nature portfolio

Corresponding author(s): Jeffrey A. Fawcett, Chengyun Li, Hideki Hirakawa, Tatsuya Ota, Yasuo Yasui

Last updated by author(s): Jun 23, 2023

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	Raw DNA sequence data for NRGene assembling were generated by sequencing using the Illumina Hiseq 2500 and the Illumina HiseqX platforms. Raw pacbio DNA sequence data were generated by sequencing using the Pacbio sequel II platform. Raw DNA sequence data for whole genome re-sequencing was generated by sequencing using the Illumina Novaseq6000 platform. RNA-seq data was generated by sequencing using the Hiseq 2500 and the Novaseq6000 platforms. Raw DNA sequence data for NGS tilling was generated by sequencing using the Illumina HiseqX platform.
Data analysis	The softwares used in this study are listed as follows: DenovoMAGIC v3.0 (https://www.nrgene.com/de-novo-magic/) CCS v4.2.0 (https://ccs.how/) hifiasm v0.7-dirty-r256 (https://github.com/chhylp123/hifiasm) Pilon v1.22 (https://github.com/broadinstitute/pilon) Ion AmpliSeq Designer v6.0 (https://AmpliSeq.com/help/startDesign.action) Torrent Suite v5.8.17 (https://github.com/iontorrent/TS) vcftools v0.1.12b (https://vcftools.github.io/) AntMap v1.0 (http://lbm.ab.a.u-tokyo.ac.jp/~iwata/antmap/) BLAST+ v2.9.0 (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/) MashMap v2.0 (https://github.com/marbl/MashMap) MUMMER v4 (https://mummer4.github.io/) LTRharvest v1.5.9 (http://genometools.org/) LTRdigest v1.5.9 (http://genometools.org/)

RepeatModeler v1.0.11 (https://www.repeatmasker.org/RepeatModeler/)
RepeatMasker v4.0.9 (https://www.repeatmasker.org/)
MAFFT v7.475 (https://mafft.cbrc.jp/alignment/software/)
CD-HIT v4.8.1 (https://sites.google.com/view/cd-hit)
HMMER v3.1b2 (http://hmmer.org/)
IQ-TREE v2.2.0 (http://www.iqtree.org/)
Salmon v1.2.1 (https://combine-lab.github.io/salmon/)
fastp v0.20.1 (https://github.com/OpenGene/fastp)
HISAT2 v2.1.0 (http://daehwankimlab.github.io/hisat2/)
Stringtie v2.1.7 (http://ccb.jhu.edu/software/stringtie/index.shtml)
BRAKER v2.1.5 (https://github.com/Gaius-Augustus/BRAKER)
DIAMOND v0.9.9 (https://github.com/bbuchfink/diamond)
minimap2 v2.24 (https://github.com/lh3/minimap2)
BUSCO v5.3.2 (https://busco.ezlab.org/)
LAI version beta3.2 (https://github.com/oushujun/LTR_retriever)
MCScanX v1.0 (https://github.com/wyp1125/MCScanX)
PAML v4.9 (http://abacus.gene.ucl.ac.uk/software/paml.html)
OrthoFinder v2.5.2 (https://github.com/davidemms/OrthoFinder)
trimAl v1.4.1 (http://trimal.cgenomics.org/)
BEAST v2.6.6 (https://www.beast2.org/)
Trimmomatic v0.32 (http://www.usadellab.org/cms/?page=trimmomatic)
BWA v0.7.17 (https://bio-bwa.sourceforge.net/)
bcftools v1.10.2 (https://samtools.github.io/bcftools/)
VarScan v2.3.9 (http://dkoboldt.github.io/varscan/)
SnpEff v4.3t (http://pcingola.github.io/SnpEff/)
Plink v1.90b (https://www.cog-genomics.org/plink/)
ADMIXTURE v1.3 (https://dalexander.github.io/admixture/)
PHYLIP v3.697 (https://evolution.genetics.washington.edu/phylip)
pixy v1.2.7 (https://github.com/ksamuk/pixy_analysis)
BEAGLE v5.1 (http://faculty.washington.edu/browning/beagle/beagle.html)
GMAP v2021-08-25 (http://research-pub.gene.com/gmap/)
Jellyfish v2.2.6 (https://github.com/gmarcais/Jellyfish)
Circos v0.69.8 (http://circos.ca/)

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

The final genome assemblies, annotations, RNA sequences, and raw genome sequence data generated in this study have all been deposited in the DNA Data Bank of Japan (DDBJ) database under BioProjects PRJDB15031 (assembly of F. esculentum PL4: BSUD01000001-BSUD01003041, assembly of F. esculentum GS1: BSUE01000001-BSUE01042256, raw reads of various F. esculentum accessions: DRR438014-DRR438137, DRR477348-477353) and PRJDB15175 (assembly of F. homotropicum: BSWB01000001-BSWB01000436, raw reads of F. homotropicum: DRR438312). The final genome assemblies and annotations are also available from the Buckwheat Genome DataBase (BGDB) (http://buckwheat.kazusa.or.jp/), as well as the initial scaffolds and contigs used to construct the final assemblies. Publicly available datasets from the following databases and websites were used in this study: NR database of NCBI (http://www.ncbi.nlm.nih.gov/RefSeq), UniProtKB (https://www.uniprot.org), PFAM (https://pfam.xfam.org), Phytozome (https://phytozome-next.jgi.doe.gov), TAIR (www.arabidopsis.org), MBKbase (https://www.mbkbase.org), CARNIVOROM (https://www.biozentrum.uni-wuerzburg.de/carnivorom). Source data are provided with this paper.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	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	One plant was used for the genome assembly of F. esculentum PL4, F. esculentum GS1, and F. homotropicum because our aim was to construct assemblies of single individuals to use as references and structural variation between individuals was not the focus of our study. For linkage mapping, 112 F2 progenies from a single parental pair were used. This was because using F2 progenies derived from a single parental pair is most straightforward for genetic linkage analysis. The number of progenies was expected to be sufficient for constructing a linkage map to be used to anchor the scaffolds to pseudomolecules. Indeed, we were able to construct a linkage map and use this for anchoring the scaffolds. For NGS-TILLING, mutation populations consisting of 5,801 plants. We expected this number to be sufficient for detecting nonsense mutations in all four genes of interest, which we indeed succeeded. For population genetics analyses, whole genome sequences from 104 plants, including 11-12 wild accessions from each geographical location, and 57 cultivated accessions. We expected these numbers to be sufficient to distinguish the wild accessions from each geographical location and the cultivated accessions. Indeed, clusters corresponding to the cultivated accessions and the wild accessions from each geographical location were identified by clustering analyses.
Data exclusions	No data were excluded from the analyses
Replication	No replication was applied for our genome sequencing experiments because each region/base was sequenced to a ~10-100x coverage. For RNA-seq, no replication was applied for the data used for gene prediction because determining the expression level of all genes was not relevant. For analysis of the s-haplotype genes, three replicates were used as the expression level was of our concern and the results were consistent across each replicate as shown in Supplementary Fig. 27. Two replicates were performed for each FISH analyses and at least 10 cells per replicate were examined, and it was confirmed that the results are consistent across all cells of both replicates (>20 cells).
Randomization	No randomization was applied in this manuscript since this is a genome sequencing study and the data were generated from a single individual.
Blinding	No blinding was applied in this manuscript since this is a genome sequencing study and the data were generated from a single individual.

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

Methods

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