nature portfolio

Corresponding author(s): Brande B.H. Wulff, Simon G. Krattinger

Last updated by author(s): May 13, 2024

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 for data collection was used.
Data analysis	The software and tools used in this study are as follows: Core Hunter (v3), hifiasm (v0.16.1), QUAST (v5.0.2), Merqury (v1.3), BUSCO (v5.3.1) with embryophyta_odb10 database for Ae. tauschii and bread wheat accession CWI 86942 primary assemblies, LJA assembler (v0.2), Juicer (v1.6), 3D-DNA (v190716), Juicebox (v2.20.00), RagTag (v2.1.0), MashMap (v3.0.6), QUAST (v5.0.2), BUSCO (v5.3.1), Trimmomatic (v0.38) for Illumina reads, Trimmomatic (v0.40) for RNAseq reads, STAR (v2.7.10b), samtools (v1.8) and (v1.10.0), Braker (v3.0.3), Busco (v5.4.7) with poales_odb10.2019-11-20 database for three Ae. tauschii genome annotations, BLAST+ (v2.9.0-2) and BLAST+(v2.12.0), AGAT (v1.2.1), R (v4.2.0), RStudio (v1db809b8, 2022-05-16) InterProScan (v5.64-96.0), RepeatMasker (v4.1.2-p1) for annotation of Ae. tauschii genomes, InterProScan (v5.55-88.09) for bread wheat accession CWI 86942, liftoff (v1.6.1) and (v1.6.3), gffread (v0.11.7), DIAMOND (v2.1.8), BWA mem (v0.7.17), Bcftools mpileup (v1.9), vcftools (v0.1.16), vcfkit (v0.1.6), pbsv (v2.9.0). Other software utilized to analyse the data are: Python (v3.8), SciPy library (v1.8.0), seaborn Python library (v0.11.2), pbmm2 (v1.10.0), RunAssociation_GLM.py (https://github.com/wheatgenetics/owwc/tree/master/kGWAS), BLAST+ (v2.12.0), MEGA (v11), MapDisto (v2.0), MapChart (v2.32), siFi21-1.2.3-0008, Jellyfish (v 2.3.0), comm bash command, KMC (v3.1.2), IBSpy (v0.4.6), Persephone® Web 0.82 genome browser, QGIS (v3.32.3), OrthoFinder (v2.5.4) R packages used in this study are as follows: ggplot2 (v3.4.2), karyoploteR (v1.20.3), LEA (v3.10.2) package Custom pipelines or scripts generated and used in this study:

Custom script for producing k-mer count matrices for large GWAS panels (https://github.com/githubcbrc/KGWASMatrix)
Custom scripts for missing link finder pipeline and haplotype analysis are available at GitHub (https://github.com/emilecg/wheat_evolution).

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 sequencing data and genome assemblies generated in this study were submitted to NCBI under Bioproject number PRJNA956839, including the raw Illumina reads for 350 Aegilops tauschii accessions, the raw PacBio reads, the Hi-C data, the raw RNAseq reads from five tissues of Ae. tauschii accessions TA10171, TA1675 and TA2576, the raw Illumina reads for 59 wheat landraces and the raw PacBio reads and Omni-C data of the wheat landrace CWI 86942.
The genome assemblies of the 46 Ae. tauschii accessions, the assemblies and annotations for CWI 86942, TA10171, TA1675 and TA2576, the variant call (SNP) file, the k-mer matrix for 920 Ae. tauschii accessions, the phylogenetic tree for 493 non-redundant Ae. tauschii accessions, the structural variant call (SV) files, the IBSpy variation tables, the predictions of the subpopulations contributing to the 17 hexaploid wheat assemblies, an excel file containing the RagTag scaffold output agp files and the dot-plots produced by MashMap used to validate the RagTag scaffolding are available at DRYAD (https://doi.org/10.5061/dryad.vmcvdnd0d; https://doi.org/10.5061/dryad.wm37pvmvd; https://doi.org/10.5061/dryad.wpzgmsbvm; https://doi.org/10.5061/dryad.p5hqbzkvx).
The Lr39 genomic sequence was deposited in NCBI Genbank under accession number OR567850.
The TA10171 (L1), TA1675 (L2) and TA2576 (L3) genomes are available for online BLAST, Jbrowse visualisation and synteny analysis with the currently available Triticinae genomes at (https://wheat.pw.usda.gov/GG3pangenome/wheat/D/taus_home.php).
The Ensembl nrTEplants repetitive element database (June 2020) was used for repeat content prediction. Viridiplantae protein models from OrthoDB v.11 were used to predict de novo gene models for the annotated Ae. tauschii genomes. The predicted translated proteins were annotated using the following databases: FunFam, SFLD, PANTHER, Gene3D, PRINTS, Coils, SUPERFAMILY, SMART, CDD, PIRSR, ProSitePatterns, AntiFam, Pfam, MobiDBLite, PIRSF, NCBIfam. We downloaded sequencing data for 306 accessions from NCBI BioProject number PRJNA685125, 275 accessions from NCBI BioProject number PRJNA705859 and 24 accessions from the China National Center for Bioinformation - National Genomics Data Center under accession number PRJCA005979.

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	No human research participants
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	The 59 bread wheat landraces were sequenced to be representative of the gradient of Lineage 3 introgression dected with the missing link finder pipeline. 46 Ae. tauschii accessions, including representative accessions for each subpopulation and accessions carrying traits of interest, and one bread wheat landrace were selected for genome assemblies. A total of 920 Ae. tauschii accessions were used for the population genomics analyses, sufficient to span the geographical distribution of the species and represent all the different subpopulations with an adequate depth. A total of five different plant tissues per three Ae. tauschii accessions were used to extract RNA for RNA-Seq. A bi-parental mapping population of 123 F2 progenies were generated by crossing leaf rust resistant and susceptible Ae. tauschii accessions. The sizes of the mapping population was based on literature and based on the calculated recombination frequency

Data exclusions	920 out of 955 accessions for which sequencing data was available were included in the population genomic analyses. Accessions were excluded due to low sequencing coverage (less than 5-fold) or duplicate accessions in the different datasets.
Replication	For the RNAseq data, 45 tissue samples were collected: From each of the three accessions, three biological replicates were taken from each of: young leaf, root, stem, flag leaf and inflorescence; all replicates were successful and none were discarded for the analysis. For VIGS, five biological replicates were used to test each silencing probe and the experiment was repeated three times, showing every time the same result. For rust phenotyping Ae. tauschii bi-parental mapping population, at least 15 seedlings of F2:3 families were screened to access the homozygous resistant, homozygous susceptible and segregating lines.
Randomization	Viral and rust inoculated plants were allocated randomly among groups.
Blinding	Blinding was performed when rust phenotyping plants (i.e., the genotype of the plant was not known when phenotyping the mapping populations)

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

Plants

Seed stocks	Seed of 228 non-redundant accessions were obtained from the Open Wild Wheat Consortium Ae. tauschii Diversity Panel collection deposited at the Germplasm Resource Unit (GRU) of the John Innes Centre; 48 accessions from the Cereal Crop Wild Relatives (Triticeae) collection of the GRU; 19 accessions from the DFW Wheat Academic Toolkit collection of the GRU that have been used as synthetic hexaploid wheat D-genome donors; 223 accessions from the Wheat Genetics Resource Center (WGRC) of Kansas State University; 34 accessions from the Plant Gene Resources of Canada (PGRC); 84 accessions collected from Tajikistan and donated by the Institute of Botany, Plant Physiology and Genetics of the Tajikistan National Academy of Sciences; 20 accessions donated by the Azerbaijan National Academy of Sciences; and 37 accessions collected from Pakistan and donated by Quaid-i-Azam University. Accession P-99.95-1.1 was obtained from the Deposited Published Research Material collection of the GRU.
Novel plant genotypes	N/A
Authentication	N/A 57 bread wheat landraces were received from International Maize and Wheat Improvement Center (CIMMYT) and three bread wheat landraces were obtained from the International Center for Agricultural Research in the Dry Areas (ICARDA).