Supplementary material for:

Comprehensive genomic resources related to domestication and breeding traits in Lima bean

Tatiana Garcia[1,7,8], Jorge Duitama[2,7], Stephanie Smolenski Zullo[3], Juanita Gil[2,9], Andrea Ariani[3,10], Sarah Dohle[3], Antonia Palkovic[3], Paola Skeen[1,3], Clara Isabel Bermudez-Santana[4], Daniel G. Debouck[5], Jaime Martínez-Castillo[6], Paul Gepts[3], Maria Isabel Chacón-Sánchez[1,*]

[1] Departamento de Agronomía, Facultad de Ciencias Agrarias, Universidad Nacional de Colombia, Bogotá, Colombia
[2] Systems and Computing Engineering Department, Universidad de los Andes, Bogotá, Colombia
[3] Department of Plant Sciences / MS1, University of California, Davis 95616-8780, CA, USA
[4] Departmento de Biología, Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia
[5] Centro Internacional de Agricultura Tropical, Cali, Colombia
[6] Centro de Investigación Científica de Yucatán, Yucatán, Mexico
[7] Co-first authors
[8] Present address: Biochemistry and Molecular Biology, Michigan State University, East Lansing, 48824, MI, USA
[9] Present address: Department of Entomology and Plant Pathology, University of Arkansas, 72701, Fayetteville, AR, USA.
[10] Present address: BASF BBCC - Innovation Center Gent, Belgium
[*] Corresponding author: michacons@unal.edu.co
Correspondence can also be addressed to ja.duitama@uniandes.edu.co

Supplementary tables

Supplementary table 1. Physical position and genetic distance of markers mapped on each chromosome (Separate Excel file).

Supplementary table 2. Recombination rates across euchromatic and pericentromeric regions for each chromosome. (Separate Excel file).

Supplementary table 3. Summary of repetitive elements identified in the Lima bean genome assembly. (Separate excel file).

Supplementary table 4. QTL peaks, LOD scores, % variation explained and QTL effect for select traits in the UC 92 x UC Haskell population (Separate Excel file).

Supplementary table 5. Genomic locations of genes with agronomic annotations (Separate Excel file).

Supplementary table 6. Genes related to resistance to biotic stresses (Separate Excel file).

Supplementary table 7. List of accessions genotyped in this study, including latitude, longitude and altitude of the collection site and phenotypic information if available (Separate Excel file).

Supplementary table 8. Population assignment by Structure assuming 6 subpopulations (Separate Excel file).

Supplementary table 9. Basic diversity statistics of gene pools in wild and domesticated Lima bean calculated on the basis of 454 accessions and 12,398 SNP markers. A total of 28 accessions that resulted to be admixed in the Structure analysis were not included. (Separate Excel file)

Supplementary table 10. Pairwise FST distance matrix among gene pools in wild and domesticated Lima bean, as measured on the basis of 451 accessions and 19,384 SNP markers. (Separate Excel File)

Supplementary table 11. List of introgressions identified in 52 accessions. The table includes the genomic location, sample id, haplotpye similarity score for each major population, the population of origin and the population of the introgressed haplotype (Separate Excel file).

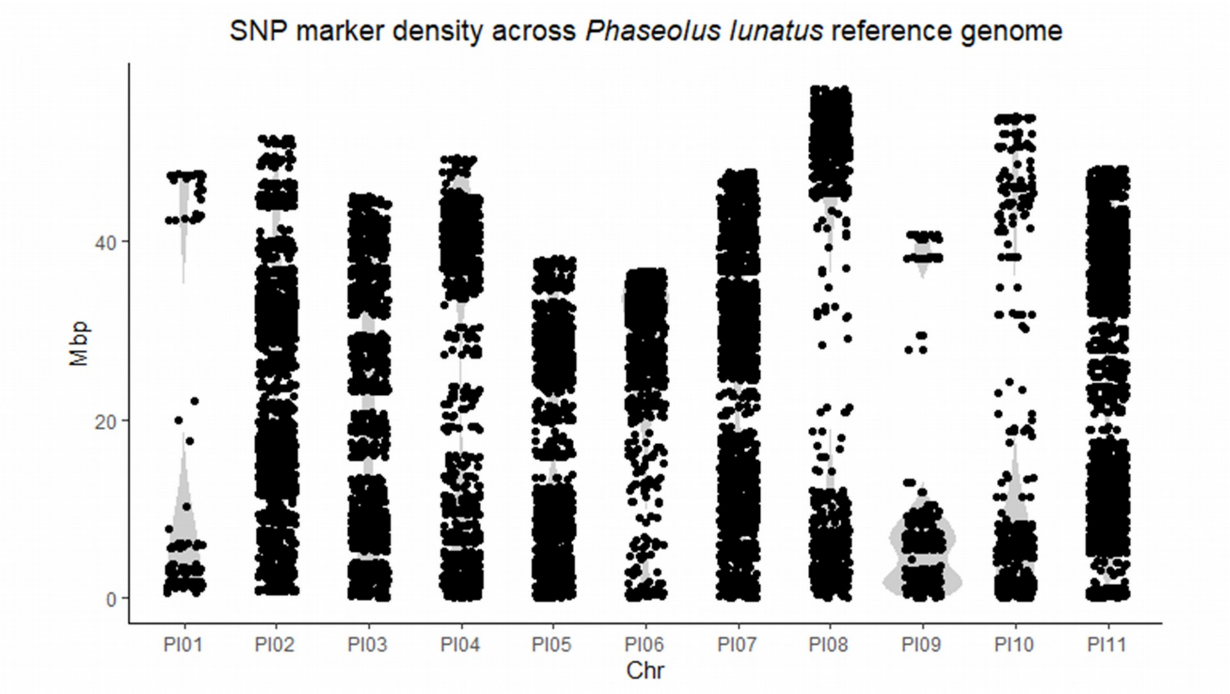
Supplementary table 12. List of genes with differential expression values (DEGs) either between developmental stages or between accessions. Functional annotation and expression TPM values are included for each gene (Separate Excel file).

Supplementary table 13. Summary of sequencing data used to generate the de novo genome assembly

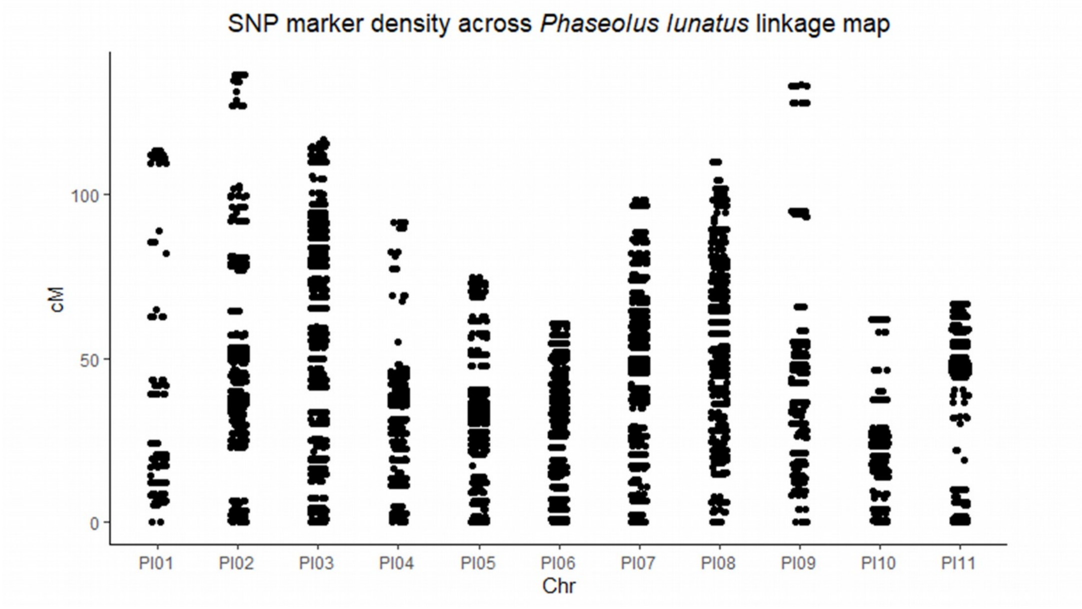
Technology	Protocol	Number of reads (Million)	Raw data (Gbp)
Pacbio	SMRT	2	25.6
Illumina	Paired-end	2 x 103	31
Illumina	Linked reads (10x)	2 x 138	41
Total		243	97.6

Supplementary figures

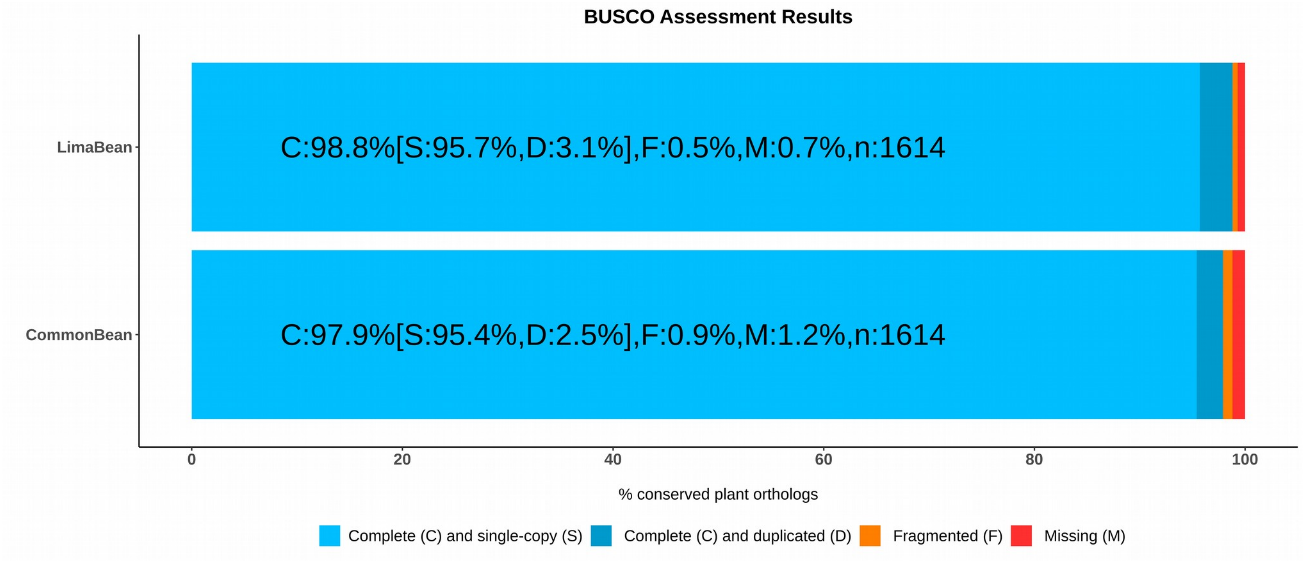
Supplementary figure 1. SNP density for the biparental population



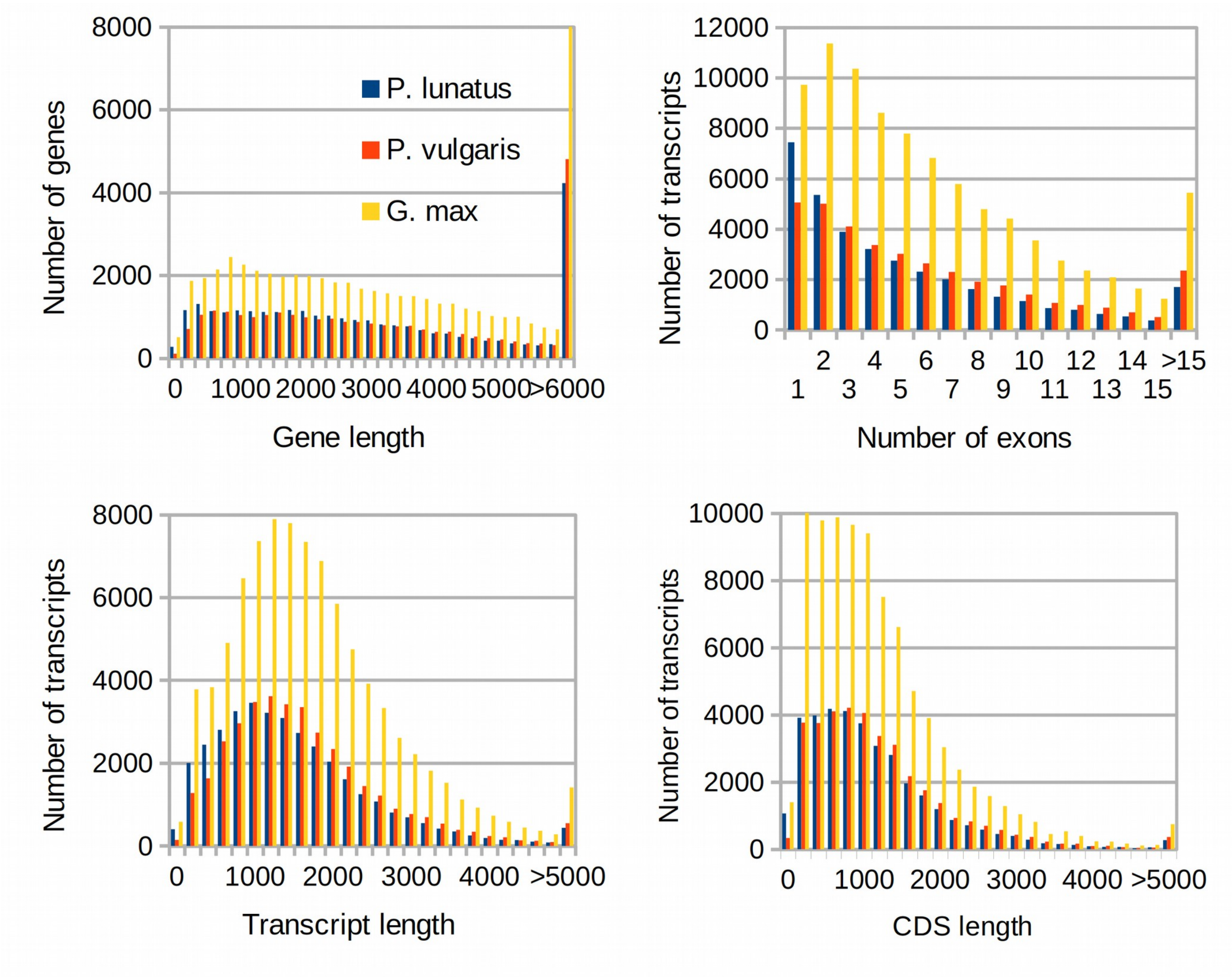
Supplementary figure 2. Locus distribution on the *Phaseolus lunatus* genetic map for the UC 92 x UC Haskell RIL population



Supplementary figure 3. Percentage of conserved genes in the plant kingdom (assessed by Busco) that can be found in the Lima Bean (*Phaseolus lunatus*) and common bean (*Phaseolus vulgaris*) genomes

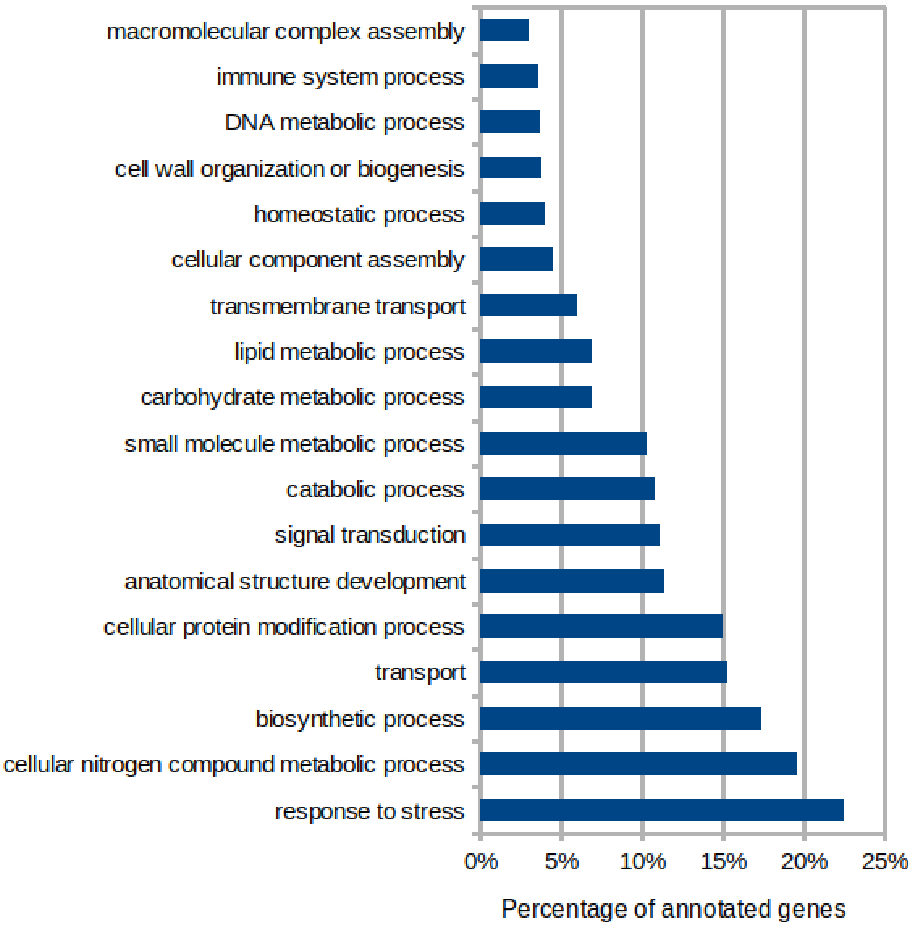


Supplementary figure 4. Distribution of gene, exon, transcript ans CDS lengths, for annotated genes in the *P. lunatus* genome. The same distributions for *P. vulgaris* and *G. max* (Soy bean) are shown for comparison.

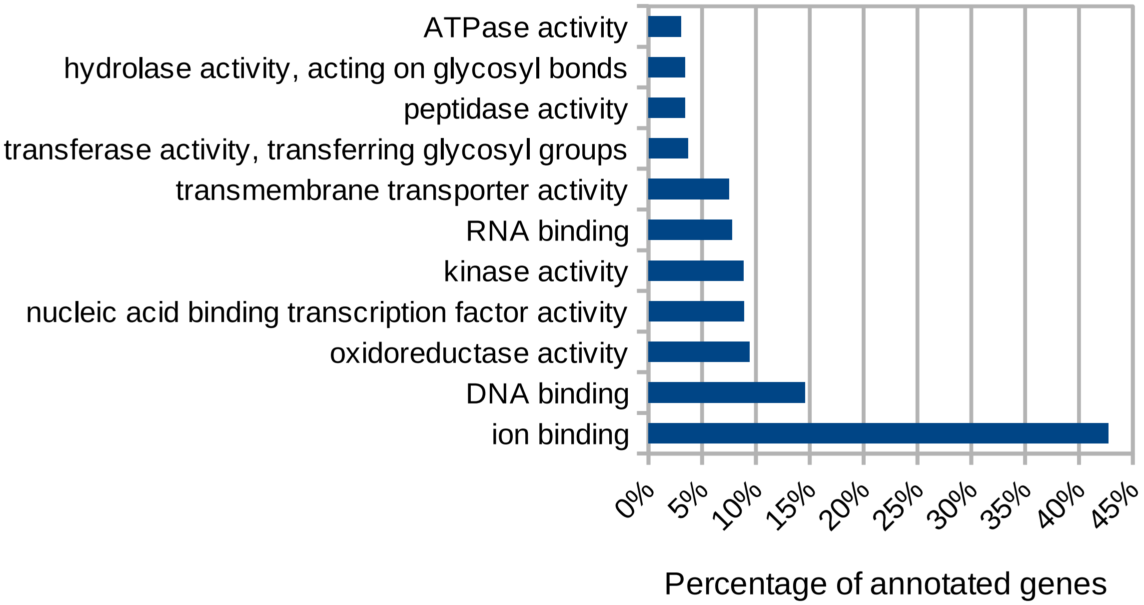


Supplementary figure 5. Gene ontology annotations assigned to at least 3% of the gene models. A. Biological processes. B. Molecular functions

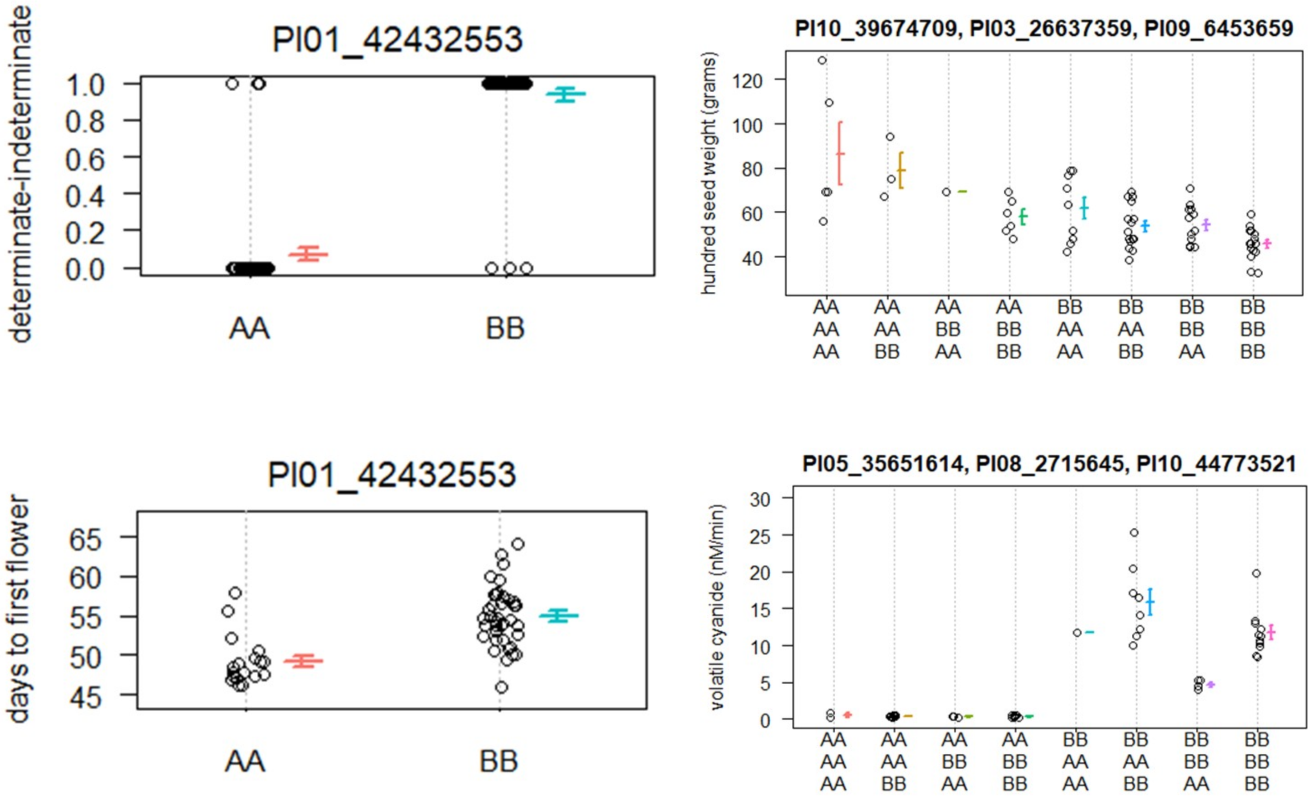
A.



B.

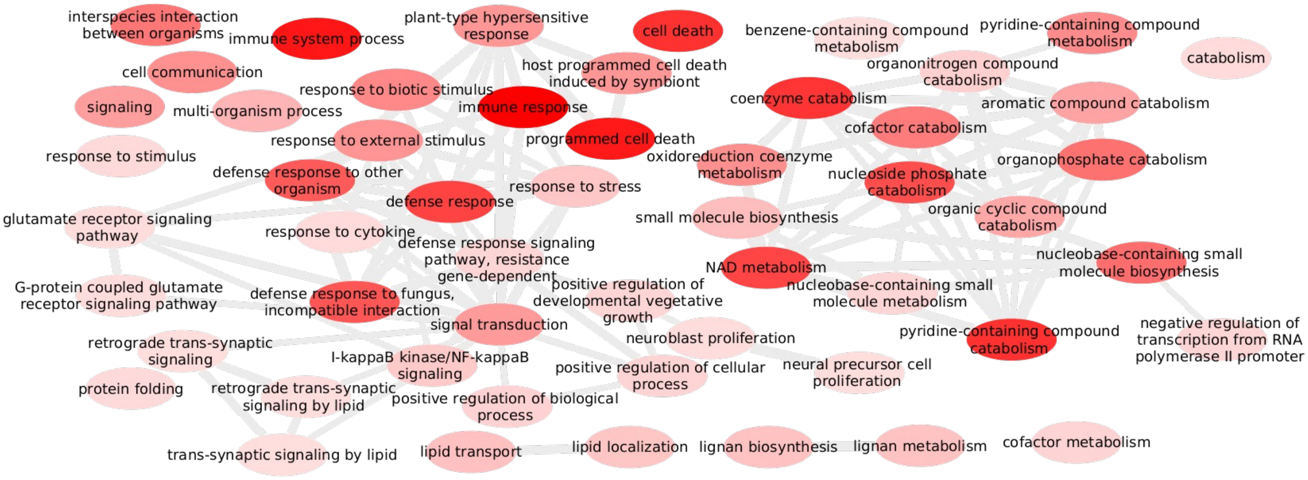


Supplementary figure 6. QTL effect for significant QTL peaks for each trait where ‘AA’ represents the UC 92 parental haplotype and ‘BB’ represents the UC Haskell parental haplotype.

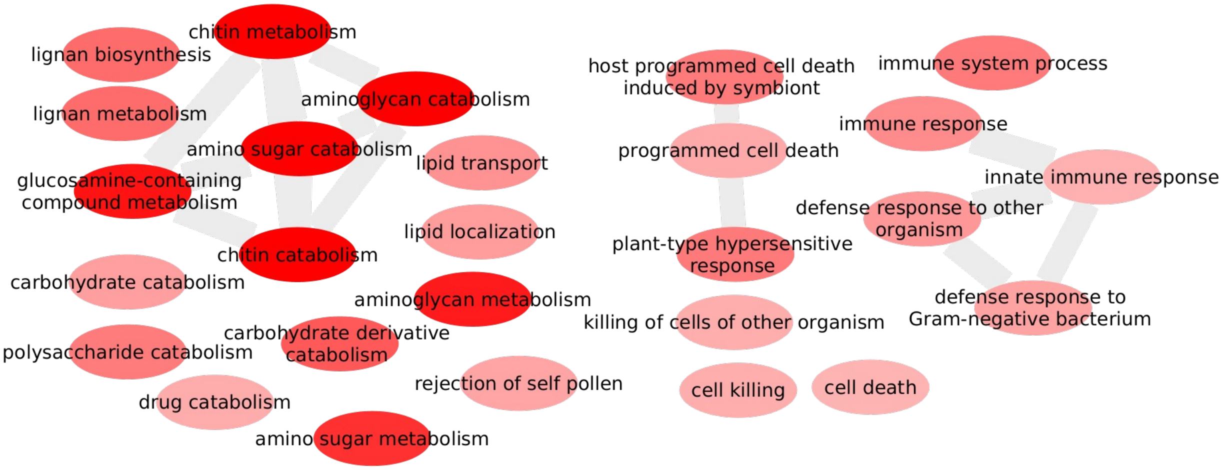


Supplementary figure 7. Enrichment of GO terms in A. same chromosome paralogs and B. synteny orthologs with Ka/Ks values above 1.

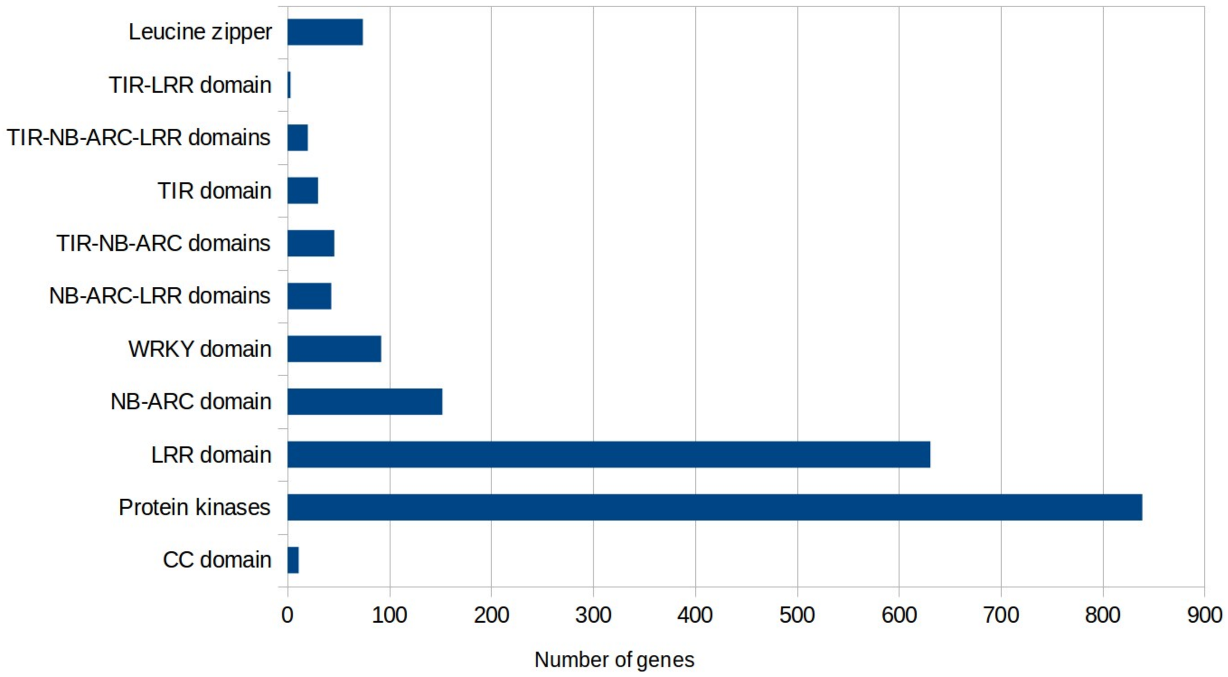
A.



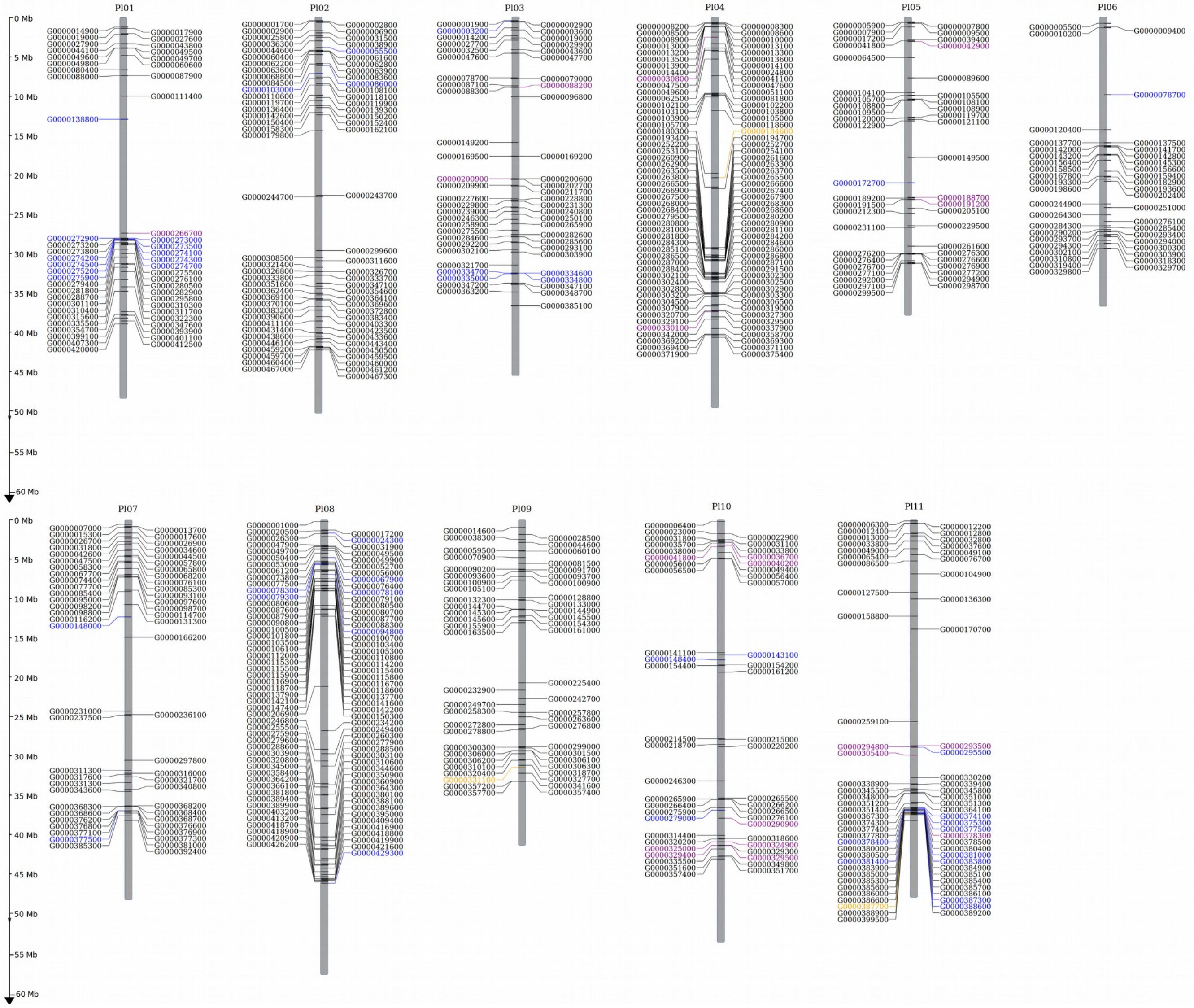
B.



Supplementary figure 8. LRR gene statistics



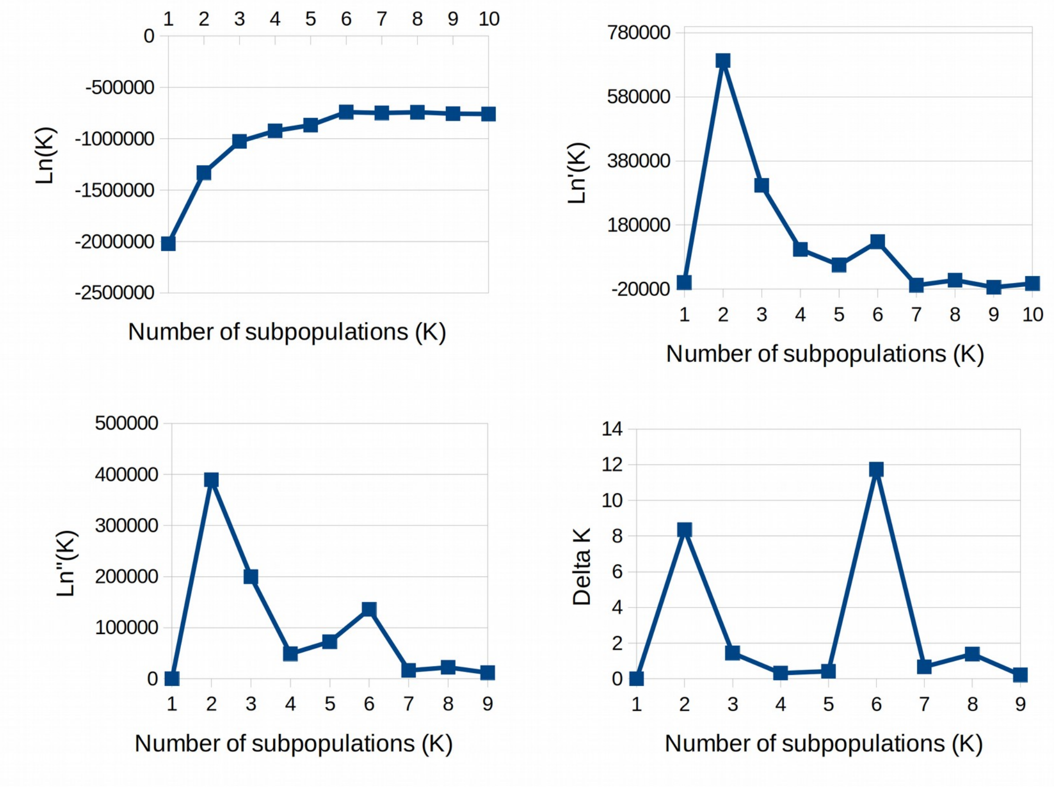
Supplementary figure 9. Physical map of the 11 chromosomes of the Lima bean genome showing resistance related genes with the LRR domain. Each gene has a unique label corresponding to the last 11 digits from the annotation ID (for example, gene G0000138800 on chromosome Pl01 corresponds to gene Pl01G0000138800). Black are genes only with the LRR domain, blue are genes with domains NB-ARC and LRR, purple are genes with domains TIR, NB-ARC and LRR, orange are genes with the TIR and LRR domains.



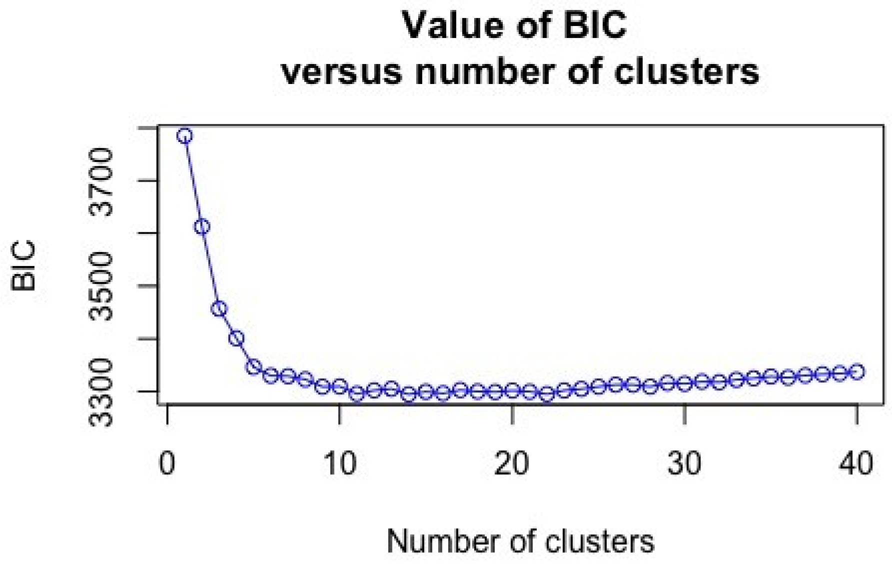
Supplementary figure 10. General geographic distribution of accessions genotyped in this study. Wild accessions are shown in the top map and domesticated accessions are shown in the bottom map.

Supplementary figure 11. Results of A. the Evanno test and B. The DAPC Bayesian information criterion test, to infer the optimal number of subpopulations

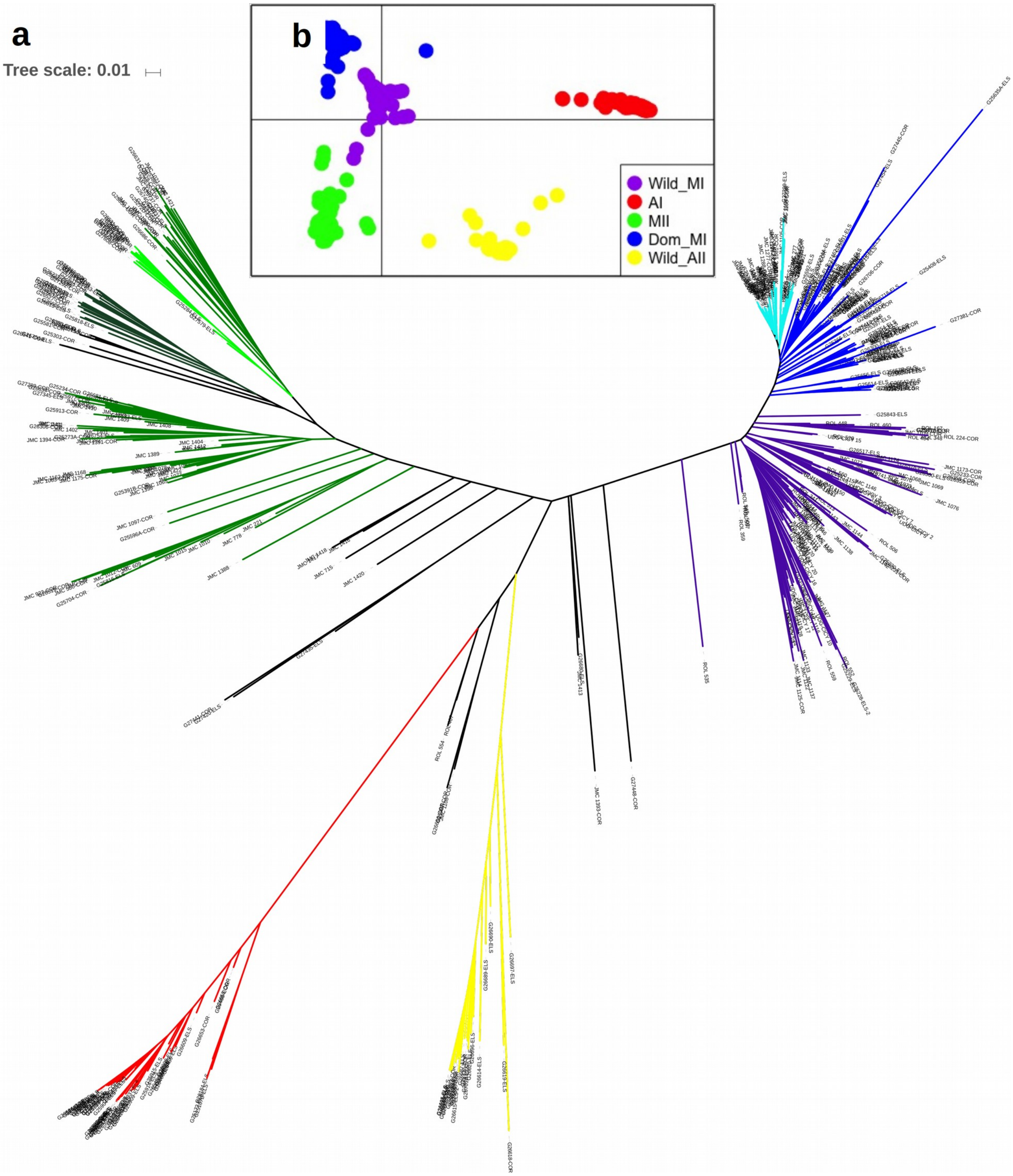
A.



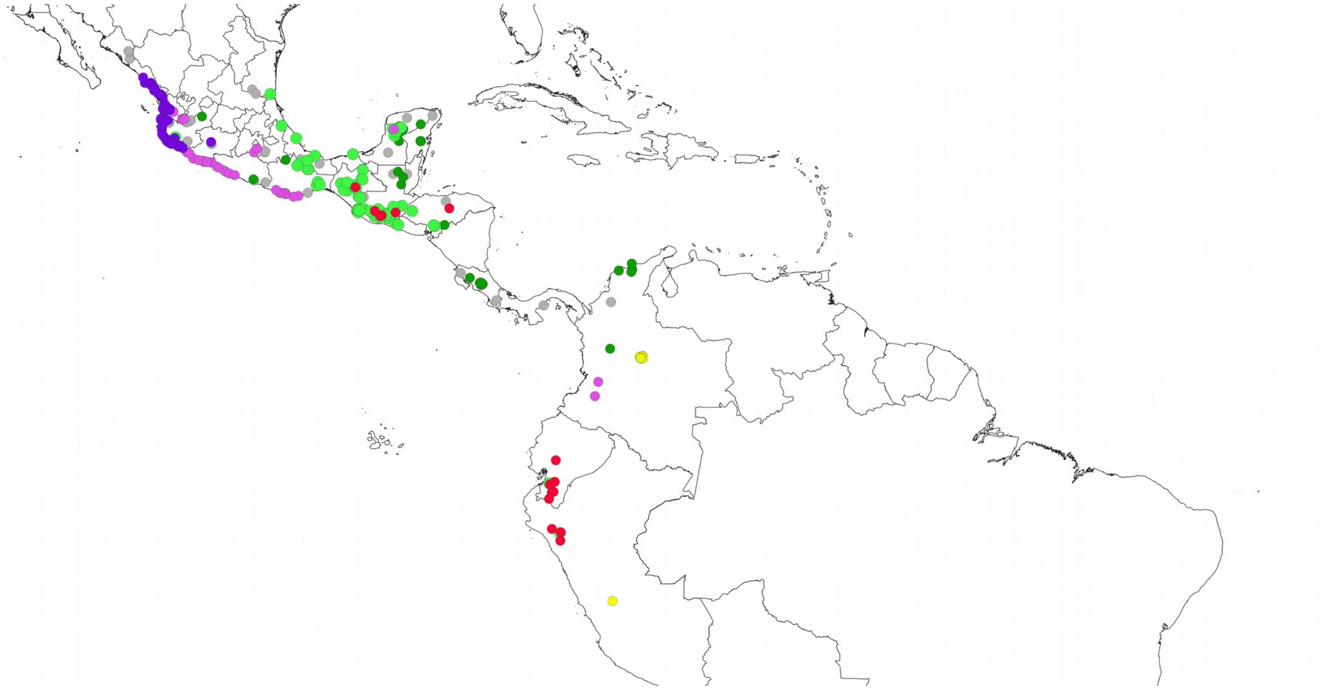
B.

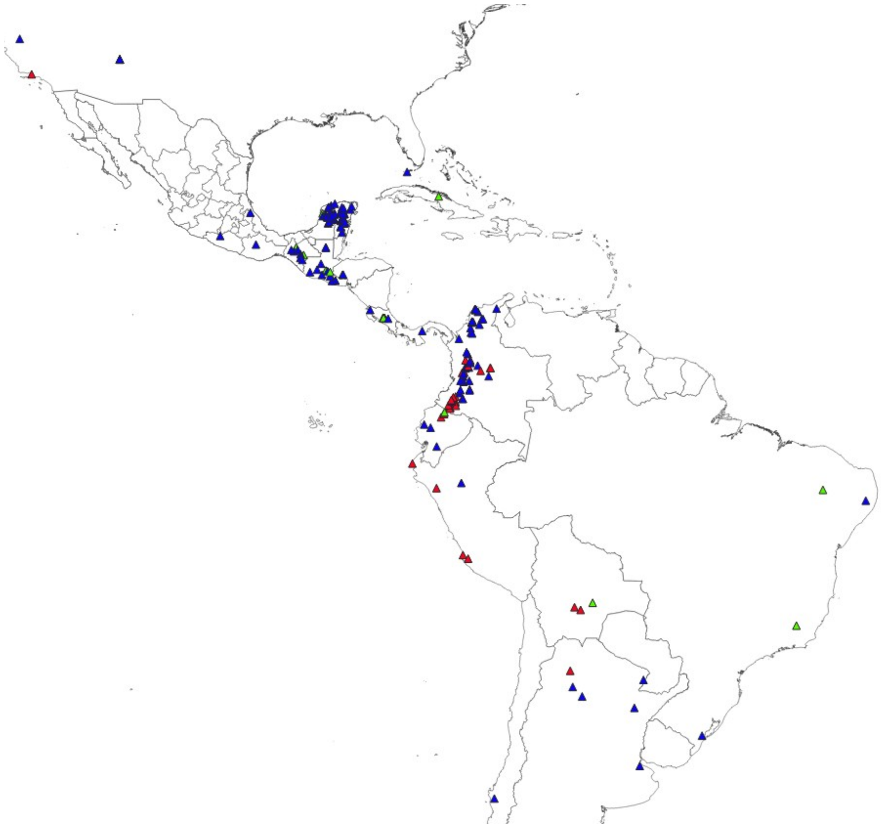


Supplementary figure 12. **a.** Neighbor-joining tree and **b.** discriminant analysis of principal components (DAPC) analysis of the genetic variability between 482 wild and domesticated Lima bean accessions collected across the Americas. Color coding for branches in the NJ tree is according to gene pools of accessions as followed: purple: wild MI, red: AI, dark green: wild MII, light green: domesticated MII, dark blue: domesticated MI, light blue: domesticated MI from Yucatan Peninsula and yellow: AII.

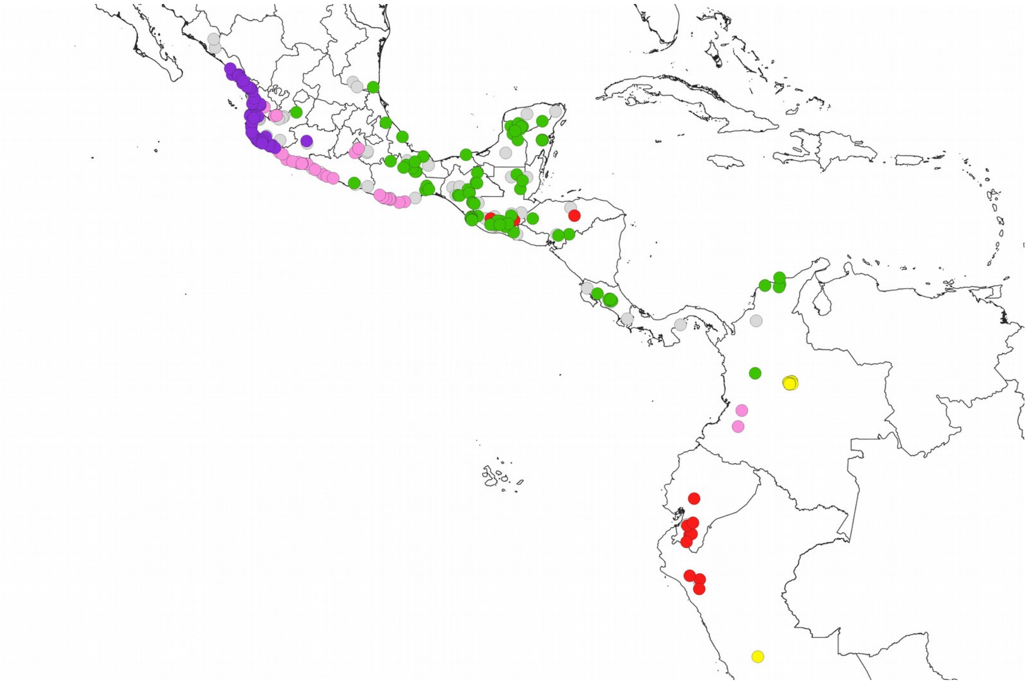


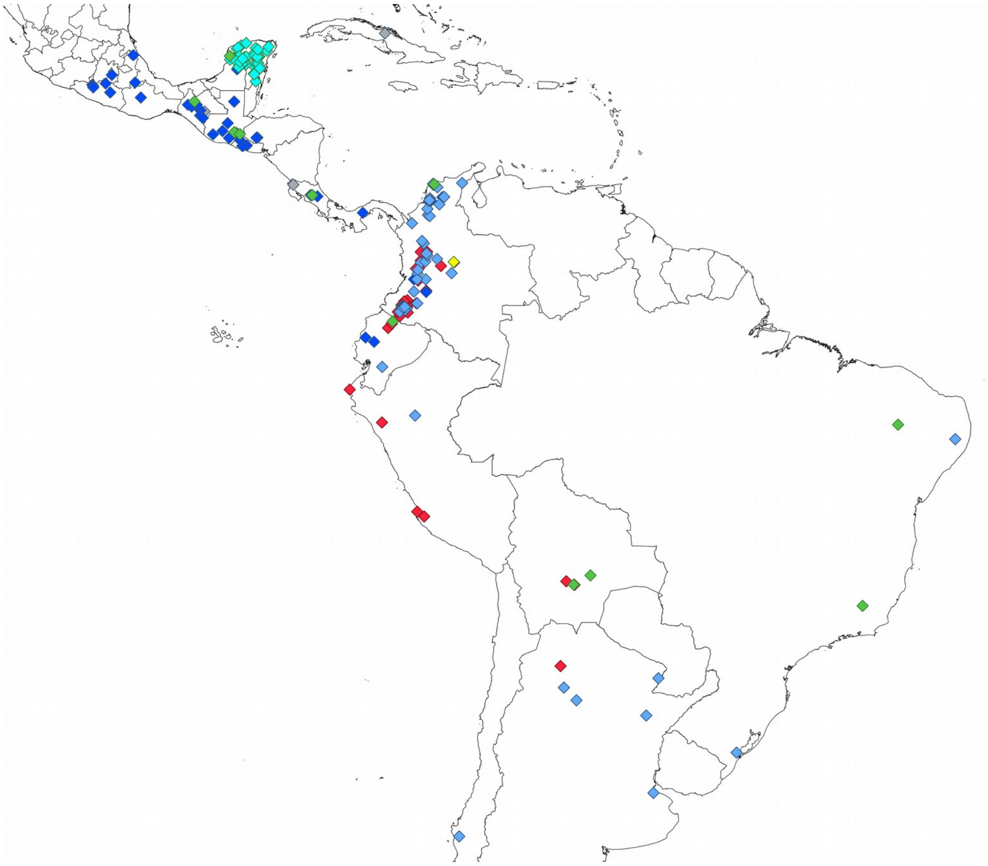
Supplementary figure 13. Geographic distribution of accessions genotyped in this study, clustered according to the high level population structure (K=6). Wild accessions are shown in the top map and domesticated accessions are shown in the bottom map.



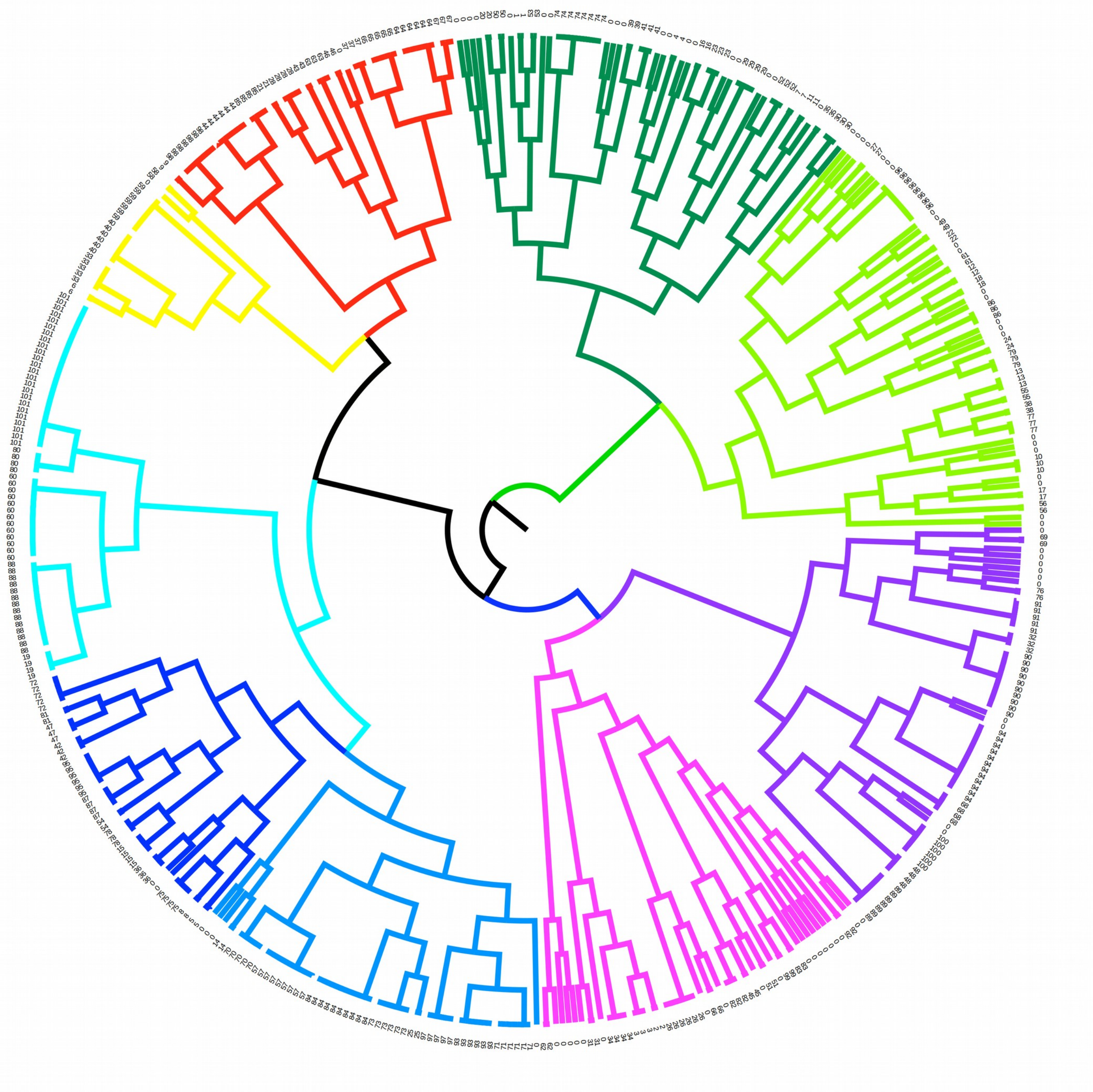


Supplementary figure 14. Geographic distribution of accessions genotyped in this study, clustered according to the clustering inferred by FineStructure. Wild accessions are shown in the top map and domesticated accessions are shown in the bottom map.

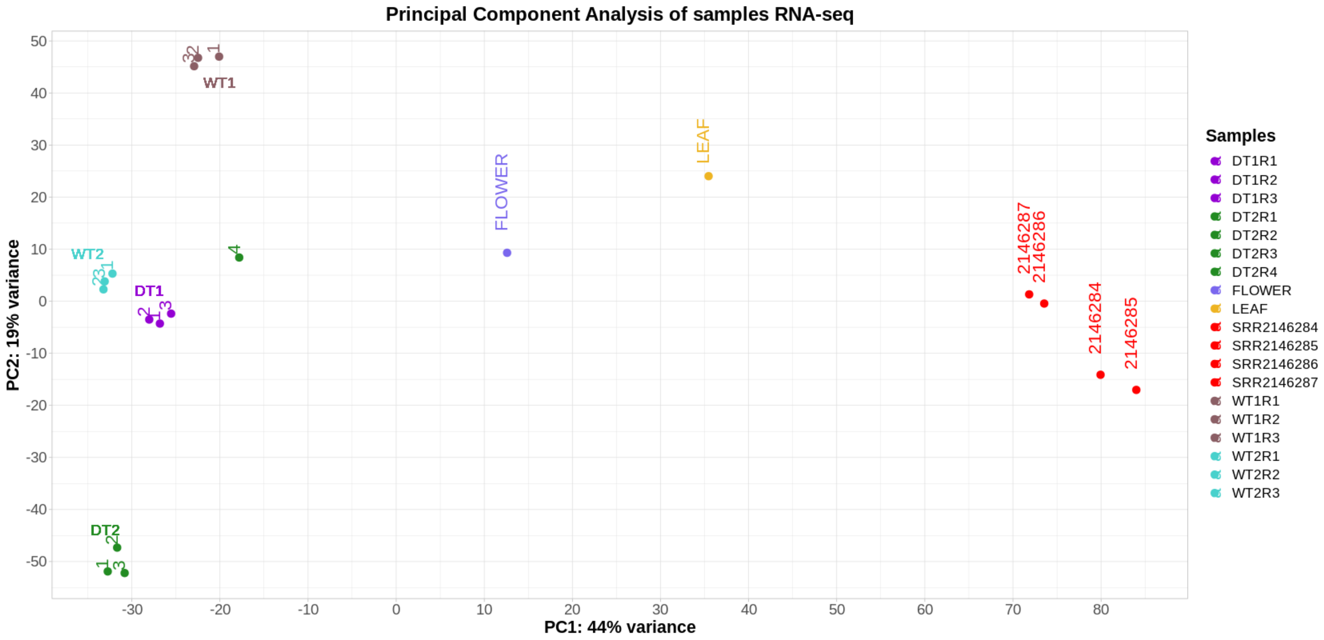




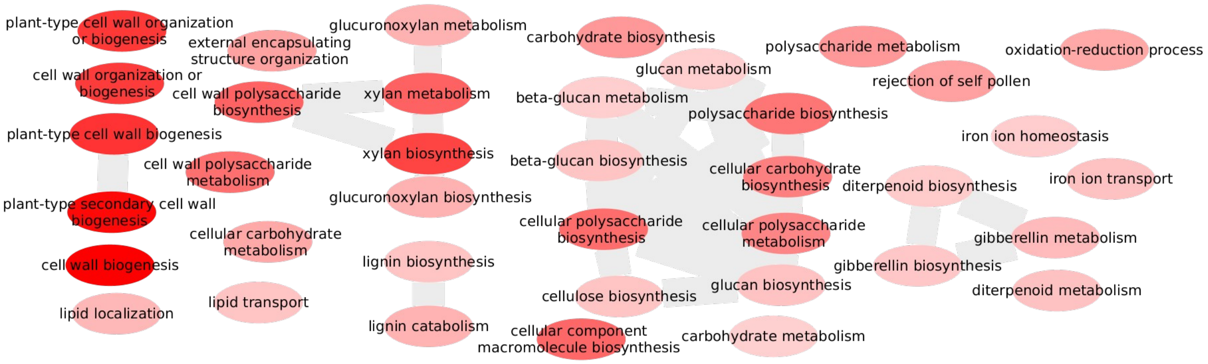
Supplementary figure 15. Radial clustering of the 482 accessions according to the analysis performed by fineSTRUCTURE. Major gene pools are shown by different colors (purple cluster: wild MI from northern-western Mexico; pink cluster: wild MI from southern-western Mexico; medium blue cluster: domesticated MI from South America; dark blue cluster: domesticated MI from Mexico and Central America (CA); light blue cluster: domesticated MI from Yucatan Peninsula; yellow cluster: AII gene pool; red cluster: AI gene pool; green cluster: MII gene pool from Yucatan, central america and Colombia; light green cluster: MII gene pool from southern and central Mexico). Population numbers are included for each accesion



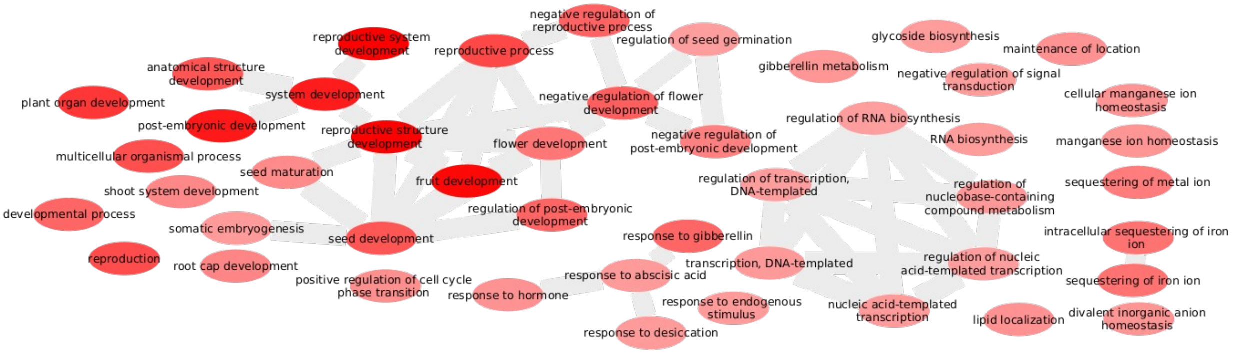
Supplementary figure 16. Principal component analysis (PCA) of normalized expression values for the complete set of samples used to annotate the reference genome. Based on this analysis, the sample DT2R4 was removed for analysis of differential expression among pod samples between developmental times and between accessions.



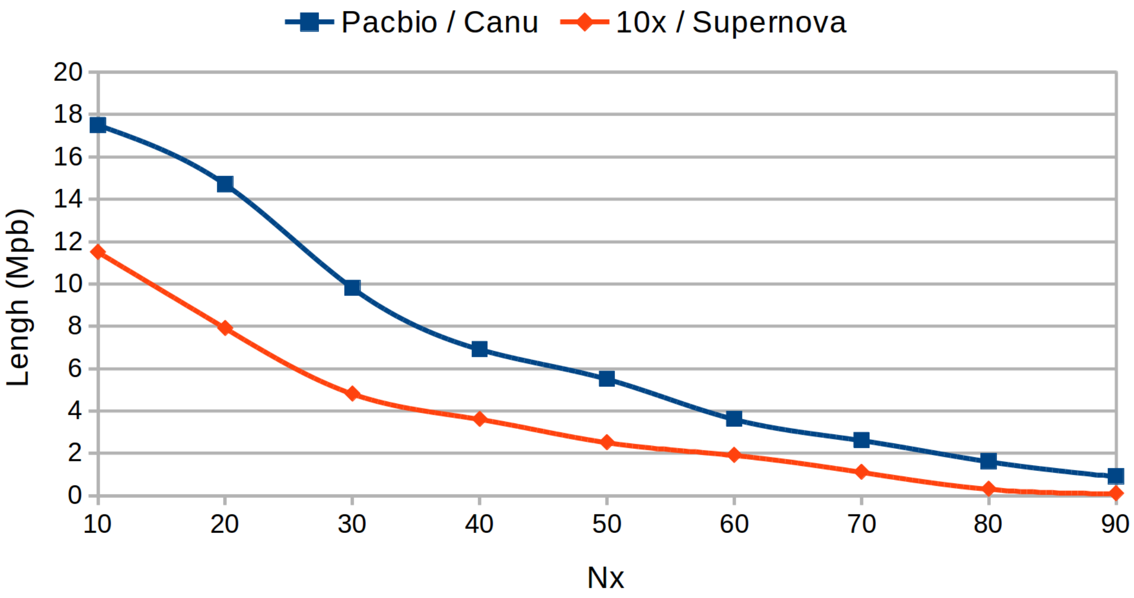
Supplementary figure 17. Functional enrichment of genes with increased expression between the first and the second developmental time for both the wild and the domesticated accessions.



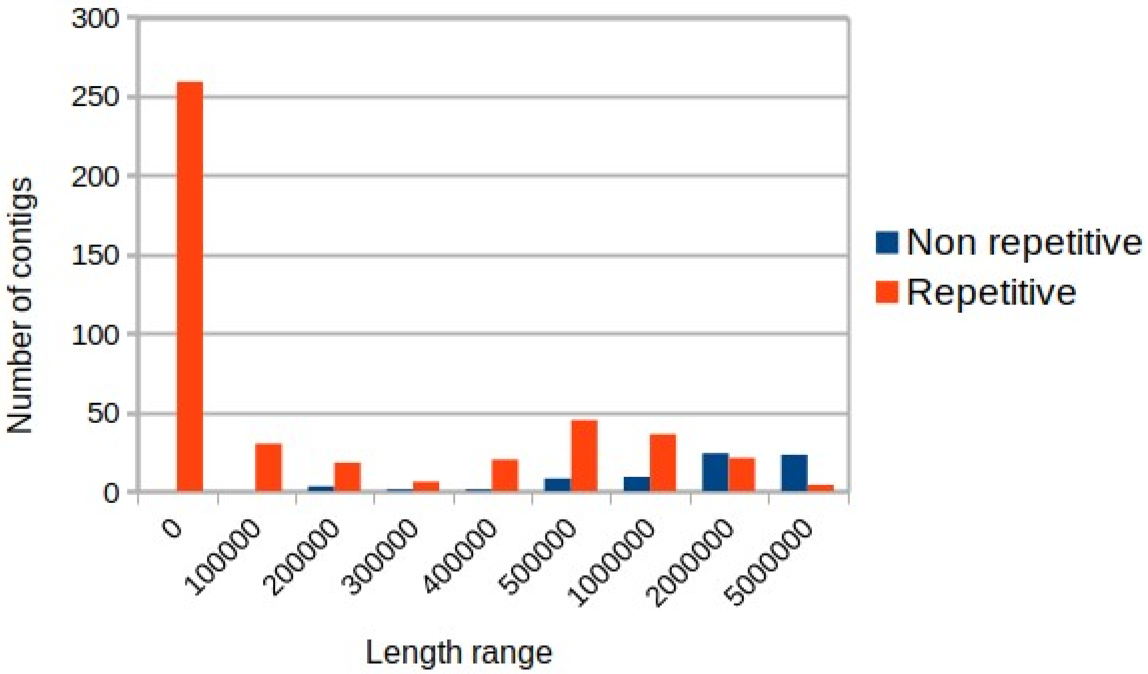
Supplementary figure 18. Functional enrichment of genes with increased expression in the domesticated accession relative to the wild accession only at the second developmental time.



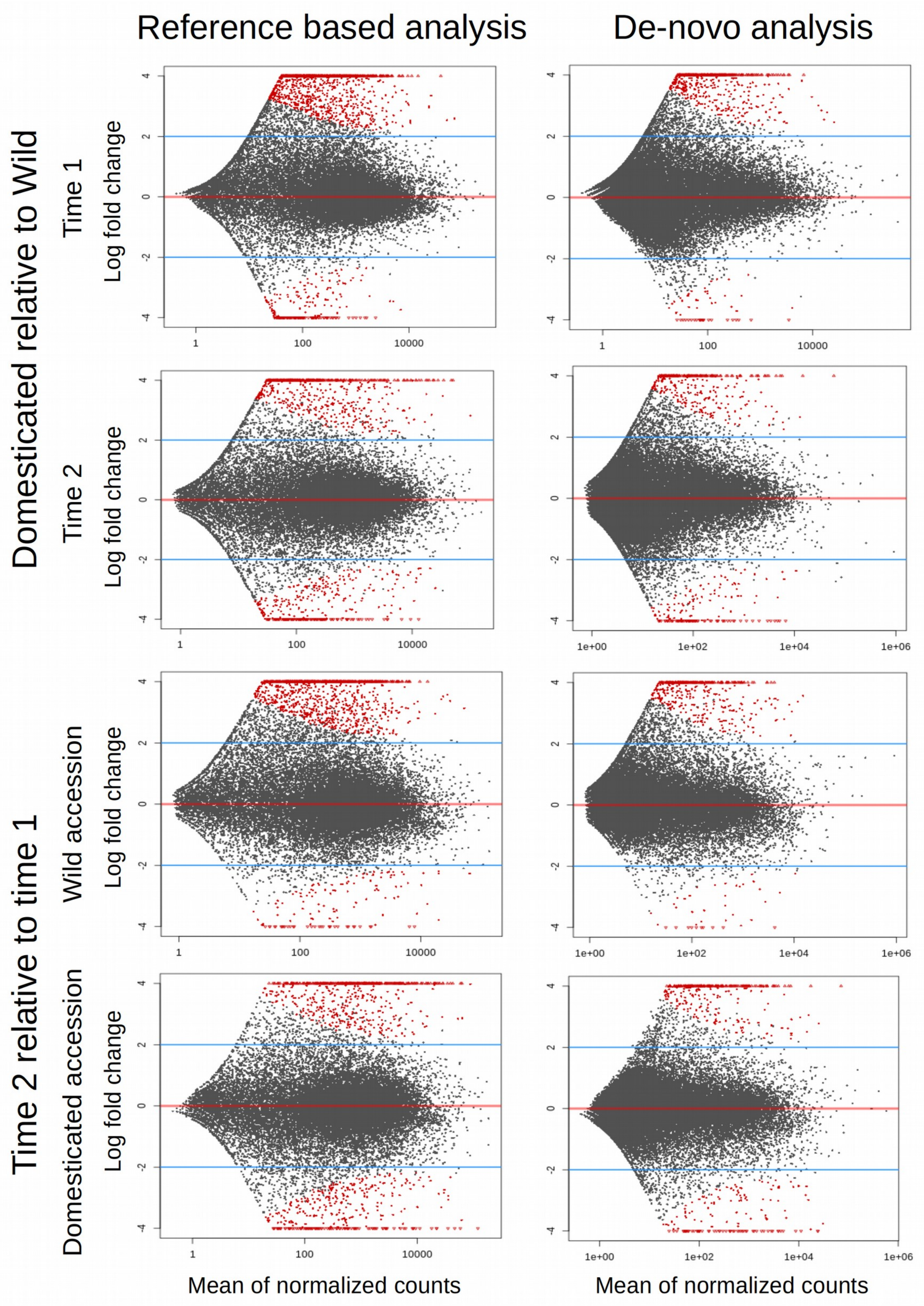
Supplementary figure 19. Nx distribution for assemblies of PacBio reads with Canu and assembly of 10x data using Supernova



Supplementary figure 20. Distribution of contig lengths discriminated as repetitive and non repetitive contigs.



Supplementary figure 21. Distribution of log fold changes in expression levels inferred from RNA-seq data for the domesticated accession taking the wild accession as a control (two upper panels) and for the developmental stage 2 taking the developmental stage 1 as a control. The red dots indicate significant changes as predicted by DESeq2.



Supplementary figure 22. Clustering of DOM-MI and wild MII accessions to show introgression of MII haplotypes within domesticated MI accessions. Red accessions have introgression at chromosome Pl07 and yellow accessions have introgression at Pl08. A. Genome-wide SNPs, B. SNPs at chromosome PL07 between 36 and 39 Mbp. C. SNPs at chromosome Pl08 between 4.6 and 7 Mbp