nature portfolio

Corresponding author(s): Noel Ellis
Noam Chayut
Shifeng Cheng

Last updated by author(s): Mar 12, 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<br>*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<br>*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 | No software used for data collection. |
| Data analysis | Alignment and SNP calling:<br>Alignment: BWA-MEM (v0.7.17)<br>Mark duplicate: picard (v 2.20.3-SNAPSHOT)<br>Sort Alignment file: SAMtools (v1.9)<br>SNP/Indel Calling: GATK (v4.1.2);Sentieon (v202112.01)<br>LD pruning: PLINK (v1.90b6.7)<br>SNP annotation: SnpEff (version 4.3t)<br>SV calling: Delly (v 0.8.7)<br>Phasing: Beagle (v 21Apr21.304)<br><br>Pea HapMaP construction:<br>Block identification: PLINK (v1.90b6.7)<br>Haplotype assignment: HAPPE (v0.1.4), https://github.com/fengcong3/HAPPE<br><br>CNV identification:<br>Blastn (v 2.8.1+)<br><br>k-mer databases count from raw reads and genome assemblies:<br>kmerGWAS pipeline https://github.com/voichek/kmersGWAS |

Jellyfish v.2.2.6
KMC v3.0.1

IBSpy variations and Long Range haplotypes:
The implementation of the complete pipeline is in the final version of IBSpy (v.0.4.6) in https://github.com/Uauy-Lab/IBSpy
Affinity Propagation (AP) methods and the API: https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AffinityPropagation.html
AP clusters were scored by Silhouette Coefficient (SC) score and the API: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.htm

Population structure analysis:
Phylogenetic tree: SplitsTree4 (V4_18_2 )
Tree visualisation: iTOL (v7; https://itol.embl.de/)
Structure: ADMIXTURE (v1.3.0)
Genetic diversity (π): vcftools (v0.1.13)
PCA: Python-sklearn (v0.24.2)
t-SNE: Python-sklearn (v0.24.2)
PopLDdecay (version 3.31)

GWAS study:
Calculate kinship and perform GWAS: GEMMA (v0.98.1)
Local details and LD: LDBlockShow (v1.40)
RTM-GWAS 2022.0-rc2 (19 Oct 2022)

Statistical analyses
R software suite (version 4.2)
Pearson correlation coefficient (r) was calculated using R package corr v0.4.4 and corrplot v0.92.
The world map is drawn using the maps package (v3.4.0) in R (version 4.2).

QTL and BSA analysis
R software suite (v3.6.1) package qtl (v1.5)
QTLseqr package (v0.7.5.2)

RNA-seq, Iso-seq and gene expression:
trimmomatic (version 0.38)
hisat2 (version 2.1.0)
stringtie (1.3.5)
Iso-Seq (v4.0.0)
ccs (v3.4.1)
lima (v2.9.0)
pbmm2 (v1.14.99)
gffread (0.12.7)

DNA Methylation analysis:
Trimmomatic (v0.39)
Bismark (v0.23.0)

qRT-PCR analysis:
R software suite (v4.2.0)

Orthologues and Gene Family Analysis:
HMMER (v3.1b1)
MAFFT (v7.475)
trimAl (v1.5.rev0)
ModelTest-NG (v0.1.7)
IQ-TREE (v2.1.2)
OrthoFinder (v2.5.4)
JCVI (v1.2.7)

Analysis pipeline is available at: https://github.com/ShifengCHENG-Laboratory/MendelPeaG2P

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

All whole-genome sequence and transcriptome data have been deposited in the National Genomics Data Center (NGDC) Genome Sequence Archive (GSA) under BioProject accession number PRJCA023166. The datasets are available under SRA accession IDs CRA014669, CRA023374, and CRA023375 (https://ngdc.cncb.ac.cn/gsa/). Phenotyping data were given in Supplementary tables and long-term phenotype curation available on SeedStor (https://www.seedstor.ac.uk). Some of the

data and analyses are also available in our established portal: https://mendelpea.com/dataAvailable, https://mendelpea.com/toolkits.

Information related to Agrobacterium tumefaciens (GV3101) can be found at:
https://www.weidibio.com/display.php?id=304
The primers and plasmid sequences used in this study are provided in Supplementary Table 47.

The SNP matrix is available at:
https://mendelpea.com/dataAvailable/SNP_matrix/

The haplotype matrix (including CNV matrix, k-mer-based IBS long-range haplotypes, and LD-based short-range haplotypes) is accessible at:
https://mendelpea.com/dataAvailable/Haplotype_matrix/

All phenotypic data in MendelG2P (including the seven classic Mendelian traits; see Supplementary tables) can be found at:
https://mendelpea.com/dataAvailable/Phenotypic_data/

The SNP matrix used for the Fa locus BSA analysis is available at:
https://mendelpea.com/dataAvailable/SNP_matrix_for_FaFa_BSA/

## 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 | N/A |
| Reporting on race, ethnicity, or other socially relevant groupings | N/A |
| Population characteristics | N/A |
| Recruitment | N/A |
| Ethics oversight | N/A |

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 | A total of 697 pea (Pisum sativum) accessions, maximizing genetic diversity, were selected from the JI Pisum Germplasm Collection (Supplementary Table 1; www.seedstor.ac.uk) and introduced to the Agricultural Genomics Institute at Shenzhen (AGIS), CAAS, China, in 2019. These lines encompass the seven classical Mendelian traits and exhibit extensive phenotypic variation, which we judged sufficient to address our research questions. The diversity panel was evaluated at three sites—Norwich, UK (52.62° N, 1.28° E), Shenzhen, China (22.61° N, 114.51° E), and Harbin, China (45.86° N, 126.83° E)—with four rounds of phenotyping in China. |
| Data exclusions | No data were excluded from the analyses. All 697 accessions met sequencing and phenotyping quality criteria. |
| Replication | Field trials with the 697 MendelPea accessions were conducted in a randomized, unreplicated augmented block design across four years at the three locations. In Shenzhen, both greenhouse and field plantings were carried out to capture differences in light and temperature conditions. For the bulked-segregant analysis (BSA) targeting the Fa locus, each pool consisted of 30 F2 individuals showing the same phenotype.<br>In laboratory experiments, RT-qPCR assays were performed with three technical replicates per sample and 3–5 biological replicates per accession. All transcriptome analyses included three biological replicates per sample. Similar replication strategies were employed in other experiments such as subcellular localization, VIGS, in situ hybridization, yeast two-hybrid assays, and electron microscopy, each repeated at least twice with consistent results. |
| Randomization | All field trial design included randomisation of the genotypes.In the laboratory, sample allocation to different treatments (e.g., VIGS, subcellular localization, qPCR assays, in situ hybridization, and yeast two-hybrid) was also randomized to control for confounding variables. |

Blinding

Phenotypic measurements on the Mendel Peas were taken independently by different scientists, across locations, and without knowledge of the underlying identifier of each accession. Hence all phenotypic measurements were blinded.

# 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 |

## Antibodies

Antibodies used

We employed an alkaline phosphatase (AP)-conjugated anti-digoxigenin (DIG) antibody (catalog no. 200-052-156, Jackson ImmunoResearch Inc., USA) for detecting the in situ hybridization signal. We followed the manufacturer's recommended working concentration and incubation conditions, as well as standard protocols102 for paraffin section preparation and probe hybridization. Sequences of the DIG-labeled antisense riboprobes are provided in Supplementary Table 47.

Validation

According to the manufacturer's documentation, this AP-conjugated anti-DIG antibody is validated for in situ detection of DIG-labeled probes in multiple applications. In our study, we confirmed its specificity using DIG-labeled antisense riboprobes on plant tissues, detecting strong and clear signals with minimal background (Brewer et al., 2006). Negative controls (probes omitted) showed no detectable signal, further supporting the antibody's specificity. Additional validation information can be found on the Jackson ImmunoResearch website and in their technical bulletins. Images were acquired and analyzed with a Pannoramic MIDI digital slide scanner (3DHISTECH Ltd., Hungary).

## 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 | A total of 697 accessions, maximising genetic diversity, were selected from the JI Pisum Germplasm Collection for this study (Supplementary Table 1 and www.seedstor.ac.uk). These germplasms were introduced in 2019 to the Agricultural Genomics Institute at Shenzhen (AGIS), Chinese Academy of Agricultural Sciences (CAAS), China, where they are grown and investigated annually. |
| Novel plant genotypes | We employed two sets of induced mutants for targeted locus studies :<br>TILLING mutants at the Gp locus. These lines were generated by chemical mutagenesis followed by a TILLING (Targeting Induced Local Lesions in Genomes) screening approach, resulting in specific point mutations in the Gp region.<br>Fast neutron (FN) mutants at the D locus. These lines were created via fast neutron bombardment, which typically introduces larger deletions or rearrangements in the targeted genomic region. |
| Authentication | Each accession was initially identified by the supplying germplasm resource center using morphological descriptors (e.g., seed shape, color, growth habit). For our study, we further confirmed species identity through whole-genome resequencing data, ensuring that the genetic profiles were consistent with Pisum sativum. We did not observe any evidence of contamination or off-target species admixture in the sequencing data.<br>All mutant lines were confirmed to carry the desired mutations through Sanger sequencing (or another genotyping method) before use in our experiments. None of these lines involve transgenic approaches, gene editing techniques, or other modern biotech modifications; rather, they rely on classical mutagenesis and subsequent genetic screening. |