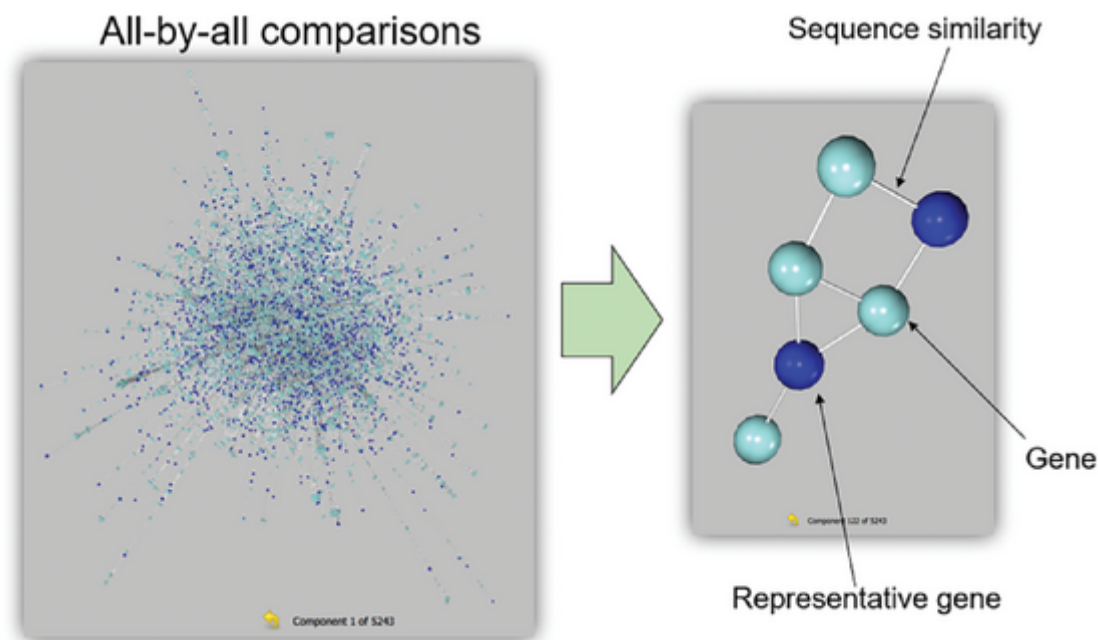# Predicting Gene Models Based on Evolutionary Conservation

December 18, 2023



*The occurrence of short sequence motifs, pinpointed from the re- sults of a GWAS, distinguished between white- and purple-flowered soybean accessions. Figure reproduced with permission from the following article by Sundaramoorthy et al. published in the Journal of the Korean Society for Applied Biological Chemistry: https://doi. org/10.1007/s13765-015-0077-z.*

With advances in DNA sequencing, it is easier to generate genome assemblies for any species. Genome annotation has not advanced at the same rate and many gene models, genomic regions that are transcribed and translated, predicted in individual plant genomes may be erroneous.

While large numbers of unique genes for species would suggest that de novo gene birth occurs at a high rate following speciation, studies suggest this is not the case, and protein-coding genes share sequence identity between species. Researchers applied the concept of evolutionary conservation to introduce a new method to support the accurate prediction of gene models. Through comparative genomic analysis, they produced a representative gene set of 15,345 gene models from 12 legume species and used it to identify potentially erroneous gene models in each species' annotation. If a gene did not have a match with the conserved set, the researchers considered it to have low confidence and to require further validation.

Using this new method, representative gene sets can be developed for any species and applied to support more accurate gene model predictions.

*prohibited.*