nature portfolio

Corresponding author(s): Jorge Duitama

Last updated by author(s): Jan 4, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Sequencing was performed following standard protocols, including software for base calling. HiFi reads for G25900 were obtained following the Pacific Biosciences circular consensus sequencing protocol at the Georgia Genomics and Bioinformatics Core (GGBC). oxford nanopore reads were obtained following the native protocol for barcoding and sequencing in a R10 flow cell and a PromethION device at the GeneCore facility of Universidad de los Andes.

| Data analysis | Genome assembly: HiFiAsm v0.16
Short Read alignment: bowtie2 v2.4.5, HISAT2 v2.2.1
Long read alignment: Minimap2 v2.28, picard v2.27
Variant calling, genotyping, filtering and imputation of missing data: NGSEP v5.0.0, SVIM v2.0.0
Assembly quality: QUAST v5.0.2, BUSCO v5.3.2
Identification of transposable elements: Inpactor v2, CD-HIT v4.8.1, EDTA v2.2.0, RepeatMasker v4.1.2, LTRFinder v1.07, LTR_retriever v3.0.1, TRF v4.09.1, NGSEP v5.0.1
Phylogenetic reconstruction of LTR transposons: Inpactor v2, CENSOR v4.2.29, MAFFT v7.475, FastTree v2.1.11
Gene annotation: Maker v2.31.9, Trinotate v3.1.1, NCBI blast v2.10.0, HMMer v3.3.1
Orthogroups and synteny: NGSEP v5.0.1
Metabolic networks: RAVEN v2.9.0, Pathway tools v 28.0
RNA read counts: Stringtie v1.3.5
Visualization: IGV v2.17, Gepard v2.1, IToL v7
Custom scripts developed for data analysis are available with the distributions of the open source software NGSEP (http://ngsep.sf.net) and Metapenta (https://github.com/FLAGlab/MetaPeNTACore). |
|---|---|

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

DNA sequencing generated for this study and the assembly of G25900 are available at the sequence read archive (SRA) database with the bioproject accession number PRJNA596114. (Biosamples SAMN45957909, SAMN45957910, SAMN45957911 and SAMN45957912). The genome annotation and the core metabolic network are provided as supplementary materials (Supplementary files 6 and 7).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	This study did not involve research with human participants
Reporting on race, ethnicity, or other socially relevant groupings	This study did not involve research with human participants
Population characteristics	This study did not involve research with human participants
Recruitment	This study did not involve research with human participants
Ethics oversight	This study did not involve research with human participants

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	HiFi Reads for the Lima bean AI genome assembly were taken from the accession with code G25900. For structural variation between wild and domesticated Lima bean, two wild MI accessions (G25228 and G25231) and one wild AI accession (G25915) were sequenced with the Oxford nanopore technology. The genomic background of each sample was determined by a previous stuudy based on Illumina sequencing (https://doi.org/10.1038/s42003-023-05144-y). No estimations of population parameters or inferences on populations were included in this study.

Data exclusions	All sequenced samples were included in the analysis. Comparisons at the gene level were performed considering the complete set genes related to each specific classification (single copy, enzymes related to a pathway group, etc).
Replication	We did not perform replication of the DNA sequencing experiments
Randomization	No randomization was applicable in this case.
Blinding	We consider that blinding is not applicable in this case.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☐	☒ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Dual use research of concern

Policy information about [dual use research of concern](#)

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes
☒	☐ Public health
☒	☐ National security
☒	☐ Crops and/or livestock
☒	☐ Ecosystems
☒	☐ Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes
☒	☐ Demonstrate how to render a vaccine ineffective
☒	☐ Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐ Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐ Increase transmissibility of a pathogen
☒	☐ Alter the host range of a pathogen
☒	☐ Enable evasion of diagnostic/detection modalities
☒	☐ Enable the weaponization of a biological agent or toxin
☒	☐ Any other potentially harmful combination of experiments and agents

Plants

Seed stocks	The accessions sequenced in this study can be requested to the germplasm bank (https://alliancebioversityciat.org/colombia-genebank) at the International Center for tropical agriculture (CIAT) with accession numbers G25228, G25231, G25900 and G25915.
Novel plant genotypes	No new plant genotypes were generated in this study
Authentication	Not applicable