Legume Generation

MADis: Genomic Analysis Tool for the Revelation of Multiple Alleles Within a Single Gene

Ivana Kaňovská[a], Jana Biová[a], Kristin Bilyeu[b], Mária Škrabišová[a*]
[a]*Department of Biochemistry, Faculty of Science, Palacký University in Olomouc, 783 71 Olomouc, Czech Republic*
[b]*United States Department of Agriculture-Agricultural Research Service, Plant Genetics Research Unit, Columbia, MO, United States*
*maria.skrabisova@upol.cz

Background

Genome-to-phenome research in agriculture aims to **improve and accelerate crop breeding. Genome-wide association study** (GWAS) has been used to identify genomic loci underlying important traits. Numerous post-GWAS analyses were developed to narrow down the associated genomic regions. However, in many cases, they are still **unable to identify candidate genes or causative mutations** (CMs).

One of the factors that complicate the GWAS-driven discoveries is the existence of **multiple independent CMs in a single gene** that has emerged either through historical natural selection or breeding-related artificial selection. These alleles can impact the resulting phenotype differently.

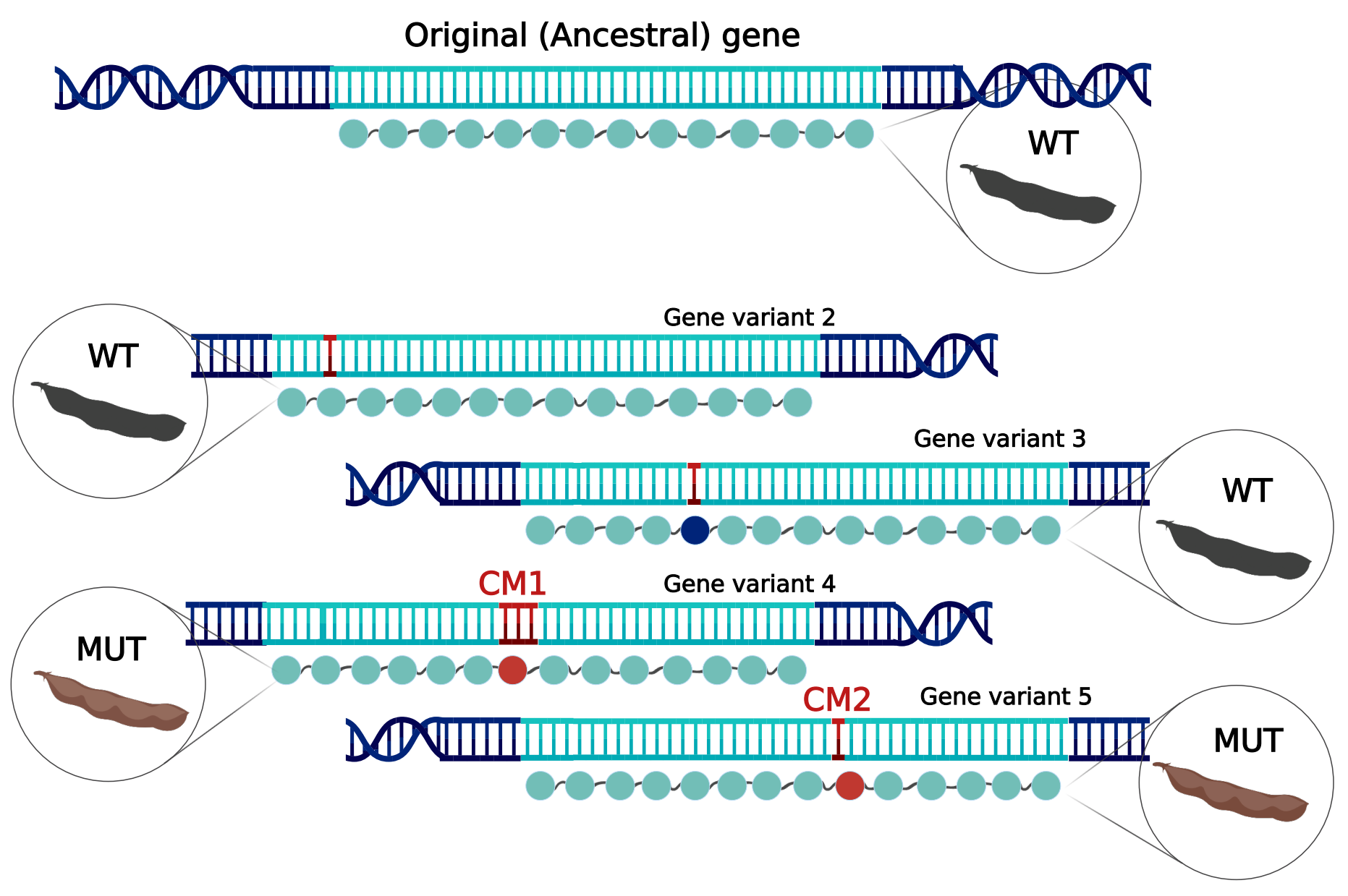
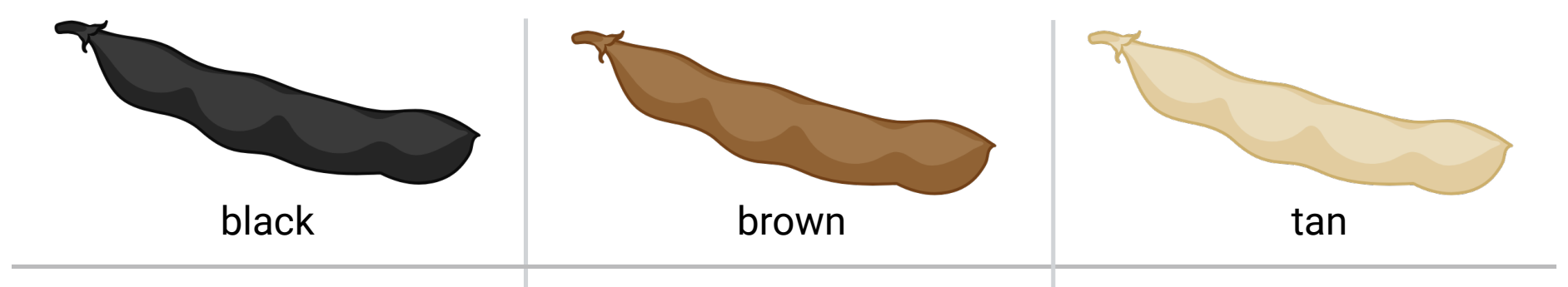


Figure 1: Scheme representing the isolated emergence of multiple alleles as a part of evolution during natural selection or domestication.

In our work, we focused on **soybean as a globally important crop** for its high protein and oil content. In soybean, multiple CMs have been discovered in many domestication-related genes - stem termination, seed coat color, pubescence color, maturity genes, and others.

Recently, a causal gene was identified for the pod color *L1* locus [1] that causes the loss of black pigments in pod walls. This may result in a brown or tan pod color, depending on **allele *L2*** (Figure 2). Several candidate genes have been proposed for *L2,* but none of them has been confirmed yet.



black	brown	tan
L1L1 L2L2 *L1L1 l2l2*	*l1l1 L2L2*	*l1l1 l2l2*

Figure 2: Soybean pod color phenotypes and their respective genetic background.

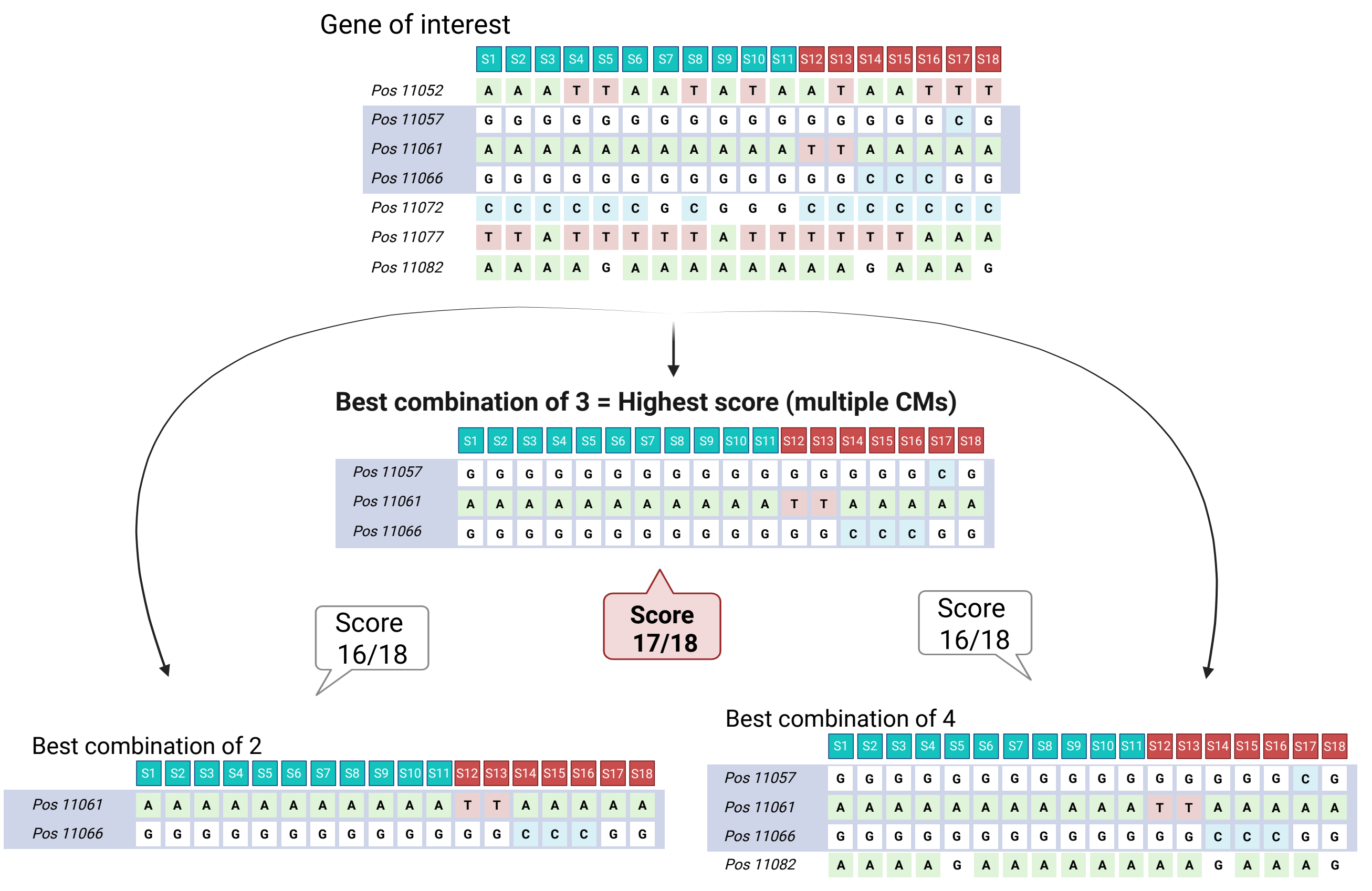
Objectives

− Develop a tool for the identification of multiple alleles
− Identify candidate gene for brown pod color in soybean

Results

In our previous work, we developed a new GWAS evaluation criterion named Accuracy that serves as a measure of direct correspondence between variant positions and phenotypes [2]. To capture all possible additional alleles that contribute to a studied phenotype change, **we developed the MADis (Multiple Alleles discovery) tool** [3].

The MADis tool tests combinations of variant positions in a gene with the aim of reaching the most probable combination of candidate CMs determined by the highest score that would explain most of the phenotypes in the dataset (Figure 3).



Gene of interest

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18
Pos 11052	A	A	A	T	T	A	A	T	A	T	A	A	T	A	A	T	T	T
Pos 11057	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	C	G
Pos 11061	A	A	A	A	A	A	A	A	A	A	A	T	T	A	A	A	A	A
Pos 11066	G	G	G	G	G	G	G	G	G	G	G	G	G	C	C	C	G	G
Pos 11072	C	C	C	C	C	C	G	C	G	G	G	C	C	C	C	C	C	C
Pos 11077	T	T	A	T	T	T	T	T	A	T	T	T	T	T	T	A	A	A
Pos 11082	A	A	A	A	G	A	A	A	A	A	A	A	A	G	A	A	A	G

Best combination of 3 = Highest score (multiple CMs)

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18
Pos 11057	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	C	G
Pos 11061	A	A	A	A	A	A	A	A	A	A	A	T	T	A	A	A	A	A
Pos 11066	G	G	G	G	G	G	G	G	G	G	G	G	G	C	C	C	G	G

Score 16/18

Score 17/18

Score 16/18

Best combination of 2

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18
Pos 11061	A	A	A	A	A	A	A	A	A	A	A	T	T	A	A	A	A	A
Pos 11066	G	G	G	G	G	G	G	G	G	G	G	G	G	C	C	C	G	G

Best combination of 4

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18
Pos 11057	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	C	G
Pos 11061	A	A	A	A	A	A	A	A	A	A	A	T	T	A	A	A	A	A
Pos 11066	G	G	G	G	G	G	G	G	G	G	G	G	G	C	C	C	G	G
Pos 11082	A	A	A	A	G	A	A	A	A	A	A	A	A	G	A	A	A	G

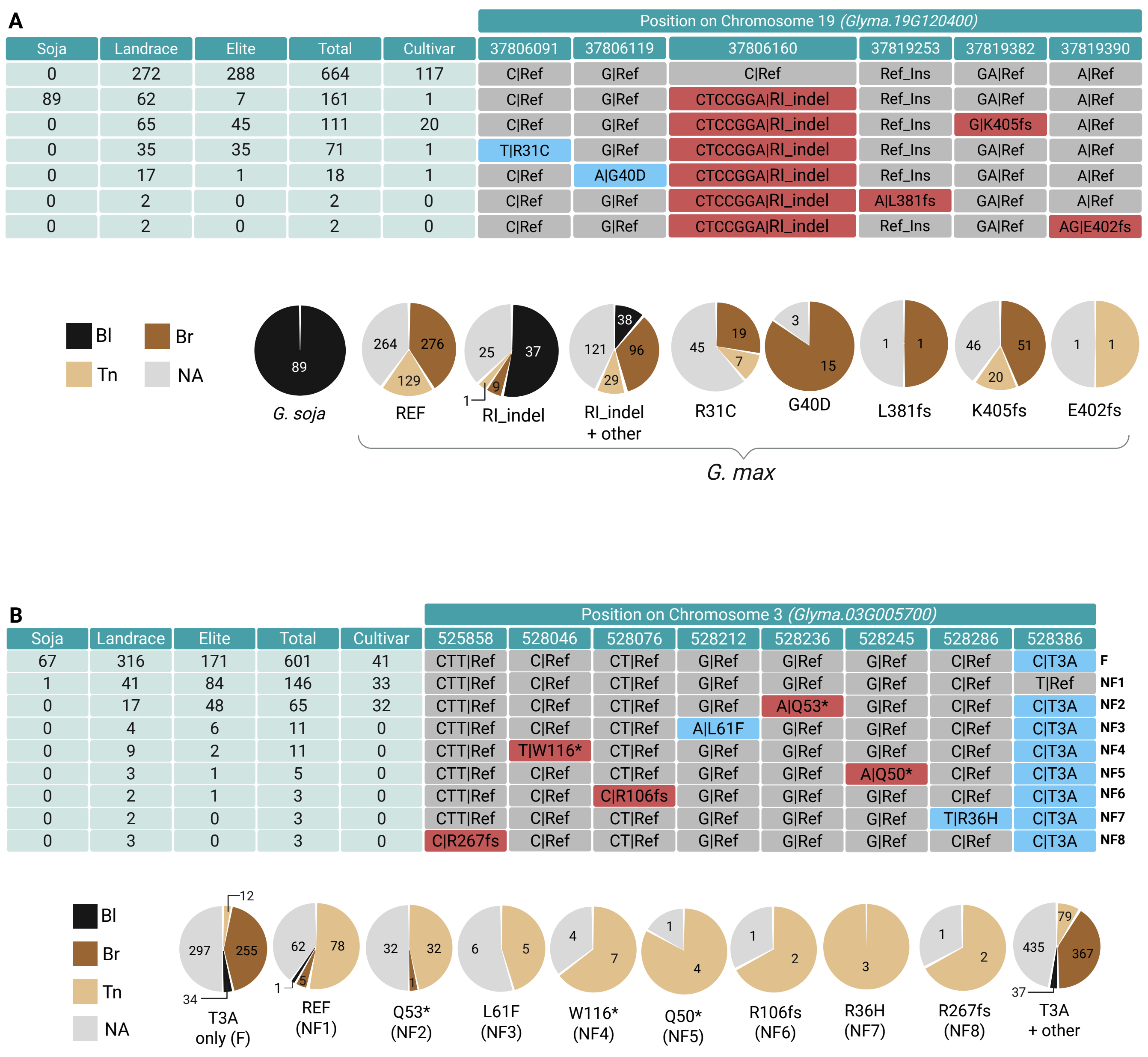
Figure 3: Simplified scheme of the MADis algorithm.

We demonstrated the validation of our genomic analysis on an example of a recently cloned gene *L1* with multiple CMs (Figure 4A).

We identified the *L2* locus responsible for brown pod color by accuracy analysis with AccuCalc.

Based on the highest average accuracy, we identified ***Glyma.03G005700*** as the candidate gene for *L2*, annotated as isopropylmalate synthase (IPMS), a homolog to the cloned *L1 Glyma.19G120400.*

Using MADis, we discovered **eight independent putative CMs** (Figure 4B) that corresponded to the complete loss of pigmentation in soybean pods, which results in the tan phenotype.



A

					Position on Chromosome 19 *(Glyma.19G120400)*					
Soja	Landrace	Elite	Total	Cultivar	37806091	37806119	37806160	37819253	37819382	37819390
0	272	288	664	117	C|Ref	G|Ref	C|Ref	Ref_Ins	GA|Ref	A|Ref
89	62	7	161	1	C|Ref	G|Ref	CTCCGGA|RI_indel	Ref_Ins	GA|Ref	A|Ref
0	65	45	111	20	C|Ref	G|Ref	CTCCGGA|RI_indel	Ref_Ins	G|K405fs	A|Ref
0	35	35	71	1	T|R31C	G|Ref	CTCCGGA|RI_indel	Ref_Ins	GA|Ref	A|Ref
0	17	1	18	1	C|Ref	A|G40D	CTCCGGA|RI_indel	Ref_Ins	GA|Ref	A|Ref
0	2	0	2	0	C|Ref	G|Ref	CTCCGGA|RI_indel	A|L381fs	GA|Ref	A|Ref
0	2	0	2	0	C|Ref	G|Ref	CTCCGGA|RI_indel	Ref_Ins	GA|Ref	AG|E402fs

B

					Position on Chromosome 3 *(Glyma.03G005700)*								
Soja	Landrace	Elite	Total	Cultivar	525858	528046	528076	528212	528236	528245	528286	528386	
67	316	171	601	41	CTT|Ref	C|Ref	CT|Ref	G|Ref	G|Ref	G|Ref	C|Ref	C|T3A	**F**
1	41	84	146	33	CTT|Ref	C|Ref	CT|Ref	G|Ref	G|Ref	G|Ref	C|Ref	T|Ref	**NF1**
0	17	48	65	32	CTT|Ref	C|Ref	CT|Ref	G|Ref	A|Q53*	G|Ref	C|Ref	C|T3A	**NF2**
0	4	6	11	0	CTT|Ref	C|Ref	CT|Ref	A|L61F	G|Ref	G|Ref	C|Ref	C|T3A	**NF3**
0	9	2	11	0	CTT|Ref	T|W116*	CT|Ref	G|Ref	G|Ref	G|Ref	C|Ref	C|T3A	**NF4**
0	3	1	5	0	CTT|Ref	C|Ref	CT|Ref	G|Ref	G|Ref	A|Q50*	C|Ref	C|T3A	**NF5**
0	2	1	3	0	CTT|Ref	C|Ref	C|R106fs	G|Ref	G|Ref	G|Ref	C|Ref	C|T3A	**NF6**
0	2	0	3	0	CTT|Ref	C|Ref	CT|Ref	G|Ref	G|Ref	G|Ref	T|R36H	C|T3A	**NF7**
0	3	0	3	0	C|R267fs	C|Ref	CT|Ref	G|Ref	G|Ref	G|Ref	C|Ref	C|T3A	**NF8**

Figure 4: Allele distribution in *Glyma.19G120400* (A) and *Glyma.03g005700* (B) in the Soy1066 dataset and its correlation with different pod-colored phenotypes.

Results

To assess the natural and artificial selection of the three main pod-colored phenotypes, we examined the **distribution of *L1/l1/ l2* phenotypes** in the context of the world soybean population (Figures 5A, B) based on our diversity panel Soy1066 [4].

The distribution of functional *L1* black pod-colored genotypes was centered more on the regions of soybean origin (China, Korea, and Japan), whereas ***l1* and *l2* genotypes were selected in geographical regions where soybeans were introduced, domesticated, and adopted**.

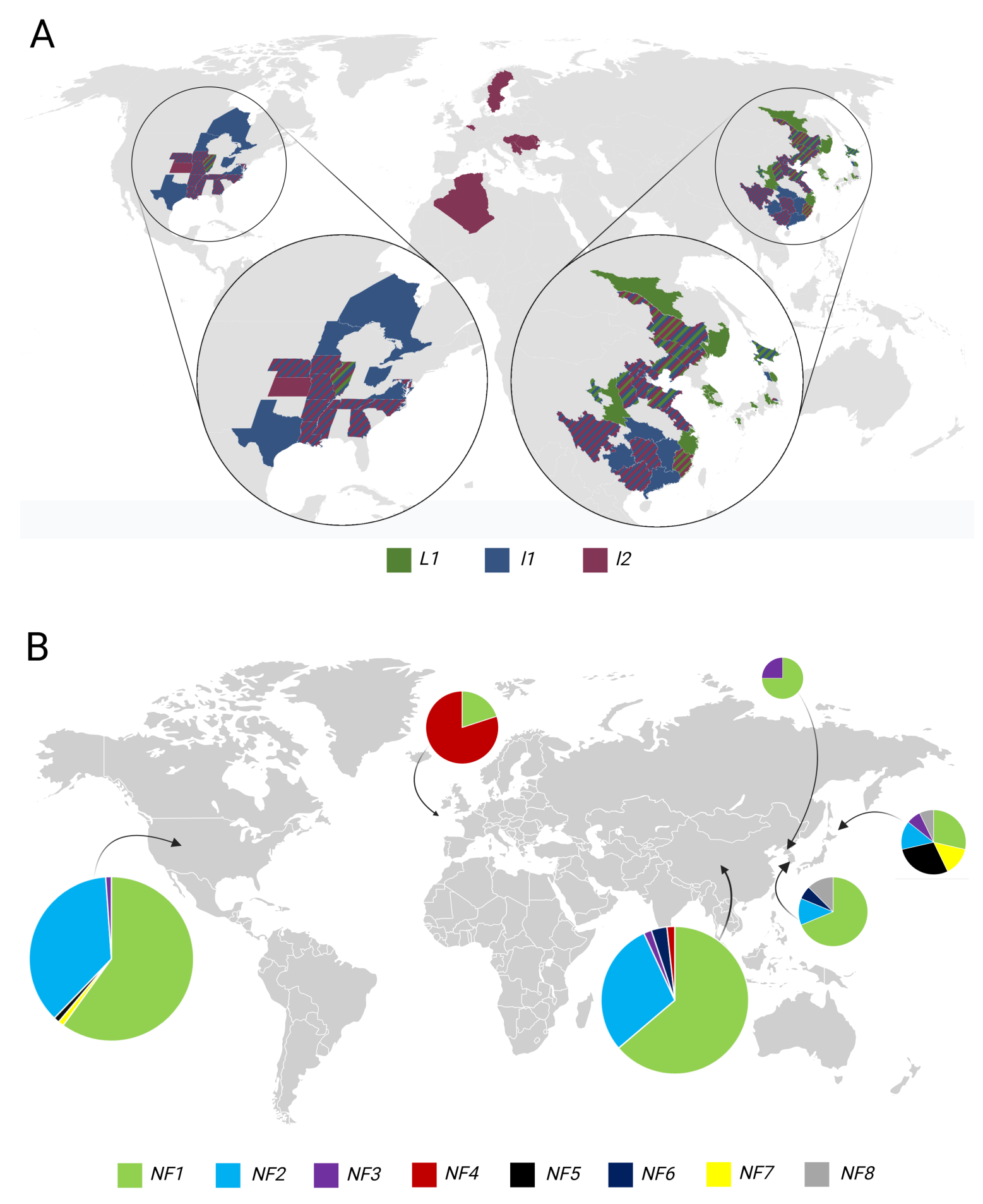


Figure 5: Geographical distribution of alleles responsible for pod color in soybean. (A) The global map displays the distribution of functional *L1*, non-functional *l1*, and non-functional *l2*. (B) Pie charts show the distribution of non-functional *l2* in six regions.

Conclusion

In this work, we solved the GWAS-limiting factor caused by the existence of parallel CMs in candidate genes that arose during natural and artificial selection. **We developed, tested, and validated the MADis tool** for the successful identification of multiple CMs for soybean. **The MADis analysis platform is publicly available for other species** to aid in the discovery of genes under selection for accelerated and, thus, improved breeding.

Access MADis Article

Soybean MADis

MADis GitHub

References

[1] Lyu, X., *et al.* (2023). Molecular Plant 16(7): 1178-1191.
[2] Skrabisova, M., *et al.* (2022). J Adv Res 42: 117-133.
[3] Biova, J., *et al.* (2024). Front Genet 14: 1320652.
[4] Chan, Y. O., *et al.* (2023). BMC Genomics 2023 24:1 24(1): 1-14.

Funding:
UP Internal Grant Agency (IGA_PrF_2024_023)

ACKNOWLEDGEMENTS

Legume Generation (Boosting innovation in breeding for the next generation of legume crops for Europe) has received funding from the European Union through Horizon Europe under grant agreement No 101081329 and co-funding from UK Research and Innovation (UKRI) from the UK government's Horizon Europe funding guarantee. It also receives support from the governments of Switzerland and New Zealand.

www.legumegeneration.eu @LegumeHubEU linkedIn.com/company/legume-hub-eu