

Applying genotypic principal component scores as latent phenotypes in genome-wide and epistatic analyses of soybean agronomic traits.

Simon Lackey^{1,2†}, Siwar Haidar^{1,2†}, Martin Charette¹, Louise O'Donoghue³, Istvan Rajcan⁴, Francois Belzile⁵, Ashkan Golshani², Elroy Cober^{1*}, Bahram Samanfar^{1,2*}

Author affiliations:

1Agriculture and Agri-Food Canada, Ottawa Research and Development Centre, Ottawa, ON, Canada

2Department of Biology, Ottawa Institute of Systems Biology, Carleton University, Ottawa, ON, Canada

3CÉROM, Centre de recherche Sur Les Grains Inc., Saint-Mathieu de Beloeil, QC, Canada

4Department of Plant Agriculture, University of Guelph, Guelph, ON, Canada

5Department of Phytology, Institut de Biologie Intégrative et des Systèmes (IBIS), Université de Laval, Québec City, QC, Canada

*Contact information for corresponding authors: Bahram Samanfar, bahram.samanfar@agr.gc.ca Elroy Cober, elroy.cober@agr.gc.ca

† co-first authors

Supplementary Table S3 Total number of proximal SNPs and/or QTLs in QTL regions identified by GWAS of gPCs of agronomic traits of soybean. Name of QTL is derived from the phenotypic trait (Mat: maturity; Oil: oil; Yld: yield; Plht: plant height; Prot: protein; Sdwt: seed weight; PO: both protein and oil), chromosome number, and principal component of the trait used in the association analysis. Chromosome:SNP is the G. max chromosome where the SNP is located, and SNP position is the position in bp. Sonah et al. (2015) shows whether the QTL was previously identified using conventional phenotypes. Citation is the original publication where the proximal SNPs or QTLs were originally identified.

QTL Name	Chromosome:SNP	Sonah et al. (2015)	Total Proximal SNPs and/or QTLs	Citation
Prot-6-PC1-1	Gm6: 20238533	Yes	5	[1-5]
Prot-14-PC1-2	Gm14: 65161	Yes	2	[6,7]
Prot-14-PC1-3	Gm14: 191942	Yes	1	[6]
Prot-19-PC1-4	Gm19: 47635025	No	6	[6,8-12]
Prot-20-PC1-5	Gm20: 31756869	Adjacent	17	[1,13-29]
Plht-6-PC1-1	Gm6:20238533	Yes	3	[5,29,30]
Plht-6-PC4-1	Gm6:43759173	No	2	[31,32]
Plht-19-PC1-2	Gm19:47635025	No	6	[33-38]
Sdwt-20-PC1-1	Gm20:43684690	Adjacent	5	[26,29,39-41]
Yld-5-PC1-1	Gm5:3114497	No	0	
Mat-5-PC1-1	Gm5:3114497	No	2	[42,43]
Mat-19-PC1-2	Gm19:39588804	No	2	[44,45]

Mat-19-PC1-3	Gm19:47635025	Yes	10	[29,38,46-50]
Mat-3-PC2-1	Gm3:2742274	No	2	[46,51]
Mat-11-PC2-2	Gm11:34564401	No	1	[52]
Mat-14-PC2-3	Gm14:6909226	No	0	
Mat-16-PC2-4	Gm16:24831342	No	3	[38,47,49]
Oil-4-PC1-1	Gm4:3466680	No	0	
Oil-4-PC1-2	Gm4:46247981	No	2	[53,54]
Oil-5-PC1-3	Gm5:31424775	Yes	2	[5,29]
Oil-14-PC1-4	Gm14:191942	Yes	2	[5,29]
Oil-15-PC1-5	Gm15:5786875	No	5	[53,55]
PO-2-Epi-4	Gm2:2362679	No	0	
PO-10-Epi-29	Gm10:2176639	No	2	[20]

- 1 Zhao, Q. *et al.* Seed Protein Genetics Linked with Nitrogen and Phosphorus Translocation Efficiency in Soybean. *Agronomy* **13** (2023). <https://doi.org/10.3390/agronomy13020598>
- 2 Zhao, Q. *et al.* Coordinate Inheritance of Seed Isoflavone and Protein in Soybean. *Agriculture (Switzerland)* **12** (2022). <https://doi.org/10.3390/agriculture12081178>
- 3 Patil, G. *et al.* Dissecting genomic hotspots underlying seed protein, oil, and sucrose content in an interspecific mapping population of soybean using high-density linkage mapping. *Plant Biotechnol. J.* **16**, 1939-1953 (2018). <https://doi.org/10.1111/pbi.12929>
- 4 Qin, J. *et al.* Identification of Candidate Genes and Genomic Selection for Seed Protein in Soybean Breeding Pipeline. *Frontiers in plant science* **13**, 882732-882732 (2022). <https://doi.org/10.3389/fpls.2022.882732>
- 5 Fu, M. *et al.* Detection of Hub QTLs Underlying the Genetic Basis of Three Modules Covering Nine Agronomic Traits in an F2 Soybean Population. *Agronomy* **12** (2022). <https://doi.org/10.3390/agronomy12123135>

- 6 Kumar, V. *et al.* Pinpointing Genomic Regions and Candidate Genes Associated with Seed Oil and Protein Content in Soybean through an Integrative Transcriptomic and QTL Meta-Analysis. *Cells* **12** (2023). <https://doi.org/10.3390/cells12010097>
- 7 Zhang, H., Jiang, H., Hu, Z., Song, Q. & An, Y.-q. C. Development of a versatile resource for post-genomic research through consolidating and characterizing 1500 diverse wild and cultivated soybean genomes. *BMC Genomics* **23**, 250 (2022). <https://doi.org/10.1186/s12864-022-08326-w>
- 8 Feng, W. *et al.* Transgressive Potential Prediction and Optimal Cross Design of Seed Protein Content in the Northeast China Soybean Population Based on Full Exploration of the QTL-Allele System. *Frontiers in Plant Science* **13** (2022). <https://doi.org/10.3389/fpls.2022.896549>
- 9 Hu, B. *et al.* Identification of quantitative trait loci underlying five major agronomic traits of soybean in three biparental populations by specific length amplified fragment sequencing (SLAF-seq). *PeerJ* **9** (2021). <https://doi.org/10.7717/peerj.12416>
- 10 Li, X. *et al.* Mapping QTLs for protein and oil content in soybean by removing the influence of related traits in a four-way recombinant inbred line population. *J. Agric. Sci.* **157**, 659-675 (2019). <https://doi.org/10.1017/S0021859620000040>
- 11 Chapman, A. *et al.* Quantitative trait loci for agronomic and seed quality traits in an F2 and F4:6 soybean population. *Euphytica* **129**, 387-393 (2003). <https://doi.org/10.1023/A:1022282726117>
- 12 Mao, T. *et al.* Identification of quantitative trait loci underlying seed protein and oil contents of soybean across multi-genetic backgrounds and environments. *Plant Breeding* **132**, 630-641 (2013). <https://doi.org/10.1111/pbr.12091>
- 13 Liu, S., Liu, Z., Hou, X. & Li, X. Genetic mapping and functional genomics of soybean seed protein. *Mol. Breed.* **43** (2023). <https://doi.org/10.1007/s11032-023-01373-5>
- 14 Kim, W. J. *et al.* Quantitative Trait Loci (QTL) Analysis of Seed Protein and Oil Content in Wild Soybean (*Glycine soja*). **24**, 4077 (2023). <https://doi.org/10.3390/ijms24044077>
- 15 Goettel, W. *et al.* POWR1 is a domestication gene pleiotropically regulating seed quality and yield in soybean. *Nature Communications* **13**, 3051 (2022). <https://doi.org/10.1038/s41467-022-30314-7>
- 16 Fliege, C. E. *et al.* Fine mapping and cloning of the major seed protein quantitative trait loci on soybean chromosome 20. *Plant J.* **110**, 114-128 (2022). <https://doi.org/10.1111/tpj.15658>
- 17 Marsh, J. I. *et al.* Haplotype mapping uncovers unexplored variation in wild and domesticated soybean at the major protein locus *cqProt-003*. *Theor. Appl. Genet.* (2022). <https://doi.org/10.1007/s00122-022-04045-8>
- 18 Wang, J. *et al.* Genetic mapping high protein content QTL from soybean 'Nanxiadou 25' and candidate gene analysis. *BMC Plant Biol.* **21**, 388 (2021). <https://doi.org/10.1186/s12870-021-03176-2>
- 19 Zhong, Y. *et al.* Identification and Mapping of QTLs for Sulfur-Containing Amino Acids in Soybean (*Glycine max* L.). *J. Agric. Food Chem.* (2022). <https://doi.org/10.1021/acs.jafc.2c05896>
- 20 Kaleri, A. A. *et al.* Recognition of qtl for seed protein and oil content in two soybean recombinat inbred lines populations. *Journal of Animal and Plant Sciences* **31**, 1669-1685 (2021). <https://doi.org/10.36899/JAPS.2021.6.0371>

- 21 Malle, S., Eskandari, M., Morrison, M. & Belzile, F. Genome-wide association identifies several QTLs controlling cysteine and methionine content in soybean seed including some promising candidate genes. *Sci. Rep.* **10**, 21812 (2020).
<https://doi.org/10.1038/s41598-020-78907-w>
- 22 Ma, Y. *et al.* QTL Mapping for Protein and Sulfur-Containing Amino Acid Contents Using a High-Density Bin-Map in Soybean (*Glycine max* L. Merr.). *J. Agric. Food Chem.* **67**, 12313-12321 (2019). <https://doi.org/10.1021/acs.jafc.9b04497>
- 23 Prenger, E. M. *et al.* Introgression of a high protein allele into an elite soybean cultivar results in a high-protein near-isogenic line with yield parity. *Crop Sci.* **59**, 2498-2508 (2019). <https://doi.org/10.2135/cropsci2018.12.0767>
- 24 Li, S., Xu, H., Yang, J. & Zhao, T. Dissecting the genetic architecture of seed protein and oil content in soybean from the yangtze and huaihe river valleys using multi-locus genome- wide association studies. *International Journal of Molecular Sciences* **20** (2019).
<https://doi.org/10.3390/ijms20123041>
- 25 Lee, S. *et al.* Genome-wide association study of seed protein, oil and amino acid contents in soybean from maturity groups I to IV. *Theor. Appl. Genet.* **132**, 1639-1659 (2019). <https://doi.org/10.1007/s00122-019-03304-5>
- 26 Zhang, Y. *et al.* Construction of a high-density genetic map and mapping of QTLs for soybean (*Glycine max*) agronomic and seed quality traits by specific length amplified fragment sequencing. *BMC Genomics* **19** (2018). <https://doi.org/10.1186/s12864-018-5035-9>
- 27 Diers, B. W., Keim, P., Fehr, W. R. & Shoemaker, R. C. RFLP analysis of soybean seed protein and oil content. *Theor. Appl. Genet.* **83**, 608-612 (1992). <https://doi.org/10.1007/BF00226905>
- 28 Chung, J. *et al.* The Seed Protein, Oil, and Yield QTL on Soybean Linkage Group I. *Crop Sci.* **43**, 1053-1067 (2003).
<https://doi.org/10.2135/cropsci2003.1053>
- 29 Sonah, H., O'Donoghue, L., Cober, E., Rajcan, I. & Belzile, F. Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. *Plant Biotechnol. J.* **13**, 211-221 (2015).
<https://doi.org/10.1111/pbi.12249>
- 30 Reinprecht, Y. *et al.* Seed and agronomic QTL in low linolenic acid, lipoxygenase-free soybean (*Glycine max* (L.) Merrill) germplasm. *Genome* **49**, 1510-1527 (2006). <https://doi.org/10.1139/g06-112> %M 17426766
- 31 Li, W., Zheng, D.-H., Van, K. & Lee, S.-H. QTL mapping for major agronomic traits across two years in soybean (*Glycine max* L. Merr.). *J. Crop Sci. Biotechnol* **11**, 171-176 (2008).
- 32 Mansur, L. M., Lark, K. G., Kross, H. & Oliveira, A. Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max* L.). *Theor. Appl. Genet.* **86**, 907-913 (1993). <https://doi.org/10.1007/BF00211040>
- 33 Kong, L. *et al.* Candidate loci for breeding compact plant-type soybean varieties. *Mol. Breed.* **43** (2023).
<https://doi.org/10.1007/s11032-022-01352-2>
- 34 Ayalew, H., Schapaugh, W., Vuong, T. & Nguyen, H. T. Genome-wide association analysis identified consistent QTL for seed yield in a soybean diversity panel tested across multiple environments. *Plant Genome* **15** (2022).
<https://doi.org/10.1002/tpg2.20268>

- 35 Lu, Y. *et al.* Identification of Genomic Regions Associated with Vine Growth and Plant Height of Soybean. *International Journal of Molecular Sciences* **23** (2022). <https://doi.org/10.3390/ijms23105823>
- 36 Tian, Y. *et al.* QTL analysis for plant height and fine mapping of two environmentally stable QTLs with major effects in soybean. *Journal of Integrative Agriculture* **21**, 933-946 (2022). [https://doi.org/10.1016/S2095-3119\(21\)63693-6](https://doi.org/10.1016/S2095-3119(21)63693-6)
- 37 Xue, H. *et al.* Mapping developmental QTL for plant height in soybean [*Glycine max* (L.) Merr.] using a four-way recombinant inbred line population. *PLoS ONE* **14** (2019). <https://doi.org/10.1371/journal.pone.0224897>
- 38 Orf, J. H. *et al.* Genetics of Soybean Agronomic Traits: I. Comparison of Three Related Recombinant Inbred Populations. *Crop Sci.* **39**, 1642-1651 (1999). <https://doi.org/10.2135/cropsci1999.3961642x>
- 39 Palomeque, L. *et al.* QTL in mega-environments: II. Agronomic trait QTL co-localized with seed yield QTL detected in a population derived from a cross of high-yielding adapted × high-yielding exotic soybean lines. *Theor. Appl. Genet.* **119**, 429-436 (2009). <https://doi.org/10.1007/s00122-009-1048-8>
- 40 Han, Y. *et al.* QTL analysis of soybean seed weight across multi-genetic backgrounds and environments. *Theor. Appl. Genet.* **125**, 671-683 (2012). <https://doi.org/10.1007/s00122-012-1859-x>
- 41 Kato, S. *et al.* A major and stable QTL associated with seed weight in soybean across multiple environments and genetic backgrounds. *Theor. Appl. Genet.* **127**, 1365-1374 (2014). <https://doi.org/10.1007/s00122-014-2304-0>
- 42 Copley, T. R., Duceppe, M.-O. & O'Donoghue, L. S. Identification of novel loci associated with maturity and yield traits in early maturity soybean plant introduction lines. *BMC Genomics* **19**, 167 (2018). <https://doi.org/10.1186/s12864-018-4558-4>
- 43 Hu, Z. *et al.* Association mapping of yield-related traits and SSR markers in wild soybean (*Glycine soja* Sieb. and Zucc.). *Breeding Science* **63**, 441-449 (2014). <https://doi.org/10.1270/jsbbs.63.441>
- 44 Wang, D., Graef, G. L., Procopiuk, A. M. & Diers, B. W. Identification of putative QTL that underlie yield in interspecific soybean backcross populations. *Theor. Appl. Genet.* **108**, 458-467 (2004). <https://doi.org/10.1007/s00122-003-1449-z>
- 45 Lee, S. H. *et al.* Molecular Markers Associated with Soybean Plant Height, Lodging, and Maturity across Locations. *Crop Sci.* **36**, crops1996.0011183X003600030035x (1996). <https://doi.org/10.2135/cropsci1996.0011183X003600030035x>
- 46 Yoosefzadeh-Najafabadi, M. *et al.* Machine-Learning-Based Genome-Wide Association Studies for Uncovering QTL Underlying Soybean Yield and Its Components. *International Journal of Molecular Sciences* **23** (2022). <https://doi.org/10.3390/ijms23105538>
- 47 Fang, C. *et al.* Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biology* **18**, 161 (2017). <https://doi.org/10.1186/s13059-017-1289-9>
- 48 Bachlava, E., Dewey, R. E., Burton, J. W. & Cardinal, A. J. Mapping and Comparison of Quantitative Trait Loci for Oleic Acid Seed Content in Two Segregating Soybean Populations. *Crop Sci.* **49**, 433-442 (2009). <https://doi.org/10.2135/cropsci2008.06.0324>
- 49 Zhang, J. *et al.* Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genomics* **16**, 217 (2015). <https://doi.org/10.1186/s12864-015-1441-4>
- 50 Specht, J. E. *et al.* Soybean Response to Water: A QTL Analysis of Drought Tolerance. *Crop Sci.* **41**, 493-509 (2001). <https://doi.org/10.2135/cropsci2001.412493x>

- 51 Xia, Z. *et al.* QNE1 is a key flowering regulator determining the length of the vegetative period in soybean cultivars. *Science China Life Sciences* **65**, 2472-2490 (2022). <https://doi.org/10.1007/s11427-022-2117-x>
- 52 Hao, D. *et al.* Identification of single nucleotide polymorphisms and haplotypes associated with yield and yield components in soybean (Glycine max) landraces across multiple environments. *Theor. Appl. Genet.* **124**, 447-458 (2012). <https://doi.org/10.1007/s00122-011-1719-0>
- 53 Han, Y. *et al.* Unconditional and conditional QTL underlying the genetic interrelationships between soybean seed isoflavone, and protein or oil contents. *Plant Breeding* **134**, 300-309 (2015). <https://doi.org/10.1111/pbr.12259>
- 54 Huang, J. *et al.* Identification and Mapping of Stable QTLs for Seed Oil and Protein Content in Soybean [Glycine max (L.) Merr.]. *J. Agric. Food Chem.* **68**, 6448-6460 (2020). <https://doi.org/10.1021/acs.jafc.0c01271>
- 55 Yao, Y. *et al.* Quantitative trait loci analysis of seed oil content and composition of wild and cultivated soybean. *BMC Plant Biol.* **20**, 51 (2020). <https://doi.org/10.1186/s12870-019-2199-7>