton, Emily Grau, Sean Buehler, Peter Richter, Risharde Ramnath, Charlie Demurjian, Abdullah Almsaeed, Jill Wegrzyn. Integrating, visualizing and analyzing tree environment, phenotype and genotype using CartograPlant, WildType and Tripal Galaxy

Audience

With whom has this work been targeted to and shared? Please describe how this project and its products have been disseminated to a community of interest. Include any outreach activity or information sharing as well as training or professional development opportunities provided in this project.

This project and its products have been disseminated to main four communities of interest (in bold and underlined). The detailed information of these dissemination activities is described in the following lines. It is also worth mentioning that all the training workshops, conferences, data submission competitions, monthly webinars and the weekly updates of CartograPlant and TPPS have been also disseminated to the international scientific community via Twitter, at @TreeGenes (<https://twitter.com/TreeGenes>):

Plant International Scientific Community:

- **Two training workshops** for plant scientists on plant data integration, visualization and analysis with CartograPlant (training materials openly available in gitlab):

1. *Introduction to CartograPlant (GxPxE)*. 22/08/2022. 2.5 hours (online). Link with openly available training materials

https://docs.google.com/document/d/1Sa81D69GKj_MCoWk2MDdmpPOKWd7GAf9P2nyEBzYyF0/edit

2. *Plant and Animal Genome Conference PAG 30*, January 13-18, 2023. San Diego (California, USA). *CartograPlant Workshop: Hands-On Training With an Analysis Framework Integrating Genotype, Phenotype, and Environmental Data* https://plan.core-apps.com/pag_2023/event/3cee484866e6d00c5e492240c4d5e8c5

- Dissemination of the project results in **five national (USA) and international conferences:**

1. *Plant and Animal Genome Conference PAG 30*, January 13-18, 2023. San Diego (California, USA)

- **Talk:** Emily Grau, Irene Cobo-Simón, Risharde Ramnath, Jill Wegrzyn. CartograPlant: Leveraging Ontologies for Landscape and Association Genetics.

2. *Botany 2022*, Anchorage (Alaska, USA), July 24-27, 2022

- **Talk:** Irene Cobo-Simón, Risharde Ramnath, Sean Buehler, Peter Richter, Charles Demurjian, Barnaly Pande, Emily Strickland, Victoria Burton, Nicole Butch, Alicia Abrams, Alex Lowe, Emily Grau, Abdullah Almsaeed, Nic Herndon, Margaret Staton, Jill Wegrzyn. CartograPlant: Cyberinfrastructure to improve plant health and productivity in the context of a changing climate

3. 1st North American Forest Genetics Society Meeting (Asilomar, California) (June 14-16, 2022):

- **Talk:** Emily Grau. *TreeGenes and CartograPlant: Resources for Forest Tree Genetics*

4. *Plant Genomes Onlile* (April 27-28, 2022):

- **Talk:** Irene Cobo-Simón, Risharde Ramnath, Sean Buehler, Peter Richter, Charles Demurjian, Barnaly Pande, Emily Strickland, Victoria Burton, Nicole Butch, Alicia Abrams, Alex Lowe, Emily Grau, Abdullah Almsaeed, Nic Herndon, Margaret Staton, Jill Wegrzyn. CartograPlant: Cyberinfrastructure to Improve Plant Health and Productivity in the Context of a Changing Climate.

5. *Plant and Animal Genome International Conference (PAG) XXIX*. Jan. 9-11, 2022. Virtual

- **Talk:** Irene Cobo-Simón, Nic Herndon, Margaret Staton, Emily Grau, Sean Buehler, Peter Richter, Risharde Ramnath, Charlie Demurjian, Abdullah Almsaeed, Jill Wegrzyn. Integrating, visualizing and analyzing tree environment, phenotype and genotype using CartograPlant, WildType and Tripal Galaxy. *Digital Tools and Resources, Session 3*.

- **Talk:** Irene Cobo-Simón, Risharde Ramnath, Sean Buehler, Peter Richter, Charlie Demurjian, Barnaly Pande, Emily Strickland, Alicia Abrams, Alex Lowe, Emily Grau, Abdullah Almsaeed, Nic Herndon, Margaret

Staton, Jill Wegrzyn. CartograPlant: Cyberinfrastructure to Improve Forest Health and Productivity in the Context of a Changing Climate. *Forest Tree Session*.

- **Poster:** [Irene Cobo-Simón](#), Nic Herndon, Margaret Staton, Emily Grau, Sean Buehler, Peter Richter, Risharde Ramnath, Charlie Demurjian, Abdullah Almsaeed, Jill Wegrzyn. Integrating, visualizing and analyzing tree environment, phenotype and genotype using CartograPlant, WildType and Tripal Galaxy

- **Two data submission competitions** to upload data in CartograPlant via TPPS, one announced via Twitter (@TreeGenes) (deadline August 22, 2022) and another performed during the 1st North American Forest Genetics Society Meeting (Asilomar, California) (June 14-16, 2022).

International scientific Community Working on Plant Databases:

- Dr Cobo is Chair of the Genotype-Phenotype Working Group of AgBioData. Outcomes:
 - White paper entitled *Agricultural sciences in the big data era: omics data standardization, utilization and integration in crops* (in preparation)
 - Workshop at PAG 30 conference entitled *The AgBioData Consortium: Challenges and Recommendations for FAIR Genetic, Genomic and Breeding Data*. Session entitled: *Challenges and Opportunities in Connecting Genotype to Phenotype Data* (https://plan.core-apps.com/pag_2023/event/3cee484866e6d00c5e492240c4d27482)
 - AgBioData monthly webinar presentation: <https://www.youtube.com/watch?v=63bheER6eX4>, <https://twitter.com/AgBioData/status/1631026980618481664>
- We are collaborating with the Standards of Genetic Variation Working Group of AgBioData to expand CartograPlant and TPPS standards to make them interoperable with crop and germplasm databases.

Tripal database administrators

Undergraduate students: We successfully trained two undergraduate student underrepresented in STEM in data biocuration (Victoria Burton and Meghan Myles)

CHALLENGES

Have you experienced any challenges or delays? Please provide the actions you took to resolve them, if possible.

The main challenge of the project was to accomplish the objective of generating agreements with two new journals to adopt TPPS as the FAIR framework of choice to ingest population genetic data, and generate a DOI at the time of publication. We contacted several journals and even sent reminder e-mails. Although two of them showed their interest (New Forests and Current Plant Biology), eventually they did not adopt TPPS. **In order to resolve this challenge, Dr. Cobo joined the *Applications in Plant Sciences (APPS) Journal* as a member of the Editorial Board on September 1, 2022** and she is being mentored by Dr. Gil Nelson (APPS Editorial Board) until August 2024 to gain experience in the editorial, peer-review and publication processes <https://bsapubs.onlinelibrary.wiley.com/hub/journal/21680450/homepage/editorialboard>. We expect that this experience will help us gain insight on the specific, current needs and objectives of plant journals to make their data FAIR and Open Access. Therefore, we will adapt TPPS features accordingly to fully meet the needs of plant journals, making TPPS more attractive for them, and thus, we expect to eventually accomplish this lacking objective.