

12-4-2020

## Mining Gems in Literature and Genomes: A Proof of Concept in *Artemisia tridentata*

Carlos Dave Dumaguit  
*Boise State University*

S. J. Galla  
*Boise State University*

Anthony Melton  
*Boise State University*

John M. A. Wojahn  
*Boise State University*

P. Martinez  
*Boise State University*

*See next page for additional authors*

---

**Authors**

Carlos Dave Dumaguit, S. J. Galla, Anthony Melton, John M. A. Wojahn, P. Martinez, and Sven Buerki

# Mining Gems in Literature and Genomes: a Proof of Concept in *Artemisia tridentata*

Dumaguit, C.D., Galla, S.J., Melton, A., Wojahn, J. M. A., Martinez, P., Buerki, S.  
Boise State University, Department of Biological Sciences



## Introduction

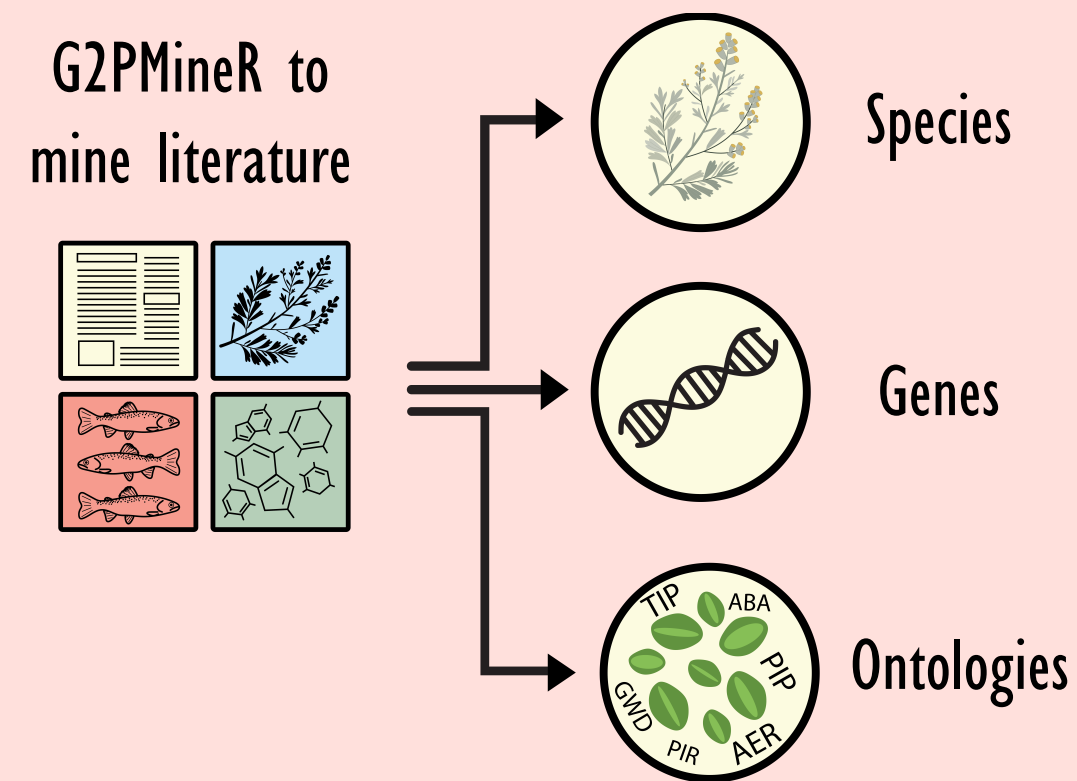
- Research predicts increased global temperatures and increased occurrences of longer droughts, which can have far-reaching impacts on crops and plant communities, including the imperiled big sagebrush (*Artemisia tridentata*)<sup>1-3,6,7</sup>
- Looking at the genetic underpinnings of drought brings insight into how plants cope with these conditions and plants' abilities to adapt to anthropogenic drought<sup>4</sup>
- While research on the genomic basis of plant resilience towards drought has been conducted in crops and model plants, fewer studies have evaluated natural plant communities
- Literature Mining provides efficient, customizable processes<sup>6</sup>

### Top Five Candidate Genes by Appearance in Literature

Gene Family	Ontology	Count
DREB (dehydration response element binding)	transcription factor	537
NAC	transcription factor	522
WRKY	transcription factor	344
MYB	transcription factor	321
SOD (superoxide dismutase)	antioxidant	229

Table 1.  
Each of the top 50 genes were manually searched in the literature for a direct relationship or empirical data linking them to plants and resistance to stress

## Process



- We used *G2PMineR*, a newly developed package to mine literature publications for genes related to drought and plant stress responses<sup>7</sup>
- After finding candidate genes, the top 50 genes/family of genes were manually validated for their relatedness to drought stress response
- One of the top candidates, the DREB (dehydration response element binding) family was analyzed more closely using a pipeline in *R* utilizing rBLAST and NCBI Gen Bank<sup>8</sup>
- DREB genes were pulled from dozens of species from NCBI Gen Bank

## Findings & Discussion

- The validation process refined one of the first estimates to determine how many genes are related to drought stress response in plants
- The abundance of transcription factors could lead future research into plant adaptability and susceptibility due to genomic variations among and within species
- DREB gene family matches were found in four scaffolds-three in *Phaseolus vulgaris* and one in *Glycine max*
- Our genome for *Artemisia tridentata* was incomplete (about half of the total) which could mean more gene matches in the future
- There was a lack of blast matches with DREB genes from species such as *Arabidopsis thaliana*, & *Oryza Sativa*
- Although not closely related by phylogeny, the relationship between legumes and sagebrush via potential drought response genes is an interesting exploration

## References

- <sup>1</sup>Dunn, R.J.H., Alexander, L.V., Donat, M.G., Zhang, X., et al. (2020). Development of an updated global land in-situ-based dataset of temperature and precipitation extremes: HadEX3. *Journal of Geophysical Research: Atmospheres*. <https://doi.org/10.1029/2019JD032263>
- <sup>2</sup>Stahle, D.W. (2020). Anthropogenic Megadrought. *Science* Vol 368, Iss 6488, 238-240.
- <sup>3</sup>Brodribb, T.J., Powers, J., Cochard, H., Choat, B. (2020). Hanging by a thread? Forest and drought. *Science* Vol 368, Iss 6488, 261-266.
- <sup>4</sup>Gupta, A., Rico-Medina, A., Cano-Delgado, A.I. (2020). The physiology of plant responses to drought. *Science* Vol 368, Iss 6488, 266-269.
- <sup>5</sup>Davies, K.W., Boyd, C.S., Beck, J.L., Bates, J.D., & Svejcar, T.J. (2011). Saving the sagebrush sea: An ecosystem conservation plan for big sagebrush plant communities. *Biological Conservation* 144, 2573-2584.
- <sup>6</sup>Jensen, L., Saric, J., & Bork, P. (2006). Literature mining for the biologist: from information retrieval to biological discovery. *Nature Reviews Genetics*, 119-129.
- <sup>7</sup>Wojahn, J.M.A., Galla, S., & Buerki, S. (2020) *G2PMineR* R Package. In preparation.
- <sup>8</sup>Melton, A. (2020) *GenomeMining* R Package. In preparation.
- Reddit User Thanatophobia666. [https://www.reddit.com/r/unixporn/comments/f1mb2g/oci\\_made\\_a\\_wallpaper\\_with\\_the\\_unix\\_logo\\_and/](https://www.reddit.com/r/unixporn/comments/f1mb2g/oci_made_a_wallpaper_with_the_unix_logo_and/)

## Future Directions

- More genes will be searched once the genome is complete to provide a full picture of *Artemisia tridentata* and its genetic tools to combat drought
- Research on natural plant communities could save on conservation efforts as most plant genetics focus on crop species and model organisms

