nature research

Corresponding author(s): Judith BURSTIN

Last updated by author(s): Jun 11, 2019

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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 data were obtained using software available with sequencing machines: RTA (1.18.64 , 1.13.48, 1.17.21.3, 1.18.54) and bcl2fastq from Illumina (1.8.3 et 2.17.4).
Data analysis	Genome assembly was performed using open-source tools : Musket 1.1, SOAPdenovo2 2.04, SSPACE 2.0, MAGUS 1.0, GapCloser v1.12, Allmaps (git commit 66165f8), BWA v0.7.4, DELLY v0.0.11. Skim GBS SNP calling was done using SGSautoSN, SOAPaligner/SOAP2 v2.21. The genetic map was built using CarthaGene v1.2 (http://migale.jouy.inra.fr/?q=fr/outils_inra). Optical maps were analyzed using the software provided by BioNano Genomics (Bionano IrysView version 2.5.1 and its associated tools). Analysis of Transposable element by RepeatExplorer v2.0, TAREAN, REPET/PASTEC v2.6, RepeatScout, DANTE - Protein domain finder, web version at https://repeatexplorer-elixir.cerit-sc.cz; PATHd8 - a program for phylogenetic dating of large trees without a molecular clock, https://www2.math.su.se/PATHd8/; R version 3.4.0 (www.r-project.org) with packages Biostrings (version 2.46.0, www.bioconductor.org), ape (version 5.1, www.r-project.org), and karyoploteR (version 1.4.2, www.bioconductor.org). Gene prediction and annotation was done using bedtools v2.26.0, JCVI utility libraries, augustus v3.0.3, fgenesh v7.1.1, blast+ v2.2.29, STAR v2.4.0j, stringtie v1.2.2, trinity-GG v2.0.6, PASA v2.0.6, EVM v1.1.1, interproscan v5.25-64.0, TrapID web version http://bioinformatics.psb.ugent.be/webtools/trapid/, euGene v4.2a, ncRNA prediction and annotation was computed by FEELnc(git commit ca37a6f), tRNAScan-SE v1.3.1, rfamscan, RNAMMER v1.2, FASTX-Toolkit v0.0.13, ShortStacks v3.8.31, TargetFinder (git commit 848b2dd). Comparative genomics was done using Orthofinder v2.1.2, diamond v0.9.14, muscle v3.8.31, PAML, bioruby-alignment. Whole genome resequencing data were analysed using BWA v0.7.12, SNPsift, PICARD tools, BCFtools v1.3, vcftools v0.1.13, plink v1.90. Phylogenetic analyses was done using IQQ-tree v1.6, R v3.5.1 with packages limma v3.38.3, VCF-kit (git commit eb45ec1). Transposable element diversity was assessed using NextGenMap v0.5.0, featureCounts v 1.5.0-p3. Chloroplast sequences were reconstructed using MITObim v1.7 and analysed using GATK and raxml v8.2.10. Translocation analysis was done using SPAdes and blat. Statistical and graphical results were computed using R v3.4.1 and SAS v9.4 unless specified differently. Circular graphic was plotted using Circos v0.69-5.

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

All raw sequencing data are available under Bioprojects at EBI (PRJEB30482) and NCBI (PRJNA507685, PRJNA507688, PRJNA509681, PRJNA510273, PRJNA285605, PRJNA431567, PRJNA509279). The pea genome reference assembly, pea genome v1a, is available for download and JBrowse at https://urgi.versailles.inra.fr/Species/Pisum. It is also available at the European Nucleotide Archive (http://www.ebi.ac.uk/ena/data/view/) under Bioproject PRJEB31320. Raw phenotyping and transcriptomics data are available upon request. All mean data are available in Supplementary data.

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	No statistical methods were used to determine sample size. The number of replicates followed what is common practice in the field. Three replicates were used for seed qPCR experiments presented in Figure 5. For some other plant tissues (info given in Supplementary data S4), the number of replicate could be 2 or one. Plant phenotyping data were averaged, for simple morphological traits, on scores made on two plants and for quantitative traits, on at least 6 plants (info given in Supplementary data S7). For germination tests, 3 replicates were done. For flow-sorted chromosomes, the number of replicates varied following blind sampling within flow-cytometry peaks. The number of samples representing each chromosome is given in the Supplementary notes.
Data exclusions	No data were excluded
Replication	Experimental findings were reliably reproduced, except for one single chromosome sample that appeared off-type.
Randomization	Plants were randomly allocated in the glasshouse.
Blinding	No phenotypic analyses, where blinding is essential for reliability of results, were carried out in this study.

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
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

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

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	In order to prepare suspension of intact mitotic chromosomes, 30 seeds were germinated in a glass petri dish on moistened filter paper. Seedlings with approximately 3 cm primary roots were transferred onto a plastic tray filled with Hoagland's solution containing 1.25 mM hydroxyurea (HU) for 18 hours. Then the roots were incubated in HU-free Hoagland's solution for 4.5 h and immediately after in 10 µM amiprophos-methyl in Hoagland's solution for 2 h. All incubations were performed in the dark at 25 ± 1°C and all solutions were aerated. Finally, the seedlings were transferred to a tray filled with ice water and incubated overnight in a refrigerator. The synchronized roots were cut 1 cm from the tip and fixed in 2% (v/v) formaldehyde in Tris buffer for 30 min at 5°C. Then the roots were washed three times for 5 min in Tris buffer and meristem tips of 25 roots were cut and transferred to a polystyrene tube containing 1 ml LB01 lysis buffer and chromosomes were released mechanically by a Polytron PT 1200 homogenizer at 13,000 rpm for 18 s. The homogenate was passed through a 20 µm pore size nylon mesh and stained by DAPI at final concentration of 2 µg/ml.
Instrument	FACSAria II SORP, BD Biosciences, San Jose, CA, USA; Firmware ver. 1.6. (BD FACSAria II)
Software	FACS Diva, ver. 6.1.3.
Cell population abundance	N/A
Gating strategy	N/A

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.