

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

No software was used in the data collection.

Data analysis

Commercial software sentieon was used for SNP calling. For all other analyses, freely available bioinformatics programs were used, including: MECAT2, Trimmomatic (v0.39), Pilon (v1.22), BWA (v0.7.17), LACHESIS, RepeatMasker(v4.0.9), RepeatModeler (v1.0.11), RECON (v1.08), RepeatScout (v1.0.5), LTR_FINDER (v1.0.7), HISAT2 (v2.2.1), StringTie (v2.0), PASA(v2.0.1), MAKER(v3.00), AUGUSTUS (v3.3.1), GENESCAN, Exonerate (v2.2.0), Blast2GO, GATK (v3.8), SnpEff (v5.0e), IQ-TREE (v1.6.12), ModelFinder, fastSTRUCTURE (v1.0), PLINK (v1.9), PopLDdecay(v3.41), ggpubr (v0.1.3), SMC++69 (v1.15.4.dev18+gca077da), SweeD (v4.0.0), VCFtools (v0.1.17), EMMAX (v20120210), TreeMix (v1.13), OptM (v0.1.6), vg (v1.40.0), minimap2 (v2.23), MUMmer (v4.0.0), Assemblytics (v1.2.1), bcftools (v1.14), Hifiasm (v0.16.0), Purge Haplotigs (v1.1.2)

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](#)

Genome assemblies, raw genome and transcriptome sequencing reads have been deposited in the National Center for Biotechnology Information BioProject database under the accession nos. PRJNA800056 and PRJNA801499. Genome assemblies and annotated genes, pan-genome graphs (Giraffe indexes), and variant data (VCF format) are also freely available at the Pepper Genomics Database we have developed (<http://ted.bti.cornell.edu/pepper>). Source data are provided within this paper.

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	Not applicable
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable
Population characteristics	Not applicable
Recruitment	Not applicable
Ethics oversight	Not applicable

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	For population genomic analyses, a core collection of 500 different pepper accessions were used, which are sufficient for population analyses based on literature. These 500 accessions were selected from a total of 1296 pepper accessions, which were further selected based on their availability in the germplasm collection and the balance and sufficient numbers of accessions in different pepper species/groups
Data exclusions	Potential hybrids and feral accessions were excluded from downstream population genomic analyses
Replication	For phenotyping of fruit length, width and shape index, three fruits from each of three independent plants were measured (three technique replicates and three biological replicates). All attempts at replication were successful.
Randomization	The samples in this study were randomized
Blinding	Not applicable. For a plant genome and population study, blinding is not necessary

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	Involvement 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	Involvement 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