# GENE HUNTERS: WHY ONTOLOGIES MATTER

**Mária Škrabišová, Ph.D.**

Palacký University in Olomouc CZ | Faculty of Science | Department of Biochemistry

Plant and Animal Genome Conference PAG32, San Diego, CA, USA
Systems Biology and Ontologies workshop, January 10, 2025

# PALACKÝ UNIVERSITY OLOMOUC

# FACULTY OF SCIENCES

# BIOLOGICAL DISCIPLINES CENTRE OLOMOUC - HOLICE

Department of Biochemistry

Legume genomics research group

# SOYBEAN APPLIED GENOMICS

**Development of strategies to accelerate soybean breeding and improvement**
**…by identifying causal genes: A gene-centeric approach**

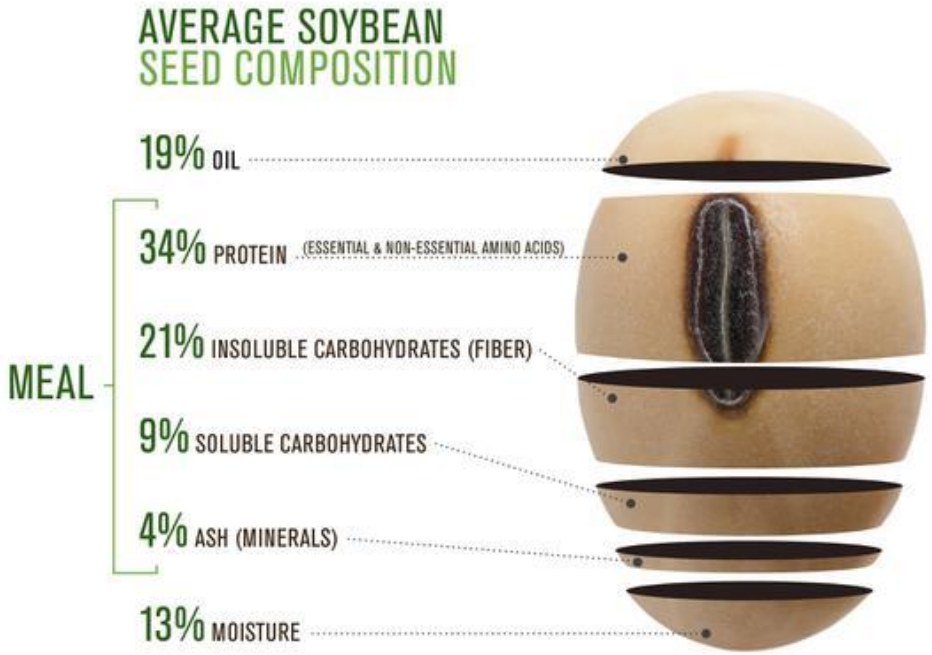
- Soybean diversity exploration
- Tools for applied genomics
- We are gene hunters

**Dr. Kristin Bilyeu**
USDA-ARS
Plant Research Unit

USDA

**Dr. Trupti Joshi**
Dep. of Biomedical Informatics
Dept. of Health Management
and Informatics

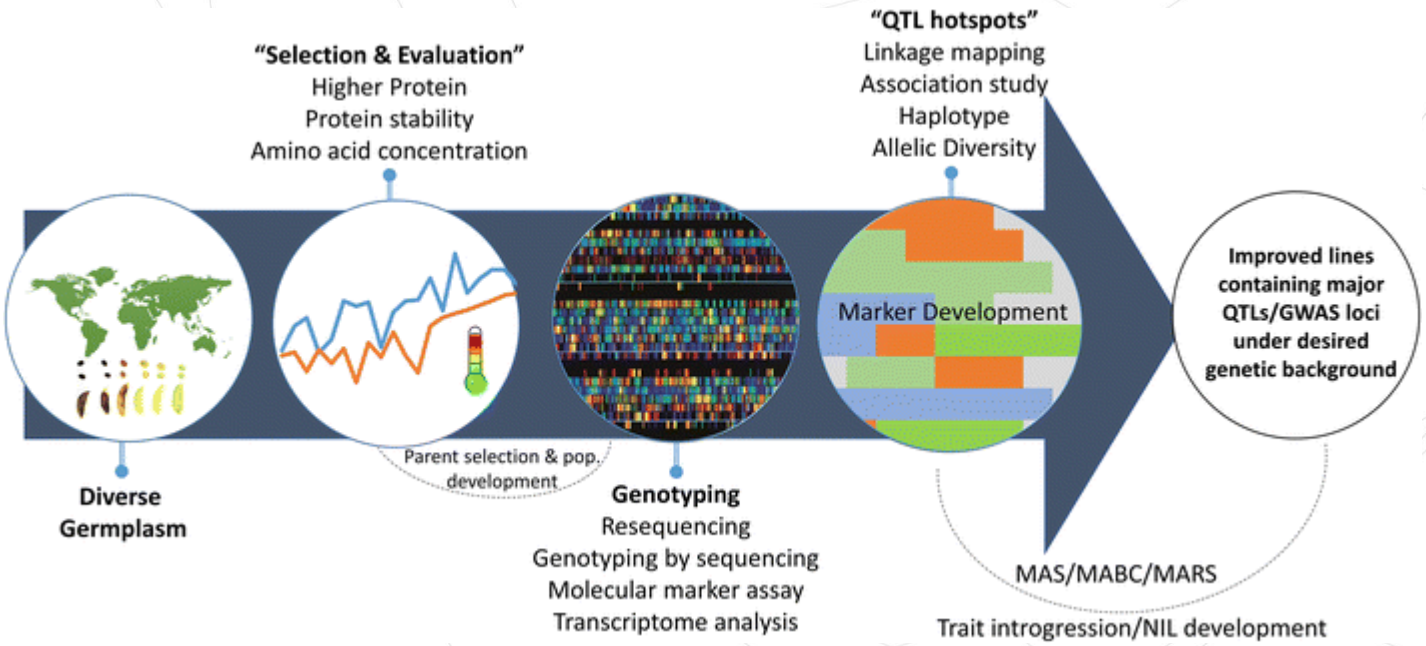
MU




Source: United Soybean Board

UNITED SOYBEAN BOARD™

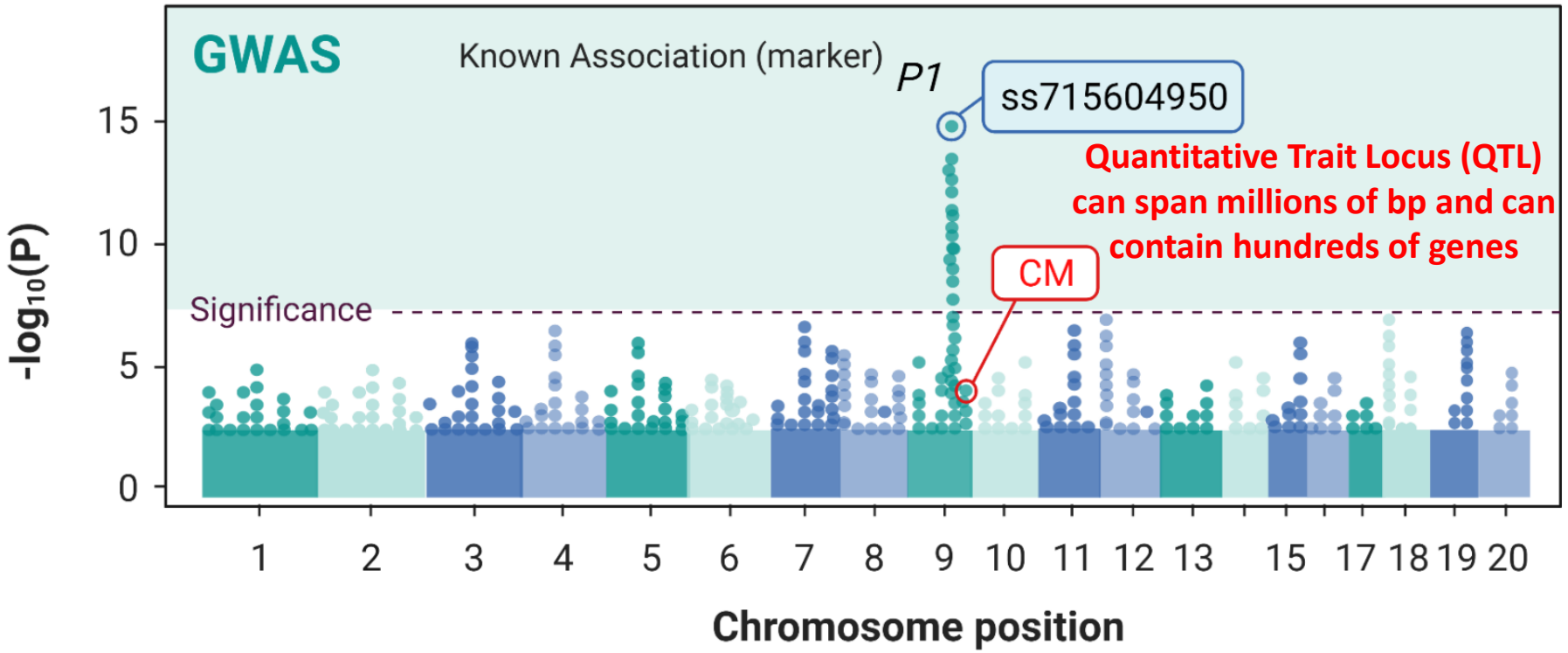
# CURRENT LIMITATIONS OF BREEDING: METHODS FOR QTL IDENTIFICATION




Gunvant et al., 2017

**Quantitative Trait Locus (QTL) can span millions of bp and can contain hundreds of genes**

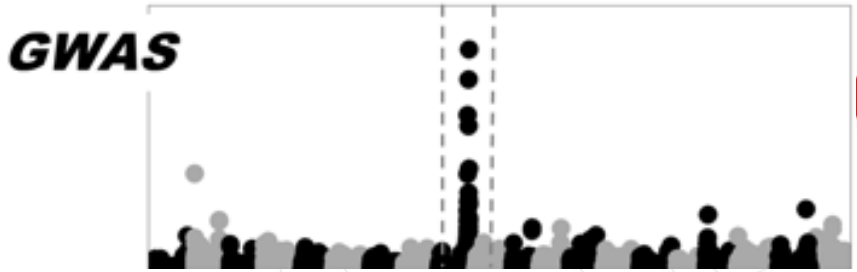
# GWAS FOR MORE PRECISE BREEDING: LIMITED.




## Varying GWAS power caused by:

- Genotype quality: density, Indels, population structure and size
- Phenotype issues: precision, data type nature (distribution of values, frequency, etc.)
- Fitting model applied
- Other factors

# HUNTING GENES FOR MORE PRECISE CROP BREEDING




Sorting the candidates

| Order | Variant position | |
|---|---|---|
| 1 | CM | Highest |
| 2 | Candidate 2 | |
| 3 | Candidate 3 | |
| 4 | Candidate 4 | |
| 5 | Candidate 5 | |
| 6 | Candidate 6 | |
| 7 | Candidate 7 | |
| 8 | Candidate 8 | |
| 9 | Candidate 9 | |
| 10 | Candidate 10 | Lowest |

**Guessing** genes…
…by proximity
**…by ontologies**

**Implementation of post-GWAS analyses to narrow down the candidates**

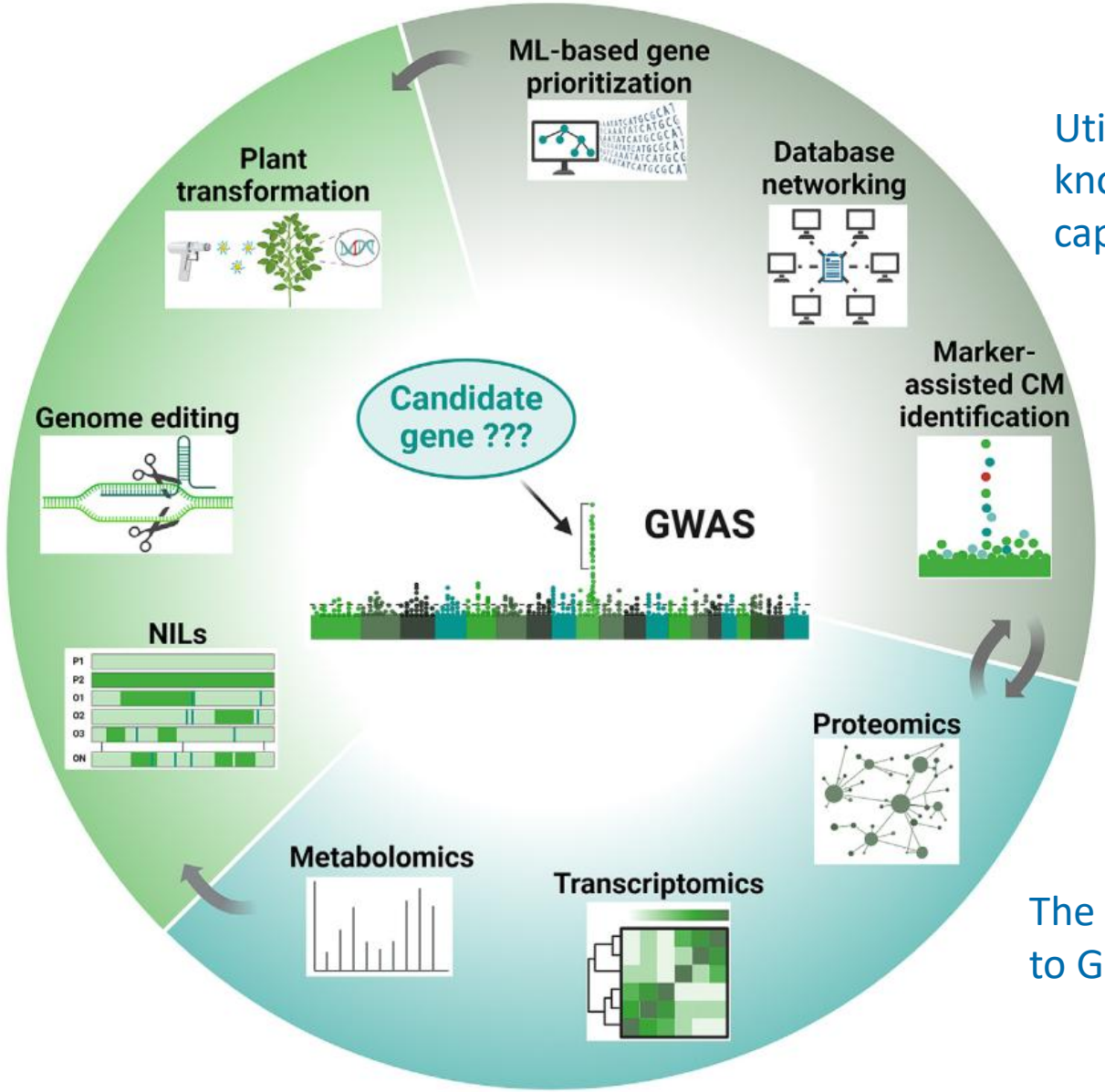
Legume Generation

# New perspectives of post-GWAS analyses:
## From markers to causal genes for more precise crop breeding

Legume Generation (Boosting innovation in breeding for the next generation of legume crops for Europe) has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No.101081329. It also receives support from the governments of the United Kingdom, Switzerland and New Zealand.

# New perspectives of post-GWAS analyses:
## From markers to causal genes for more precise crop breeding

Kanovska et al., 2024




Utilization of the previous knowledge with computational capabilities and algorithms

Engineering of genetic information

The addition of another layer to GWAS results, “omics”

# New perspectives of post-GWAS analyses:
## From markers to causal genes for more precise crop breeding

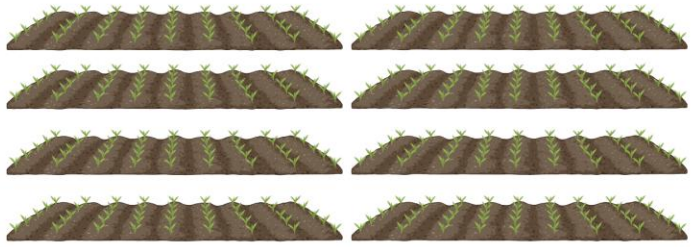
Legume Generation

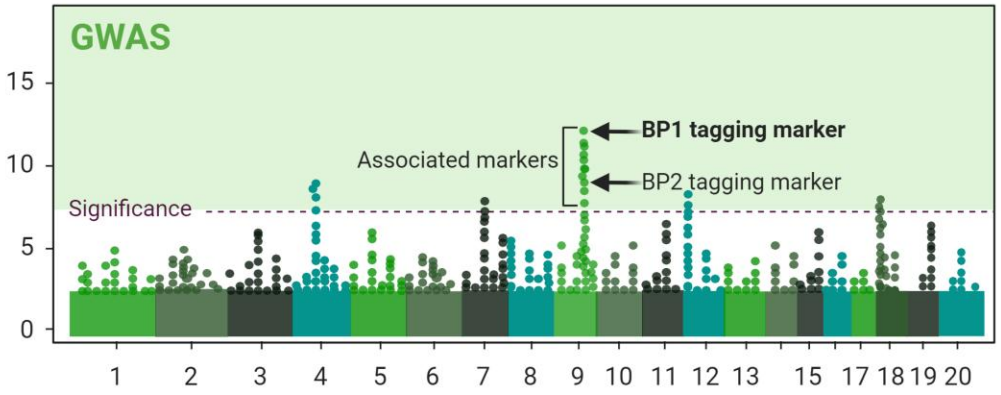
Breeding programs

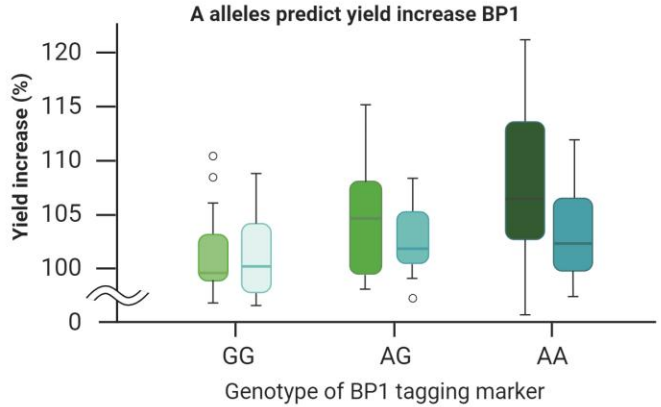
Marker identification

Marker efficiency

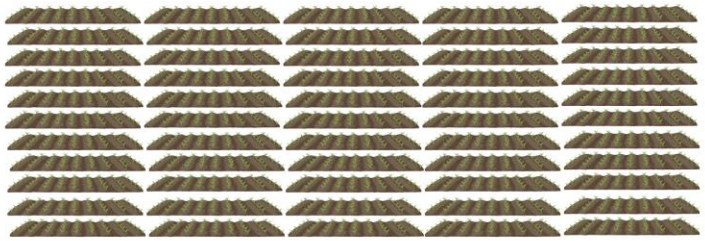
BP1

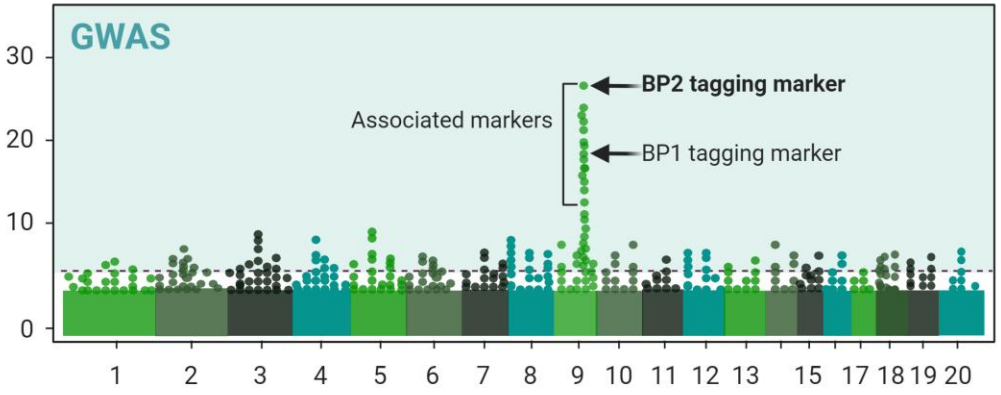


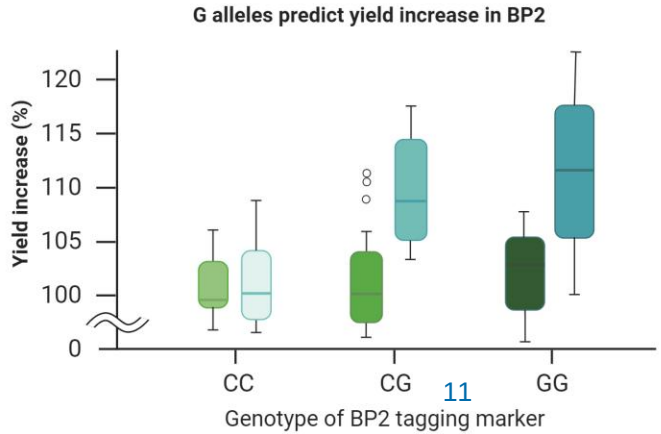



BP2



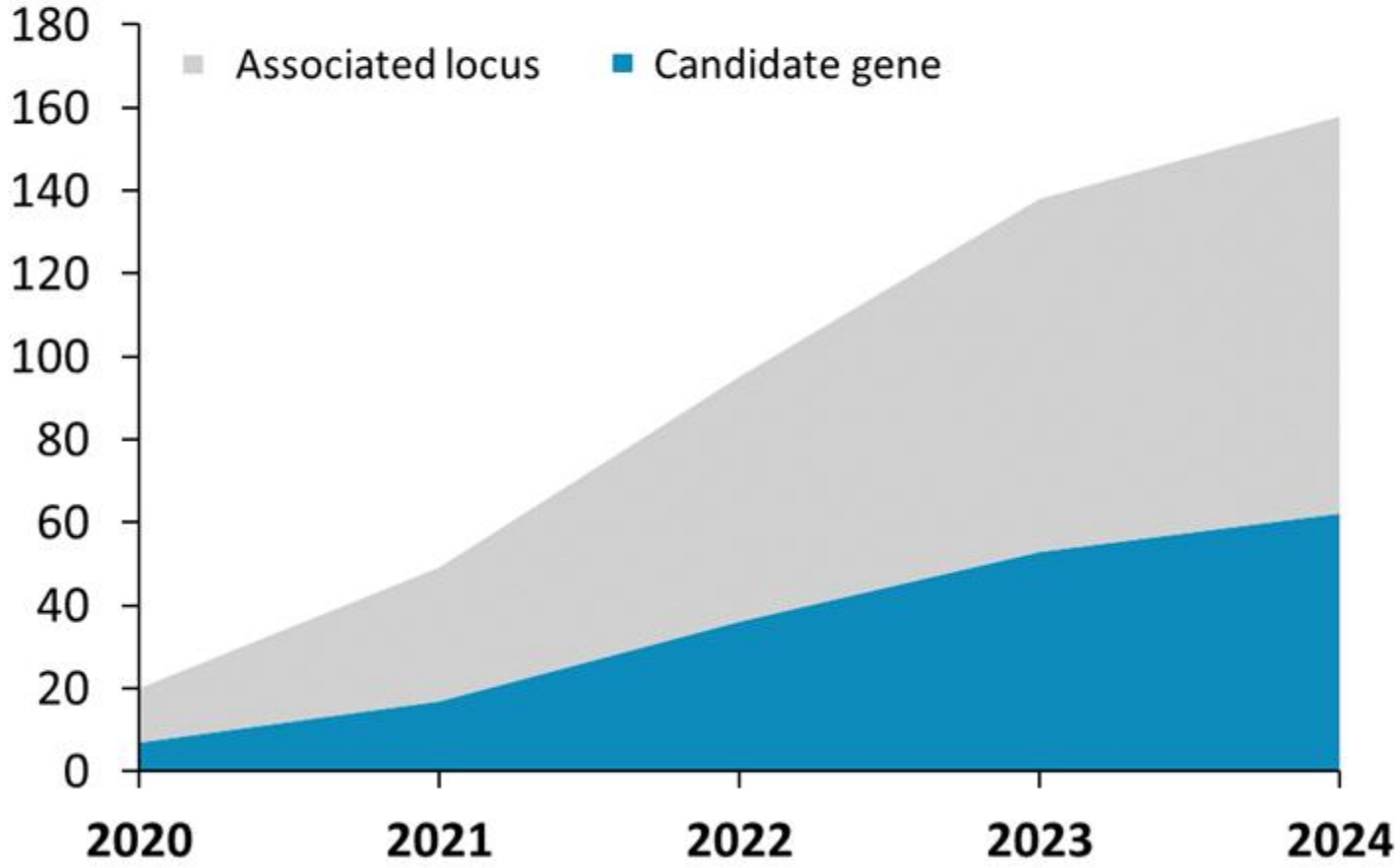


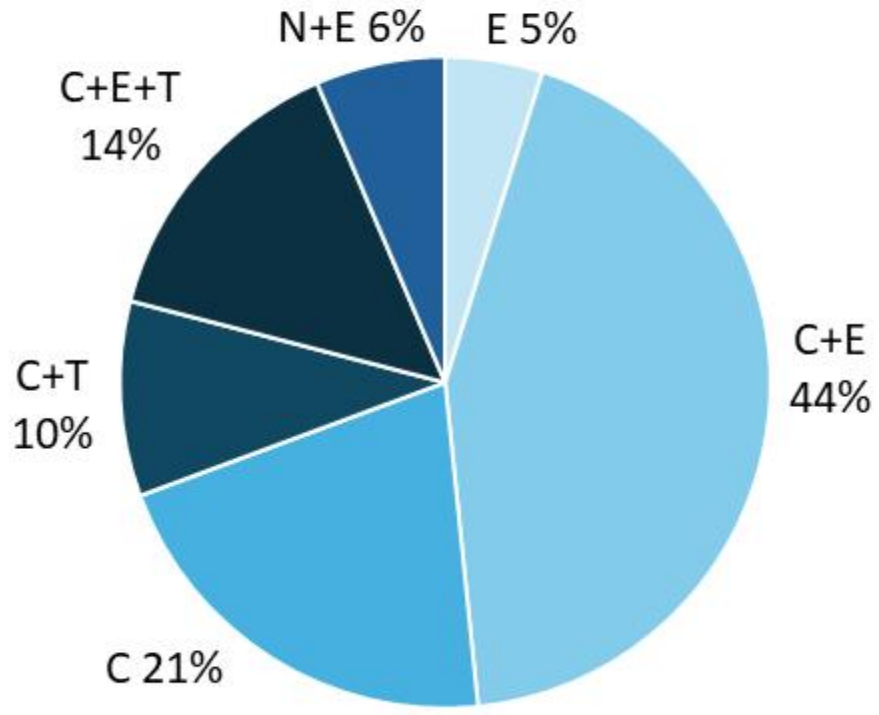
# New perspectives of post-GWAS analyses:
## From markers to causal genes for more precise crop breeding

Legume Generation

A




B



# A COMPLICATED EXAMPLE OF HUNTING A GENE …AND WHY ONTOLOGIES MATTER

## Obstacles of the gene-centric GWAS-based approach to identifying a causal gene:

1. **The most highly associated variant position is not the causative mutation**
   - skewed result of underpowered GWAS
   - complicated genetics: multiple alleles, etc.
2. **None of the highly associated variants falls into a genic region**
   - annotation issues
   - genotypic/phenotypic data issues
3. **There are more candidates but ontologies do not point to the right gene**
   - ontologies issue
   - more/parallel genes in a single locus involved
4. Other unforeseen complications

# HUNTING A GENE FOR SOYBEAN POD COLOR

## Obstacles of the gene-centric GWAS-based approach to identifying a causal gene:

1. **The most highly associated variant position is not CM**
   - skewed result of underpowered GWAS
   - complicated genetics: multiple alleles

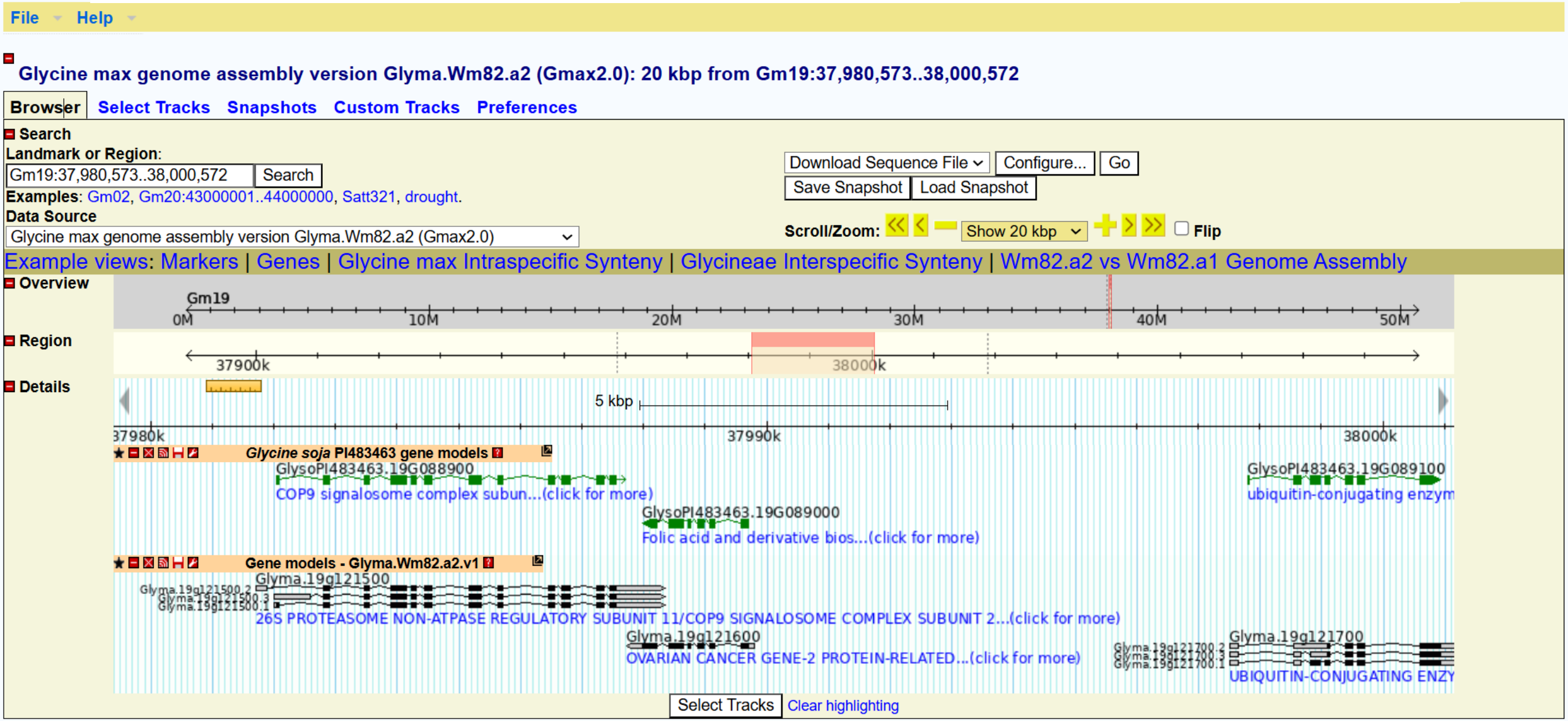
***L1* and *L2* genetic loci control pod wall pigmentation** (Nagai, 1921; Owen, 1927)

## *L1* on chromosome 19 controls the biosynthesis of “black” pigments

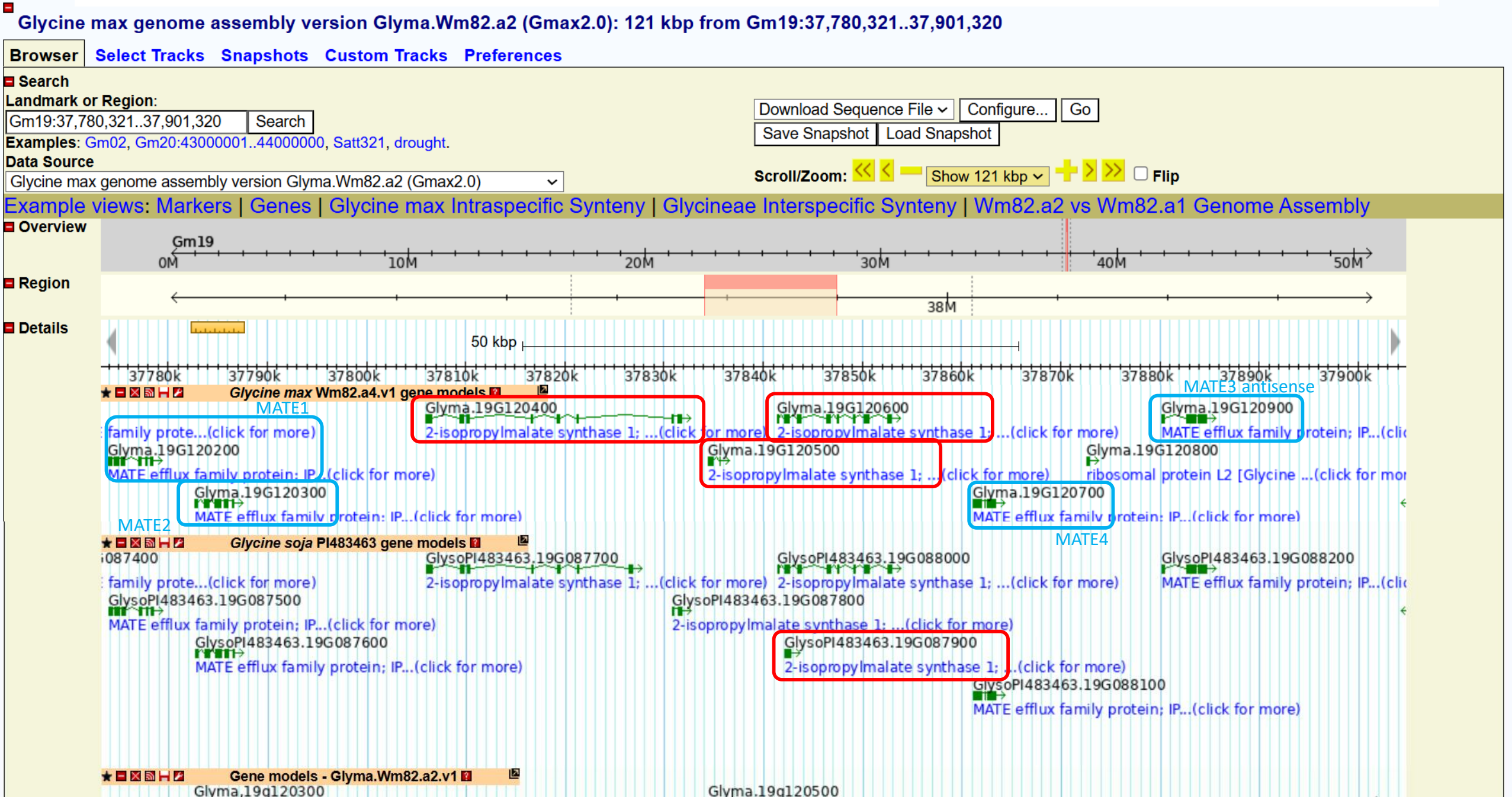
- Bandillo et al. 2015 – great GWAS, many candidate genes published afterward, no functional confirmation
- Candidate no. 1: *Glyma19g27460* (He et al. 2015 – gene expression-based evidence)
- No. 1 again: *Glyma.19g101700* (Torkamaneh et al. 2018 – domestication-related gene)
- Candidate no. 3: Search by ontologies in the wider region of *L1*
- Causal gene confirmed by transgenesis: *Glyma.19g120400* (Lyu et., 2023 – IPMS)

## *L2* on chromosome 3 controls the biosynthesis of “brown” pigments

# HUNTING A GENE FOR SOYBEAN POD COLOR – SEARCH BY ONTOLOGIES



# HUNTING A GENE FOR SOYBEAN POD COLOR – SEARCH BY ONTOLOGIES



# *L2* GENE IDENTIFIED: SoyHUB workshop on Friday 13th

Biova et al., 2024

Ivana Kaňovská
Ph.D. student

Poster 559

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 08 January 2024
DOI 10.3389/fgene.2023.1320652

Check for updates

OPEN ACCESS

EDITED BY
Li Ma,
University of Maryland, College Park,
United States

REVIEWED BY
Shoaib Ur Rehman,
Muhammad Nawaz Shareef University of
Agriculture, Pakistan
Sridhar Malkaram,
West Virginia State University,
United States

*CORRESPONDENCE
Kristin Bilyeu,
kristin.bilyeu@usda.gov
Mária Škrabišová,
maria.skrabisova@upol.cz

†These authors have contributed equally
to this work

# Natural and artificial selection of multiple alleles revealed through genomic analyses

Jana Biová[1†], Ivana Kaňovská[1†], Yen On Chan[2,3], Manish Sridhar Immadi[4], Trupti Joshi[2,3,4,5], Kristin Bilyeu[6*] and Mária Škrabišová[1*]

[1]Department of Biochemistry, Faculty of Science, Palacký University in Olomouc, Olomouc, Czechia, [2]MU Institute for Data Science and Informatics, University of Missouri-Columbia, Columbia, MO, United States, [3]Christopher S. Bond Life Sciences Center, University of Missouri-Columbia, Columbia, MO, United States, [4]Department of Electrical Engineering and Computer Science, University of Missouri-Columbia, Columbia, MO, United States, [5]Department of Biomedical Informatics, Biostatistics and Medical Epidemiology, University of Missouri-Columbia, Columbia, MO, United States, [6]United States Department of Agriculture-Agricultural Research Service, Plant Genetics Research Unit, Columbia, MO, United States

# GENE HUNTERS: WHY ONTOLOGIES MATTER WRAP-UP

- Ontologies are gene annotation related
- Gene annotations might vary between different reference genomes and between assemblies
- Ontologies are just terms limited by their key words (vocabularies)
- Combination of more post-GWAS approaches advocate for more accurate predictions
- Use CMs, not inaccurate markers for screening in breeding programs

# JOINT EFFORTS FOR SOYBEAN APPLIED GENOMICS

## Legume Genomics

Dr. Mária Škrabišová

Dr. Jana Biová

Ivana Kaňovská, Ph.D. student

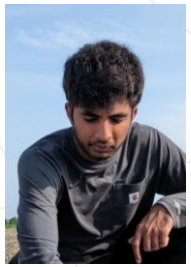
Jana Slivková, M.Sc.

## Funding

## Applied Genomics

Dr. Kristin Bilyeu

Dr. Nicholas Dietz



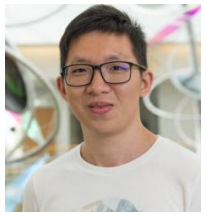
Anser Mahmood, Ph.D. student

Dr. Nathan Grant

## Bioinformatics

Dr. Trupti Joshi

Dr. Shuai Zeng



Dr. Yen On Chan

Manish Sridhar Immadi

# Acknowledgement

Legume Generation (Boosting innovation in breeding for the next generation of legume crops for Europe) has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No.101081329. It also receives support from the governments of the United Kingdom, Switzerland and New Zealand.

# THANK YOU FOR YOUR ATTENTION!