Supplementary materials for:

Recent genome evolution, domestication, and metabolism of cyanide compounds in Lima bean

Erick Duarte[1,2,+], Tatiana Garcia[3,+], Federico Zuñiga[1], Johanna Stepanian[1], Valerie Parra[1], Juan Pablo Londoño[1,5], Paula Siaucho[4], Edwin Bautista[4], Santiago Jimenez[4], Liza Romero[4], Jessica Ospina[6], Fabio Herrera[2], Natalia Duarte[1], Daniela Lozano[1], Laura Natalia Gonzalez-Garcia[1,7], Romain Guyot[7], Alejandro Reyes[5], Andres Fernando Gonzalez Barrios[2], Maria Isabel Chacón[3], Jorge Duitama[1,*]

1. Systems and Computing Engineering Department, Universidad de los Andes, Bogotá, Colombia
2. Grupo de Diseño de Productos y Procesos (GDPP), Department of Chemical and Food Engineering, Universidad de los Andes, Bogotá, Colombia
3. Facultad de Ciencias Agrarias, Universidad Nacional de Colombia, Bogotá, Colombia
4. Biotecnología y Genética SAS, Biotecgen, Bogotá, Colombia
5. Research Group on Computational Biology and Microbial Ecology, Max Planck Tandem Group in Computational Biology, Department of Biological Sciences, Universidad de Los Andes, Bogotá, Colombia
6. Facultad de Ciencias Agrarias, Universidad Nacional de Colombia Sede Palmira, Palmira, Colombia
7. UMR DIADE, Institut de Recherche pour le Développement (IRD), CIRAD, Université de Montpellier, 34394, Montpellier, France
+ Co-first authors
* Corresponding author: ja.duitama@uniandes.edu.co

Supplementary Tables (Separate excel file)

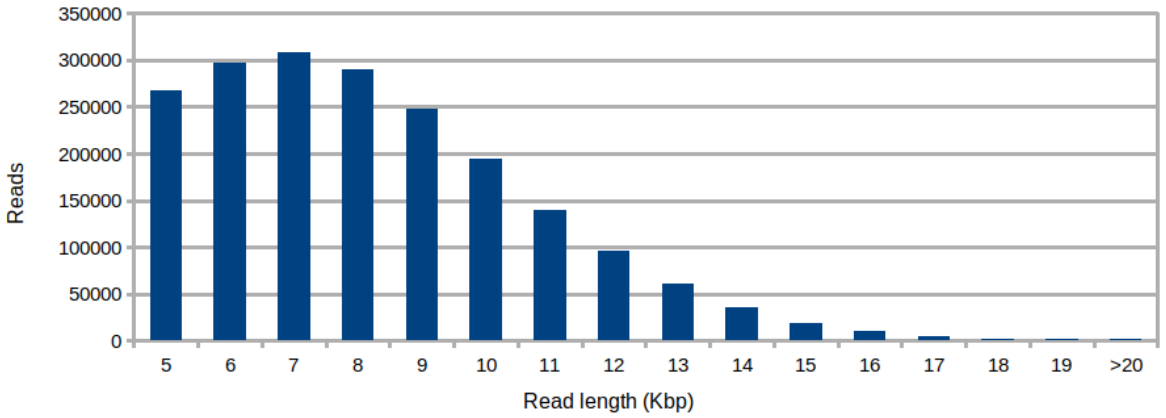
Supplementary table 1. Chromosome lengths for the genome assemblies of the Lima bean accessions G27455 and G25900.
Supplementary table 2. Summary of transposable elements annotated in the assembly of G25900.
Supplementary table 3. Number of genes and functional annotation of orthogroups.
Supplementary table 4. Genes enclosed in structural variants predicted from the sequence-based alignment of the genome assemblies of G27455 and G25900.
Supplementary table 5. Genes enclosed in structural variants predicted from the alignment of ONT reads of the wild Mesoamerican accessions G25228 and G25231, relative to the genome assembly of G27455.
Supplementary table 6. Genes enclosed in structural variants predicted from the alignment of ONT reads of the wild Andean accession G25915 relative to the genome assembly of G25900.
Supplementary table 7. Orthology of genes related to agronomic traits.
Supplementary table 8. Statistics of automated reconstructions of genome-scale metabolic models.
Supplementary table 9. P-values of comparisons of number of samples with evidence of expression for genes grouped by metabolic pathway groups.
Supplementary table 10. Location and expression information for genes in common bean related to the production of HCN.
Supplementary table 11. Location and expression information for genes in Mesoamerican Lima bean related to the production of HCN.

Supplementary Files

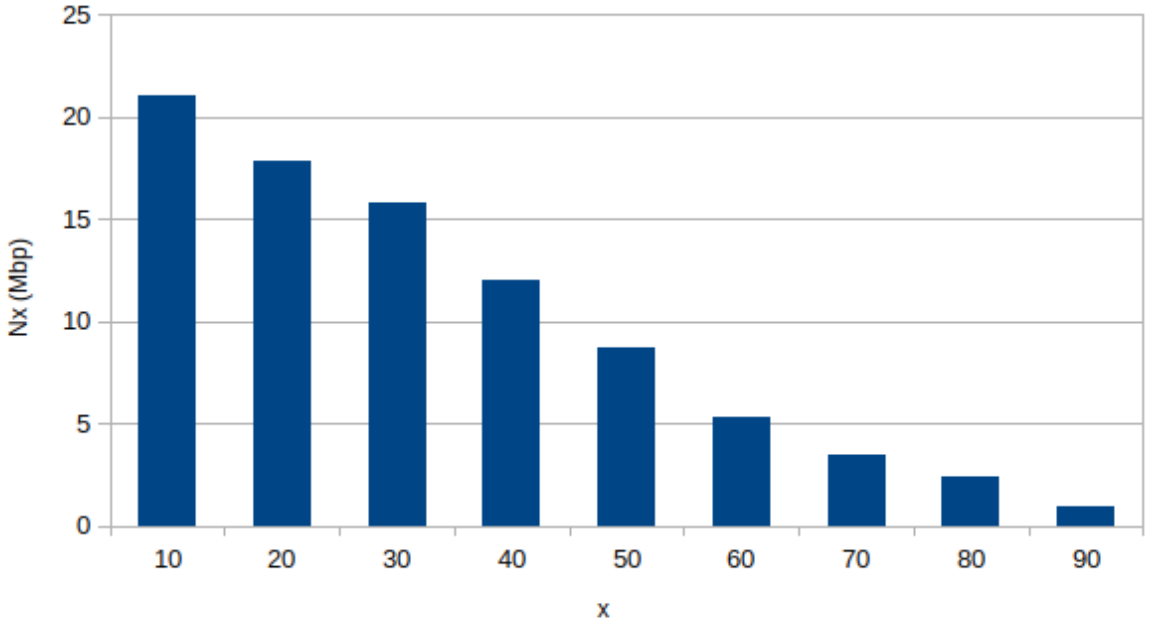
Supplementary file 1. Orthogroups between the five genomes analyzed in this study.
Supplementary file 2. Structural variants predicted from the sequence-based alignment of the genome assemblies of G27455 and G25900. Coordinates are relative to the genome assembly of G27455.
Supplementary file 3. Structural variants predicted from the alignment of ONT reads
Supplementary file 4. Visualization of read alignments around genes related to agronomic traits.
Supplementary file 5. Raw counts and TPM values for 140 common bean RNA-seq samples.
Supplementary file 6. Annotation of gene models for the genome assembly of G25900.
Supplementary file 7. Core metabolic network.

Supplementary Figures

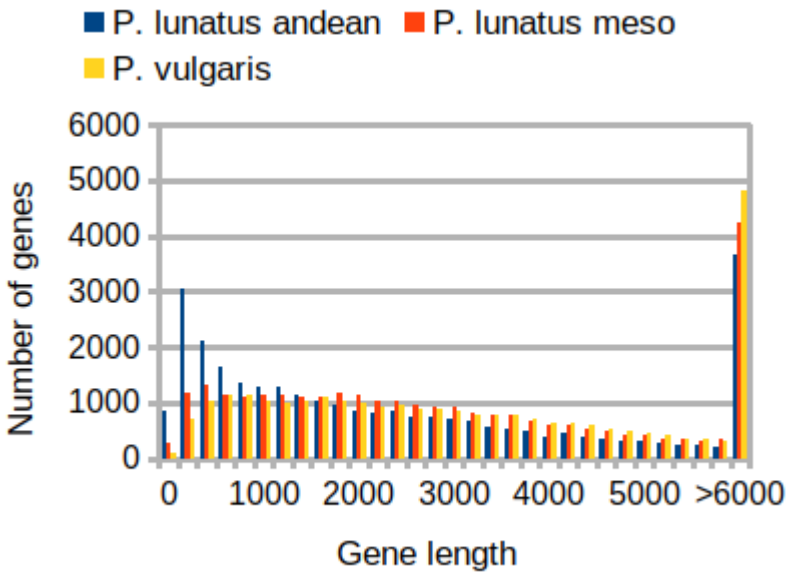
Supplementary Figure 1. Distribution of read length for PacBio HiFi reads taken from the domesticated Andean lima bean accession G25900.

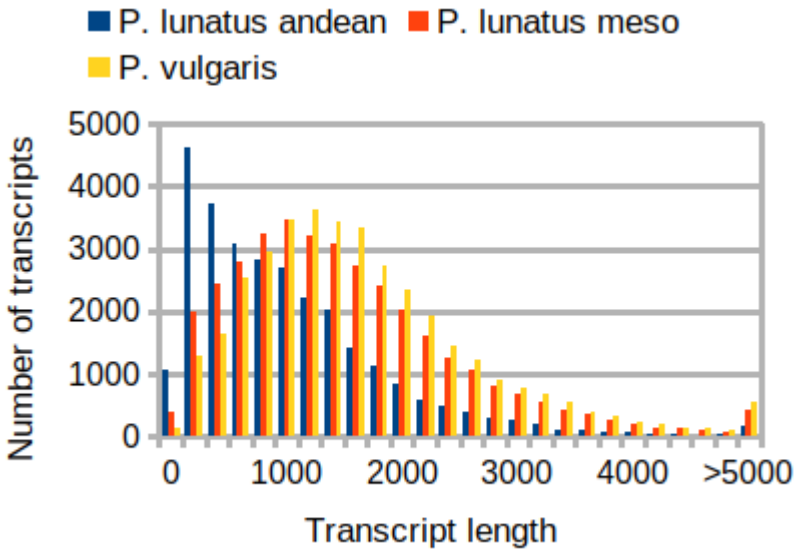


Supplementary Figure 2. Distribution of the Nx statistic for the contig level assembly of the domesticated Andean lima bean accession G25900

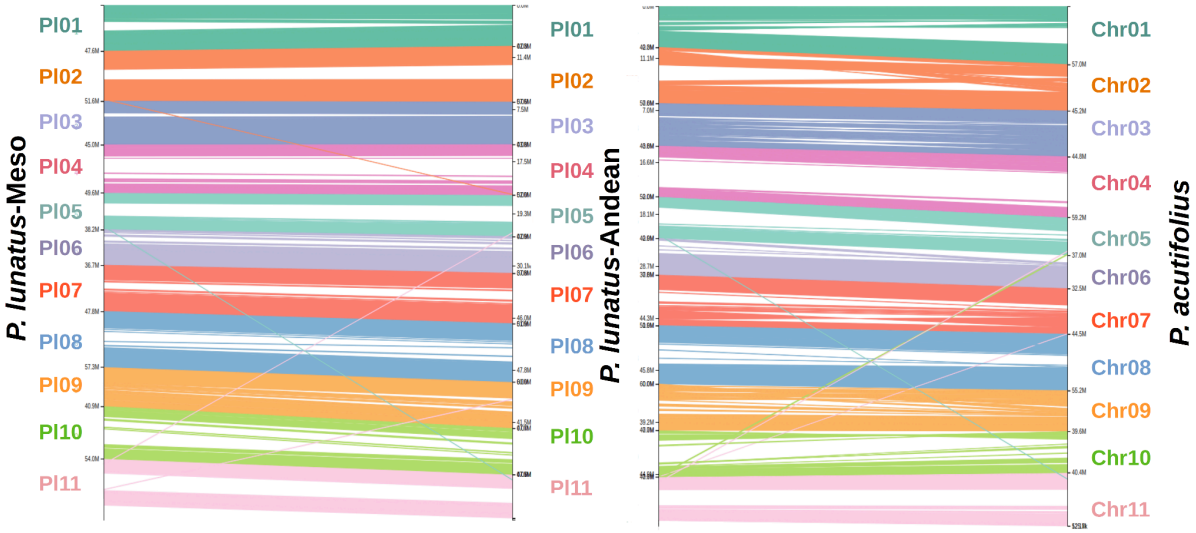


Supplementary Figure 3. Distribution of gene and transcript lengths for the Andean *P. lunatus* gene annotation.



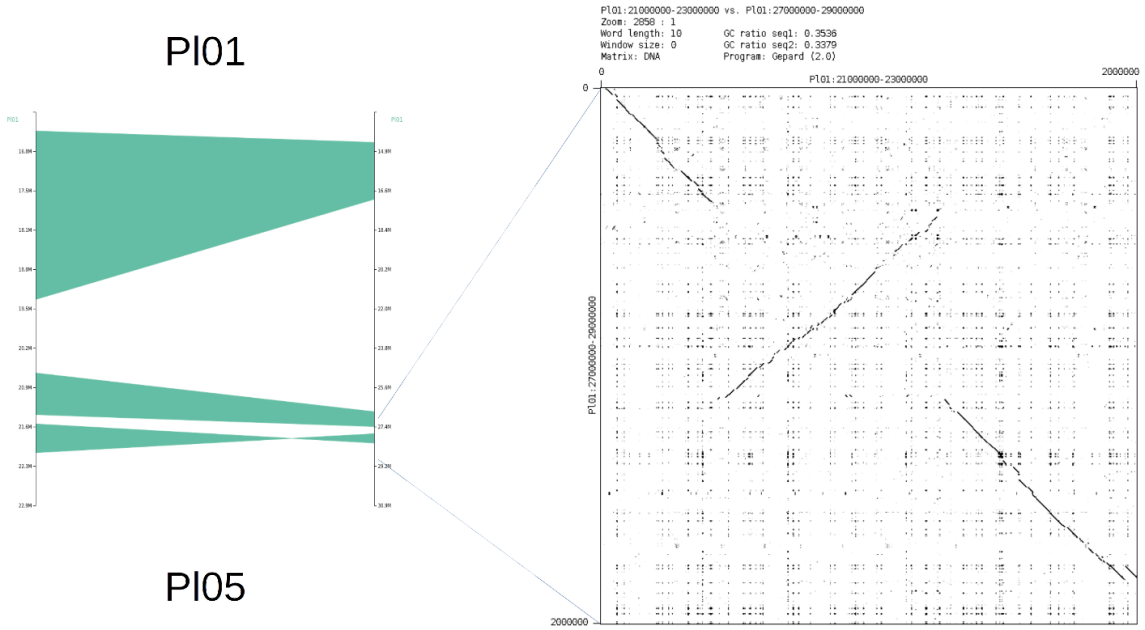


Supplementary Figure 4. Synteny between the mesoamerican and the andean genomes of lima bean.

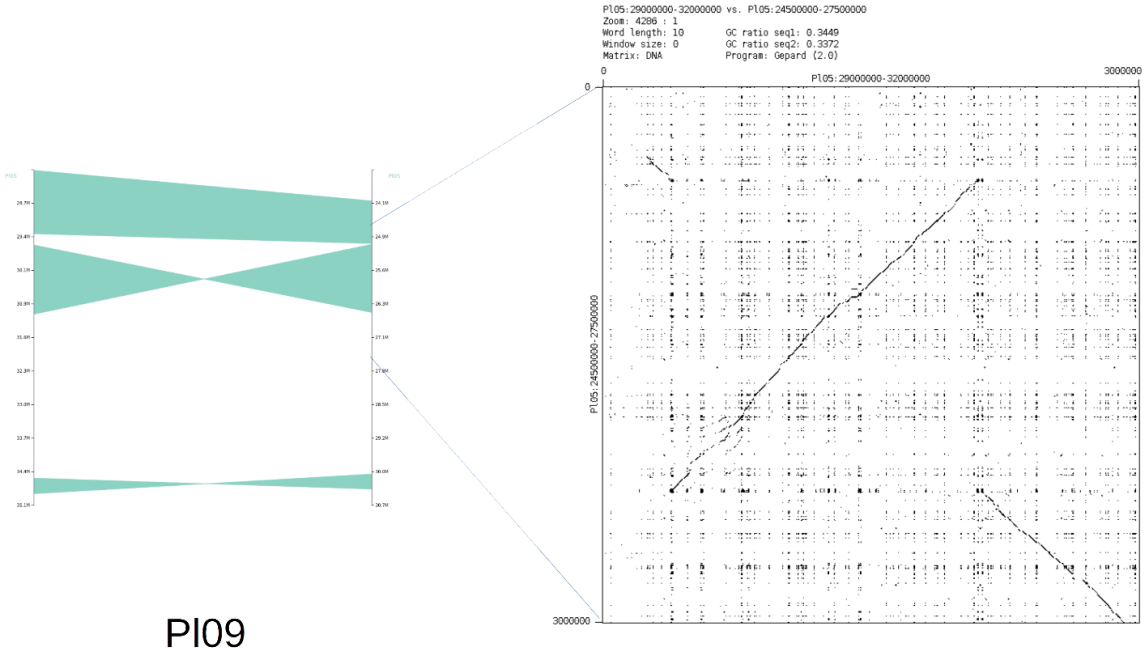


Supplementary Figure 5. Major structural variants between Mesoamerican and Andean Lima bean. The left figures show the local synteny between homologous chromosomes. The right panels show dot plots of the DNA sequence around the event

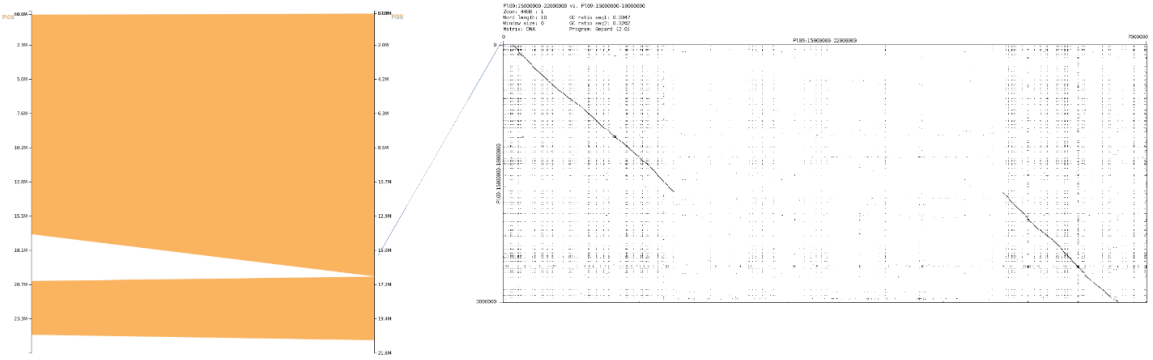
Pl01



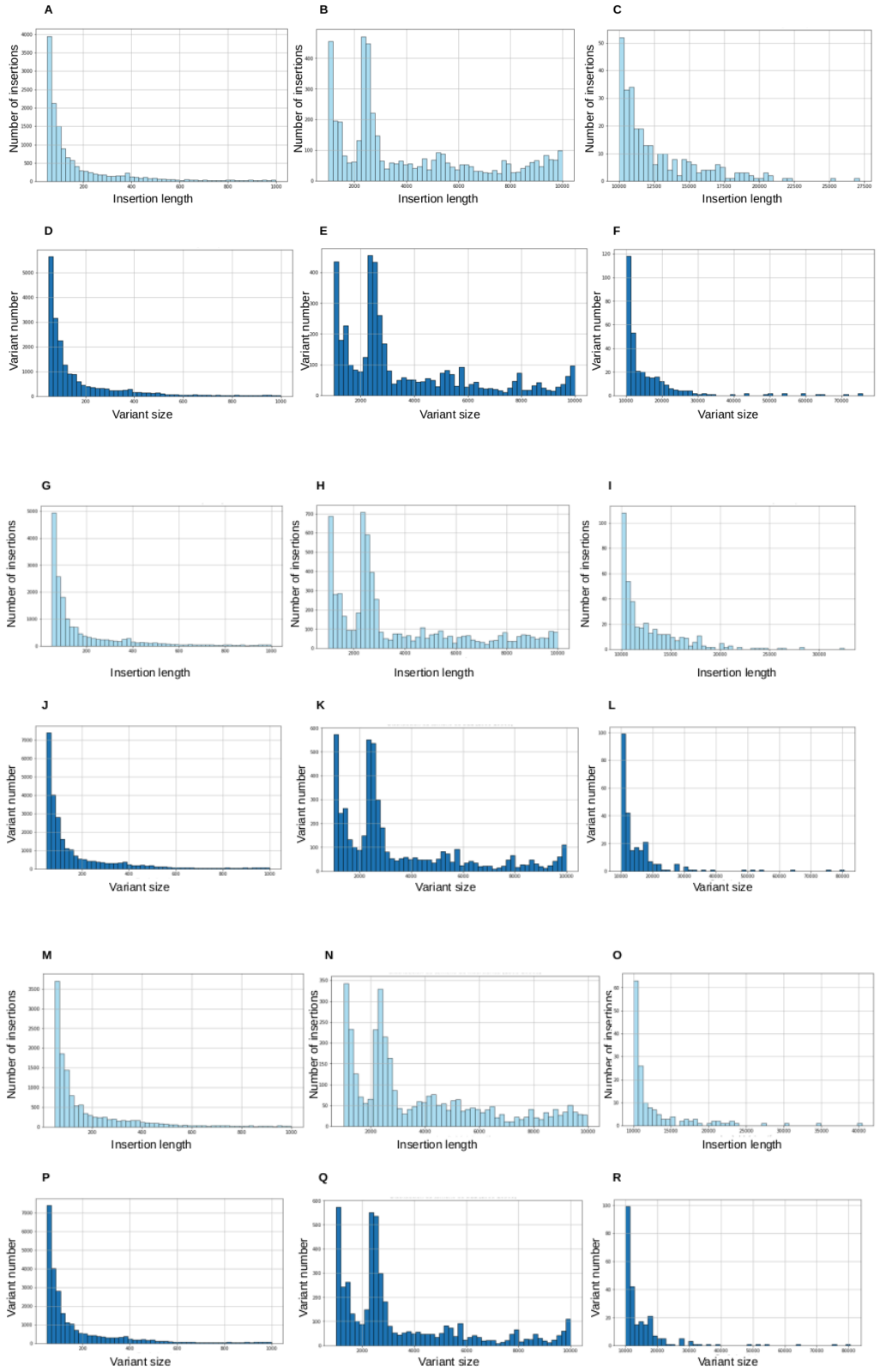
Pl05



Pl09

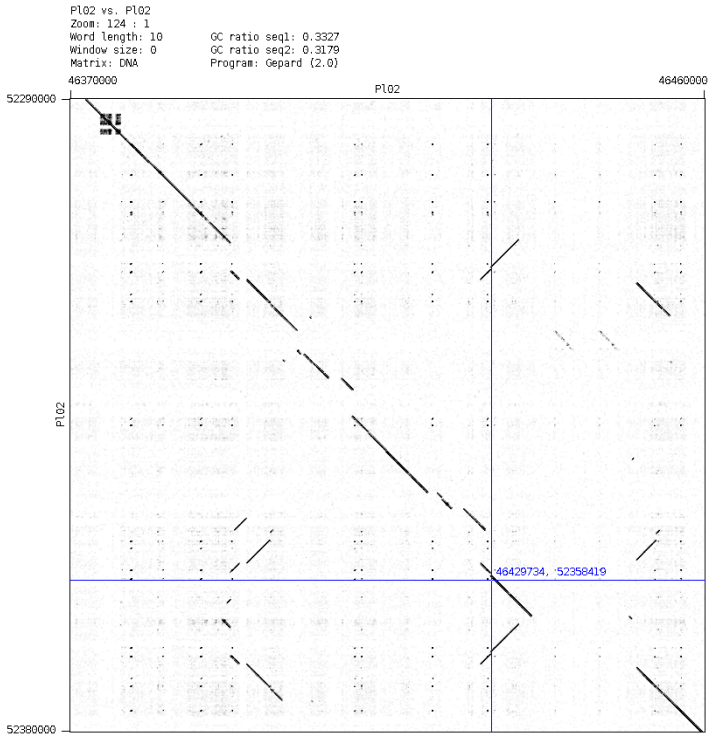


Supplementary Figure 6. Size distribution of insertions and deletions for the three accessions. Insertion events (light blue) and deletion events (dark blue) are grouped into three size categories: small (0-1000 bp), medium (1000-10000 bp), and large (greater than 10000 bp). The gene pool G25228 Wild MI (A, B, C, D, E, F), G25231 Wild MI (G, H, I, J, K, L), and G25915 Wild AI (M, N, O, P, Q, R).



Supplementary figure 7. A) Alignment between the Andean (Vertical) and the Mesoamerican (Horizontal) *P. lunatus* genomes in the region around the gene *IND* (*Pl02G0000404800*). B) Alignment of long reads from the wild Andean accession G25915 to the genome of the domesticated Andean accession G25900. The tracks display the annotations of gene models, transposable elements, large indels, sequencing depth and read alignments.

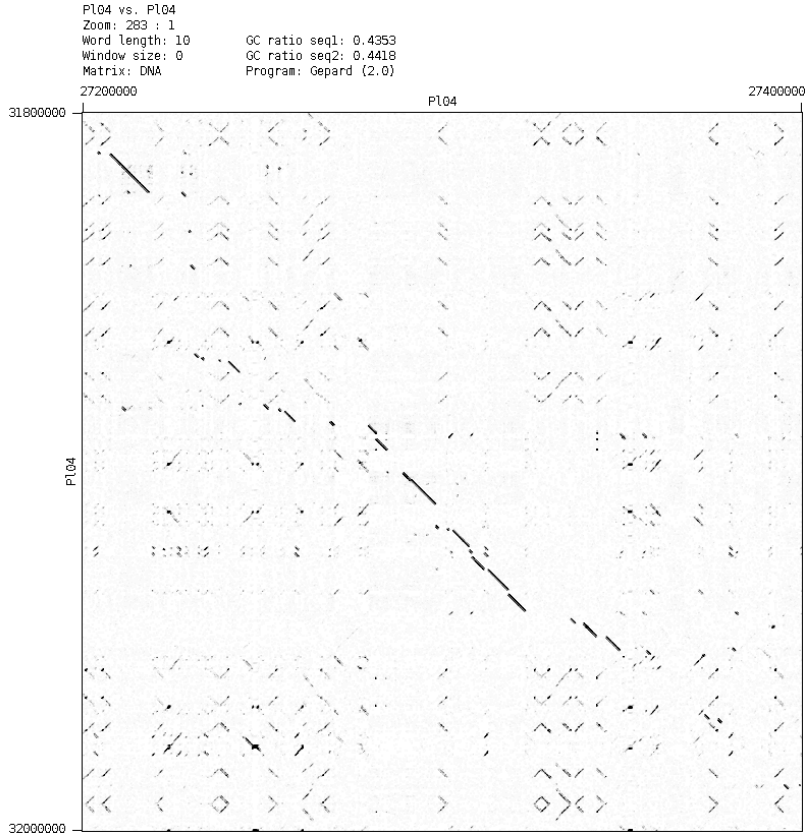
A.



B.

Supplementary figure 8. A) Alignment between the Andean (Vertical) and the Mesoamerican (Horizontal) *P. lunatus* genomes in the region around the gene GIGANTEA (*Pl04G0000200500*). B) Alignment of long reads from the wild Andean accession G25915 to the genome of the domesticated Andean accession G25900. The tracks display the annotations of gene models, transposable elements, large indels, sequencing depth and read alignments.

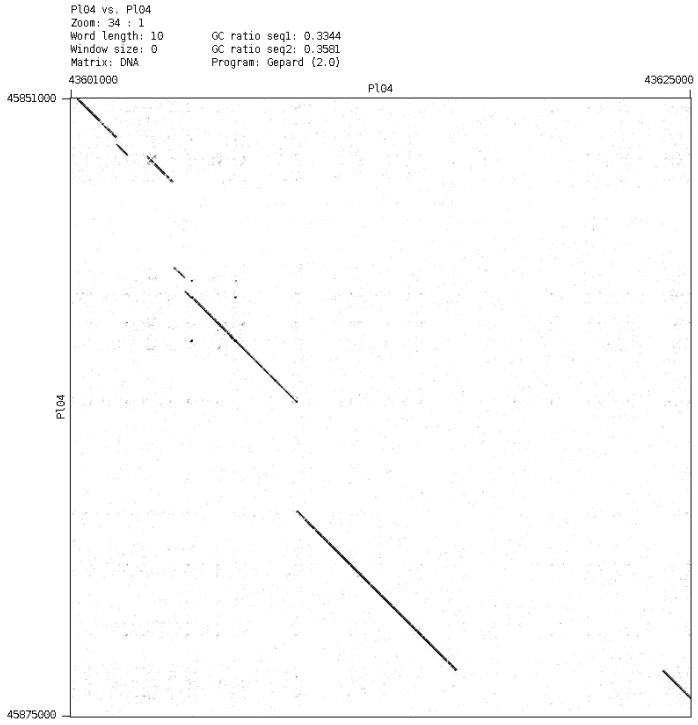
A.



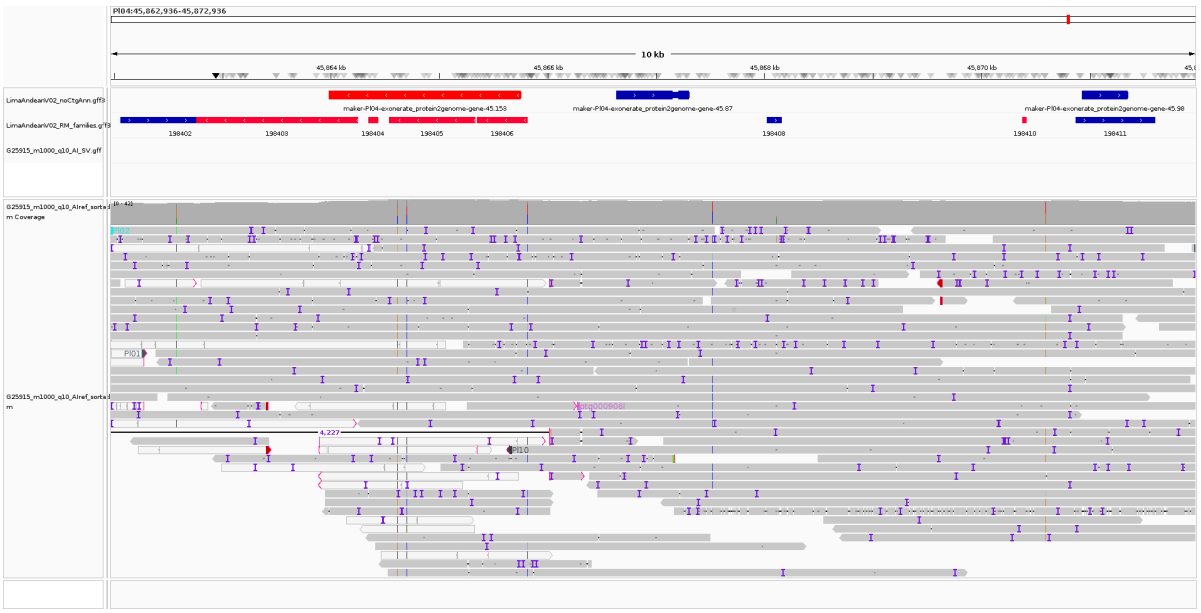
B.

Supplementary figure 9. A. Alignment between the Andean (Vertical) and the Mesoamerican (Horizontal) *P. lunatus* genomes in the region around the gene DREB1A (*Pl04G0000312600*). B. B) Alignment of long reads from the wild Andean accession G25915 to the genome of the domesticated Andean accession G25900. The tracks display the annotations of gene models, transposable elements, large indels, sequencing depth and read alignments.

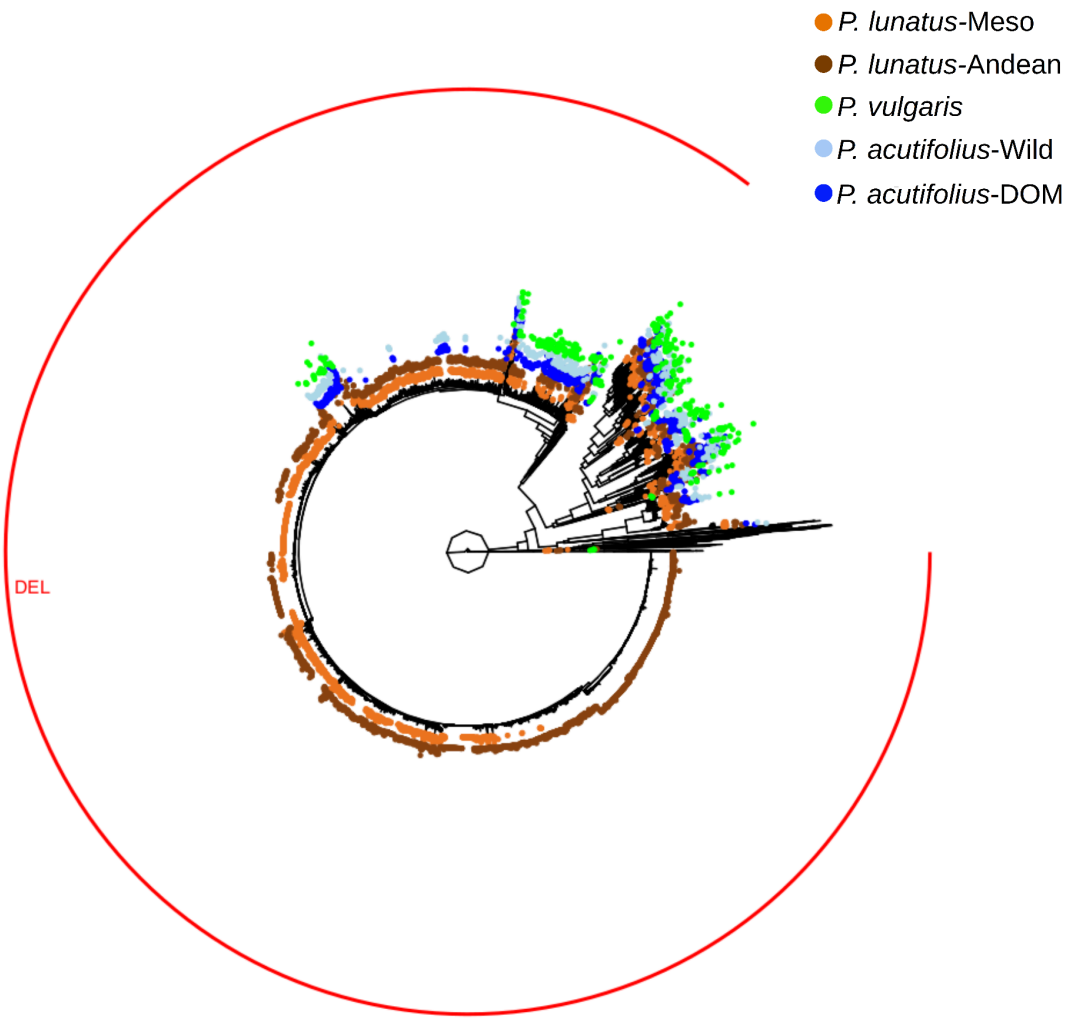
A.



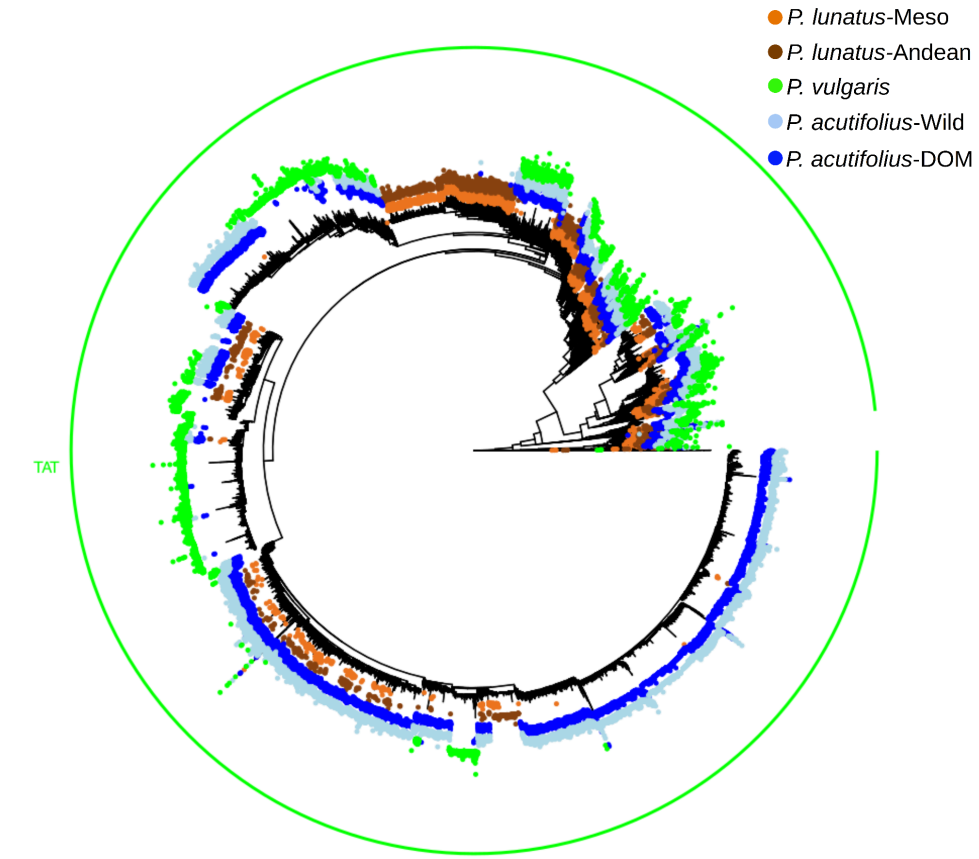
B.

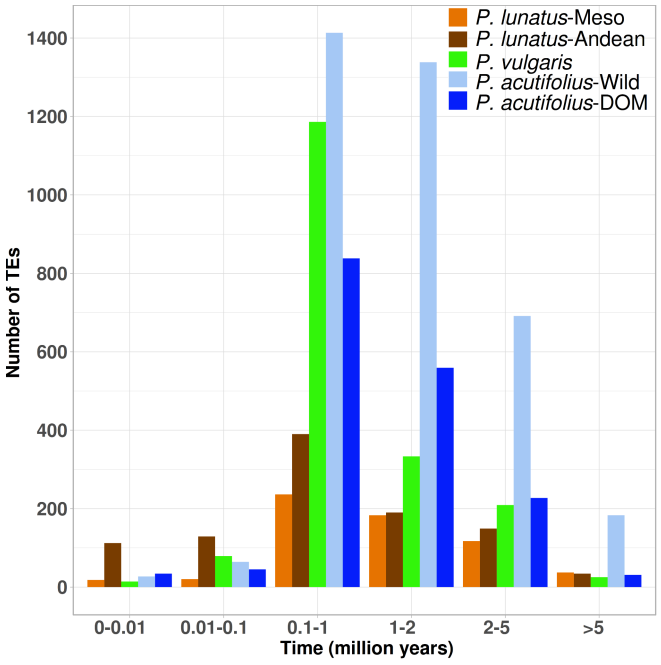


Supplementary Figure 10. Phylogenetic analysis and comparison of the *P. lunatus AI*, *P. lunatus MI, P. vulgaris, P. acutifolius* domesticated and wild LTR Gypsy/TEKAY-DEL retrotransposon sequences encoding the reverse-transcriptase (RT) domains.

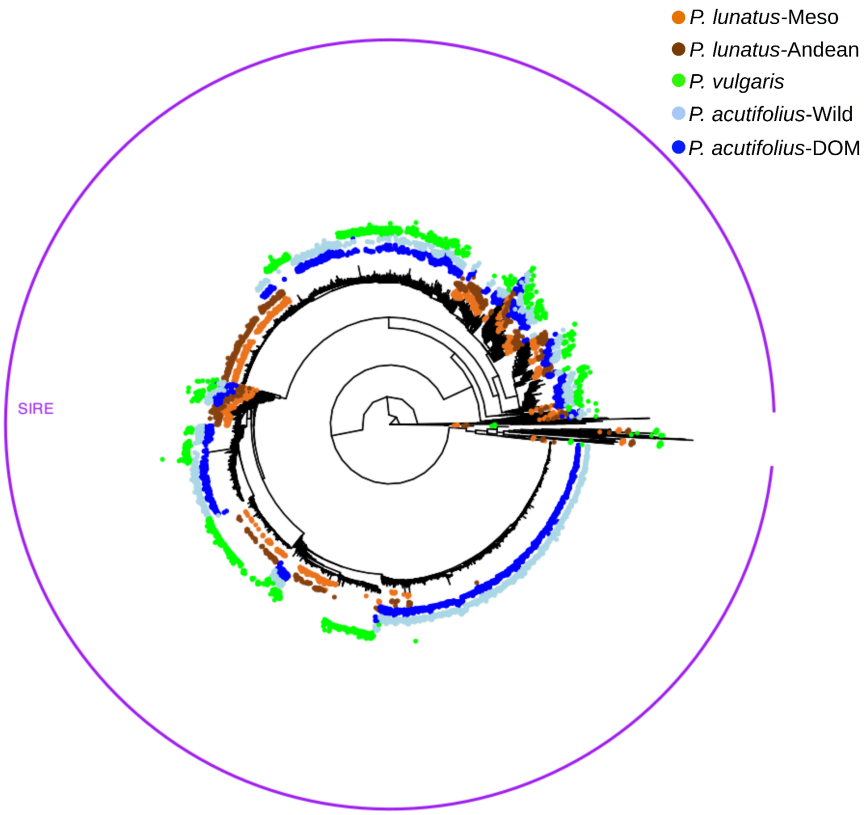


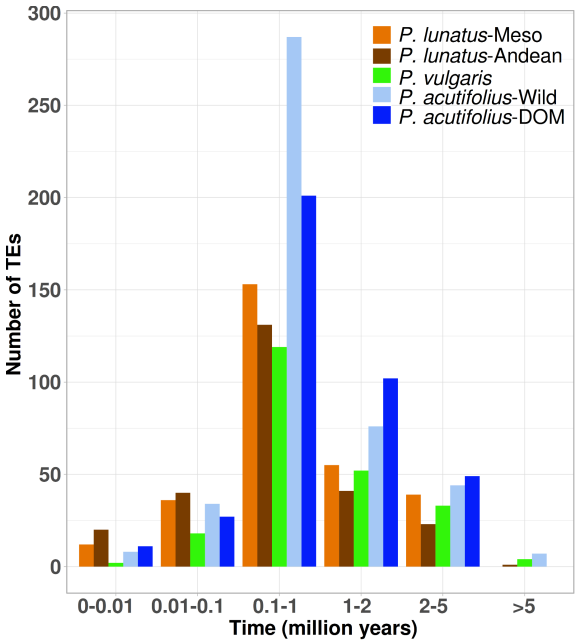
Supplementary Figure 11. Phylogenetic analysis and comparison of the *P. lunatus AI*, *P. lunatus MI, P. vulgaris, P. acutifolius* domesticated and wild LTR Gypsy/TAT retrotransposon sequences encoding the reverse-transcriptase (RT) domains. The lower figure shows a histogram of insertion times for each event.



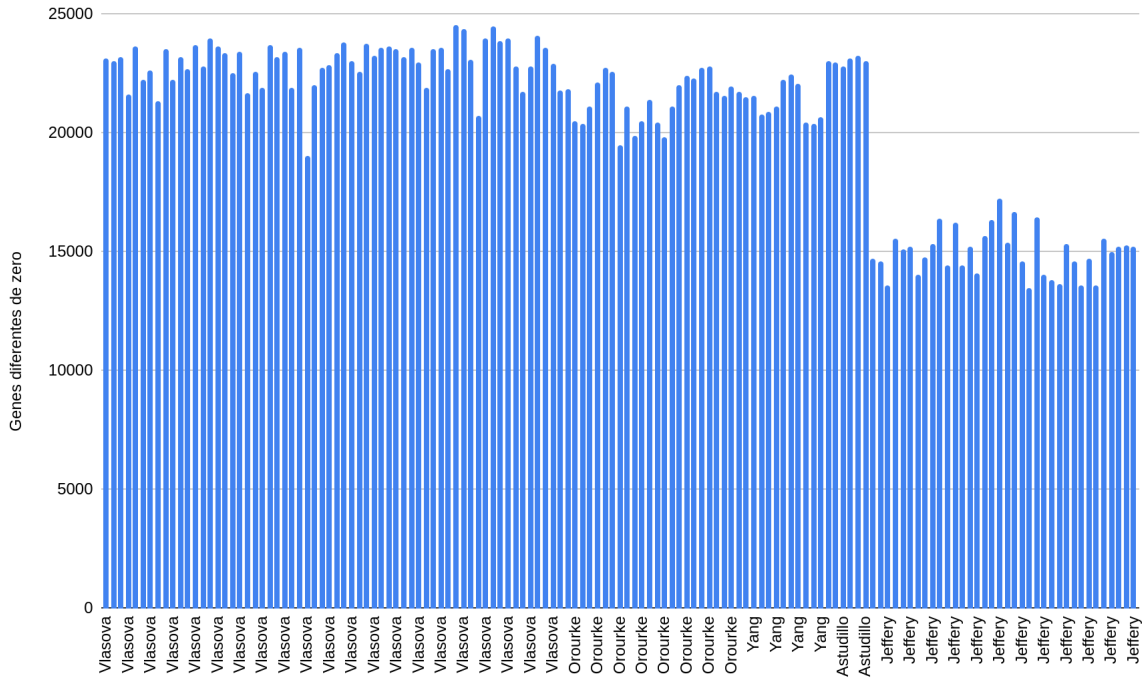


Supplementary Figure 12. Phylogenetic analysis and comparison of the *P. lunatus AI*, *P. lunatus MI, P. vulgaris, P. acutifolius* domesticated and wild LTR Copia/SIRE retrotransposon sequences encoding the reverse-transcriptase (RT) domains. The lower figure shows a histogram of insertion times for each event.





Supplementary figure 13. Number of genes with evidence of expression after reanalysis of 139 RNA-seq samples obtained in five different studies.



Supplementary Figure 14. Principal Component Analysis (PCA) of139 RNA-seq samples obtained in five different studies. Colors discriminate between the study from which each sample was taken.