

# Gramene PanMaize: One-Stop Pan-Genome Browser for Exploring the Rich Genetic Diversity in Maize



URL: <https://maize-pangenome.gramene.org>

Lu, Zhenyuan<sup>1</sup>; Chougule, Kapeel<sup>1</sup>; Olson, Andrew<sup>1</sup>; Tello-Ruiz, Marcela K<sup>1</sup>; Wei, Sharon<sup>1</sup>; George, Nancy<sup>2</sup>; Papatheodorou, Irene<sup>2</sup>; Ware, Doreen<sup>1,3</sup>

<sup>1</sup>Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724; <sup>2</sup>EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SD, UK; <sup>3</sup>USDA ARS NEA Plant Soil & Nutrition Laboratory Research Unit, Ithaca, NY, 14853, USA

## Abstract

The Gramene project has developed pan-genome subsites, as a natural extension of the Gramene and Ensembl infrastructures. Each pan-genome site is dedicated to the study of individual crop groups (e.g., maize, rice, sorghum, and grape). The Gramene project (<http://www.gramene.org>) was a key player in sequencing the first maize reference B73, more recently the 25 NAM founders, and continues to be engaged with the community. The maize pansite (<https://maize-pangenome.gramene.org>), established in 2021, hosts reference assemblies for 26 maize accessions. Each maize accession is hosted in a separate genome browser, providing access to gene-based views with entry points via text-based searches. The B73 v5 genome serves as the reference assembly for anchoring expression, population and pathway views; transcript abundance across a gene model at different tissues and stages of development and views of paralog expression support decision making on candidate gene selection. Phylogenetic analyses are based on maize-specific gene trees built from the assigned canonical protein for each gene locus. The protein-based trees allow users to rapidly traverse between the different maize accessions, where alleles are considered orthologs in the trees, as well as other species; copy-number variations within maize can also be evaluated alongside lineage-specific gene expansions. Protein homology is viewable as amino acid alignments and as gene neighborhood conservation. We adapted these views to deploy community curation tools—for the homology tab—for users to flag potential structural annotation issues. More recent work has used these gene trees to build a pan-gene index used as the inputs to improve and extend the previous gene structural annotation workflow. Release 3 will include updated annotations and access to 8 maize genomes (4 European Flint, Mo17, W22, PH207 & A188), expression, and Maize ENCODE (NSF-IO5-1445025) related studies. Funded by USDA-ARS-8062-21000-041-00D.

Funding acknowledgement: United States Department of Agriculture (USDA)

## New Search and Quick Views

Search for *lipoxygenase 9 (lox9)* gene

## Gene search results categorized

- **Location** - Genomic context
- **Expression** - Baseline gene expression from EBI
- **Homology** - Customizable Ensembl Compara gene family trees with displays
  - Alignment Overview (full-length gene)
  - Multiple-Sequence Alignment (zoom into the amino acid level)
  - Neighborhood Conservation View (zoom out +/- 10 flanking genes)
- **Pathways** - Association to pathways in Gramene's Plant Reactome
- **Xrefs** - Cross-references to other databases

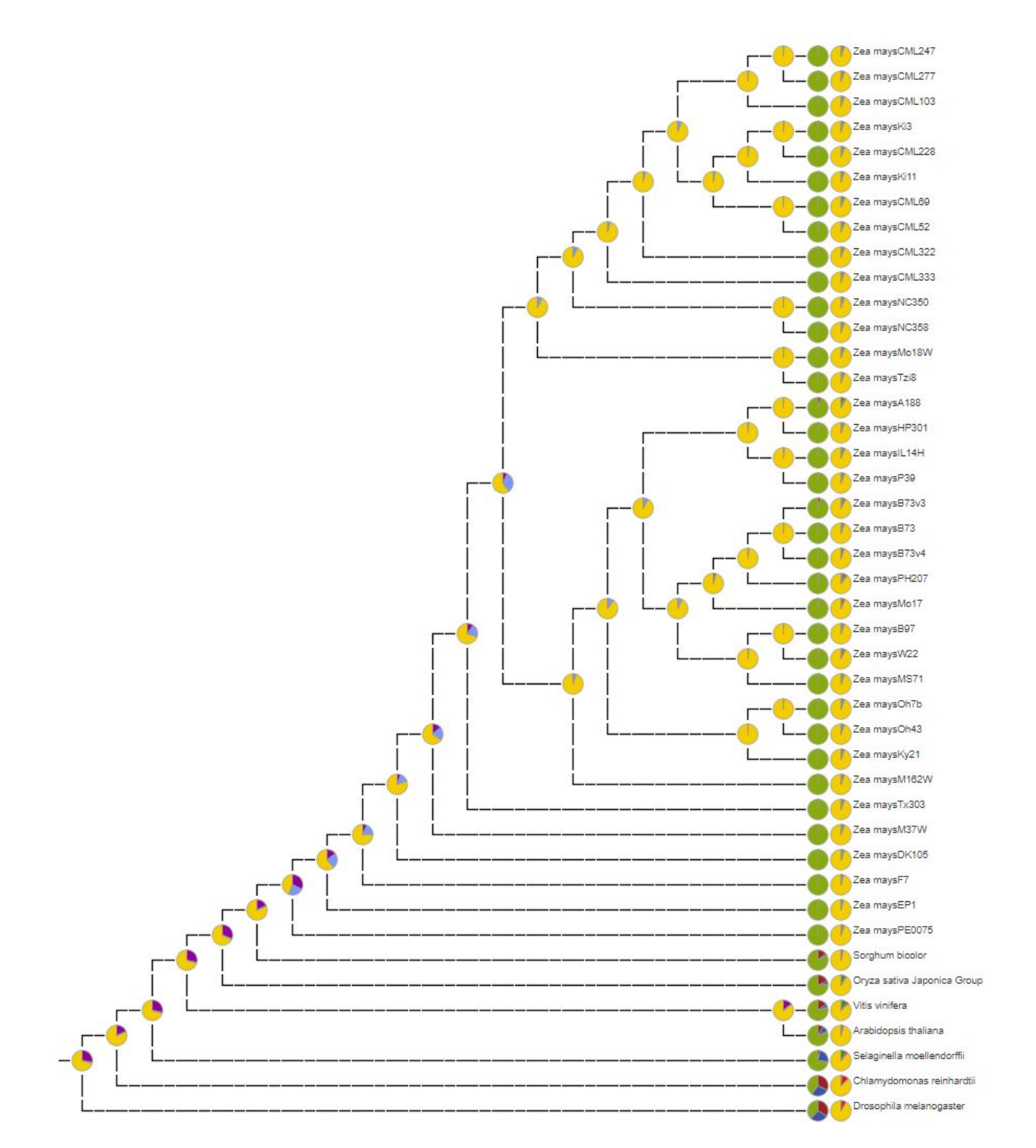
## Complex queries on the fly!

## Aligned tree view enables structural gene curation

## Gene Neighborhood Conservation View

## A multi-species phylogenomic and pathway-informed community resource

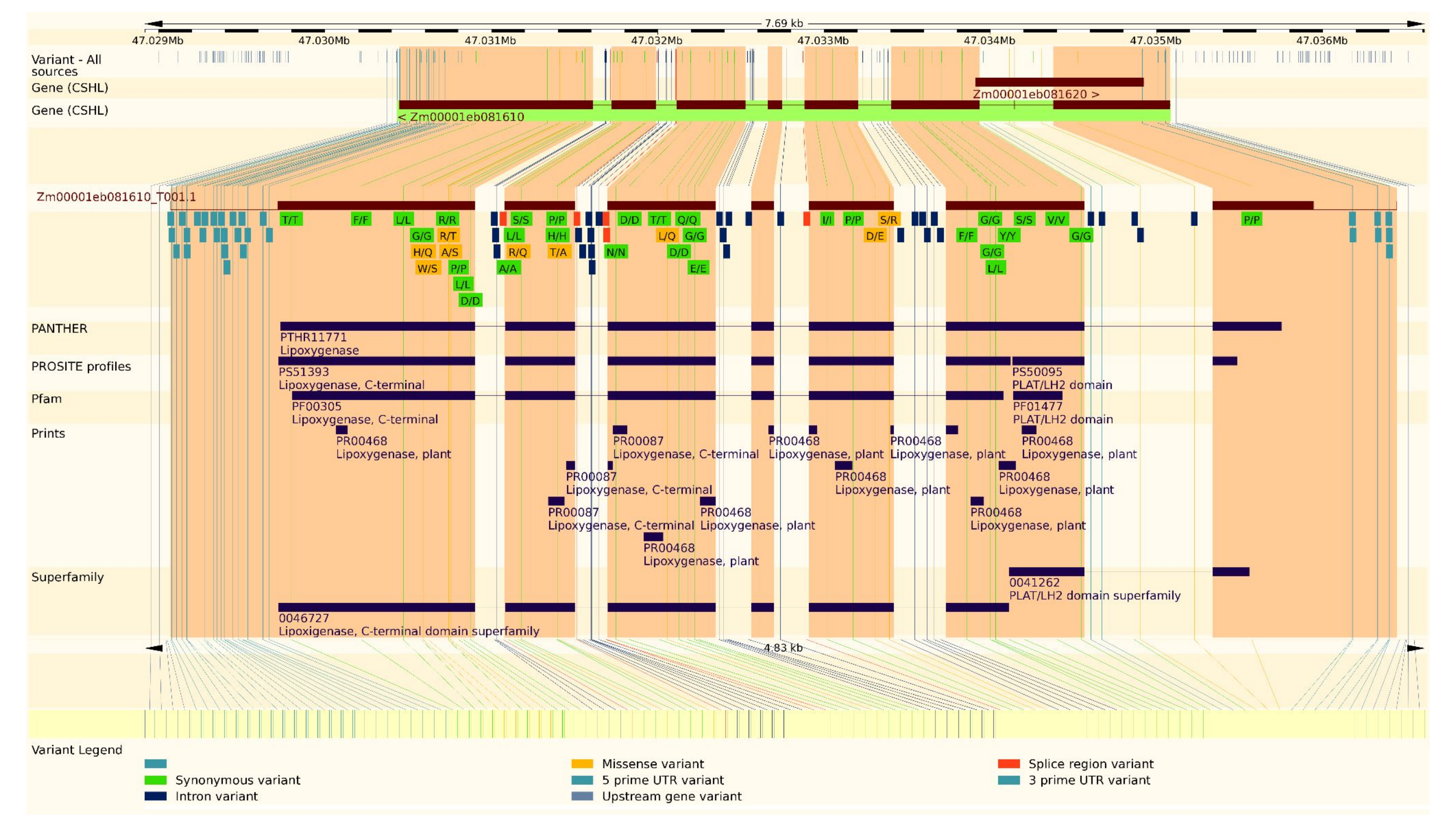
- Release 3 includes 43 plant genomes with 26 Maize NAM lines and 8 new genomes (4 European Flint, Mo17, W22, PH207 & A188)
- Integrated gene function from the literature enabling searches by TO/PO terms
- > 36K Gene Tree families with ortholog & paralog calls
- Manually curated B73 v4 gene models mapped to v5
- Gene curation user interface
- Genetic variation (Hapmap2 & Panzea) & structural variation for the NAM genomes vs B73 v5
- Visualization of variant effect prediction over protein domains
- Gene expression (EBI-Atlas) & orthology-based pathway projections (Plant Reactome) for B73 v5
- Blast service
- Programmatic access
- Quick start guide



## Structural variation for NAM genomes (B73 v5 browser)



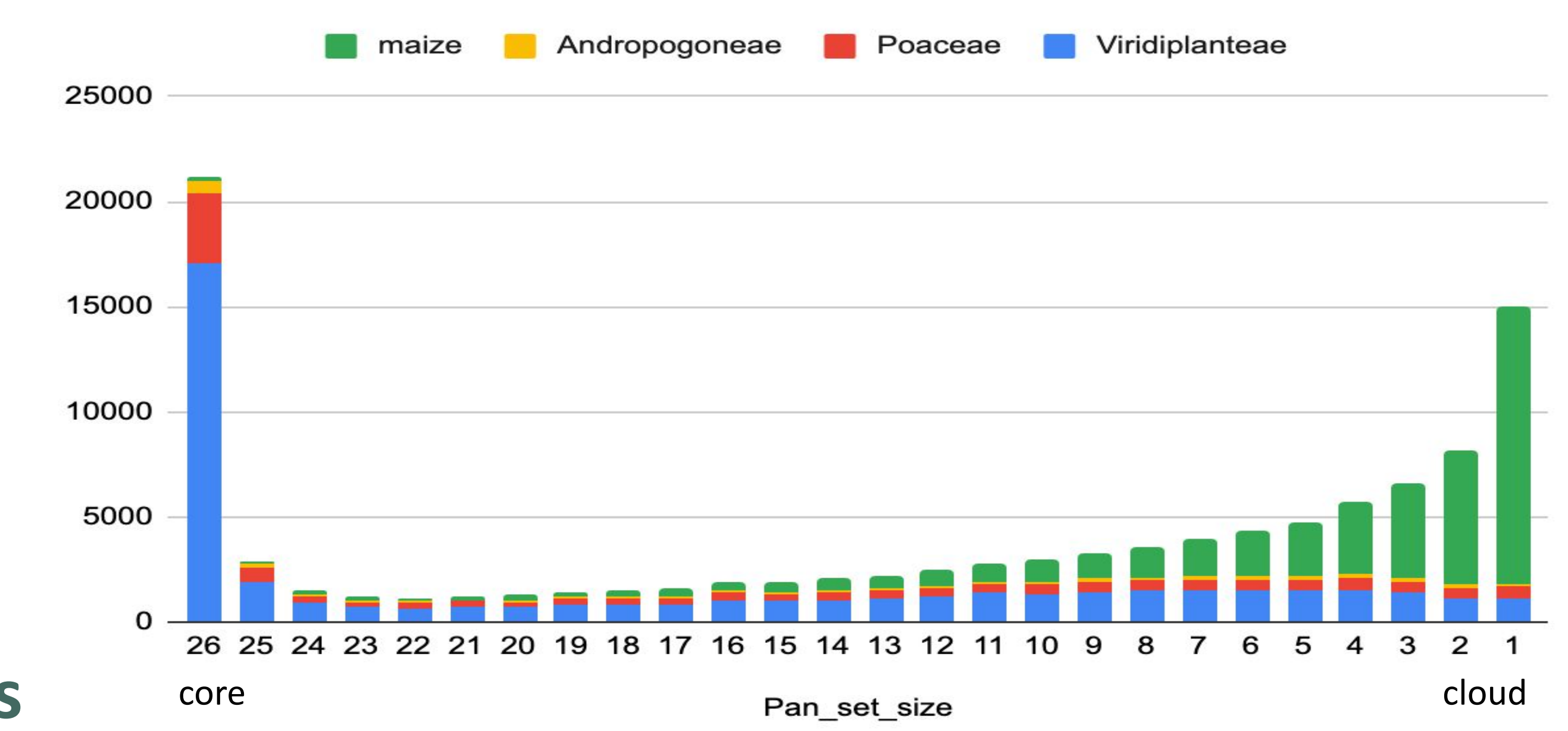
## Color-coded genetic variation overlaid on protein domains



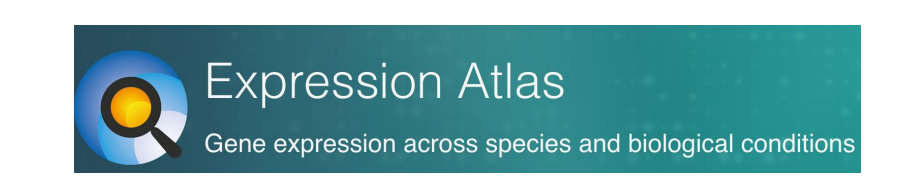
## Origin of Maize Genes

| Taxonomic Origin | % genes<br>n=39,324 | % families<br>n=9236 |
|------------------|---------------------|----------------------|
| Eukaryota        | 38.4                | 35.2                 |
| Viridiplantae    | 6.6                 | 7.2                  |
| Embryophyta      | 21.2                | 16.3                 |
| Tracheophyta     | 2.8                 | 2.9                  |
| Magnoliophyta    | 10.6                | 13.0                 |
| Commelinids      | 0.8                 | 1.3                  |
| Poaceae          | 8.1                 | 14.6                 |
| Panicoideae      | 0.9                 | 1.7                  |
| Andropogoneae    | 0.6                 | 1.5                  |
| Zea mays         | 5.3                 | 6.2                  |
| Orphan           | 4.7                 | 0.0                  |

## Pan gene distribution by age from Maize NAM genome annotations



## Expression Atlas



Baseline and differential expression analysis from a growing number of studies