nature portfolio

Corresponding author(s):	Jingquan yu; Mingfang Zhang and Yanhong Zhou
Last updated by author(s):	Oct 16, 2022

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 <u>Editorial Policies</u> and the <u>Editorial Policy Checklist</u>.

_				
U 1		-	ıst	100
_	_		_	11 \

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.
\boxtimes	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. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
\boxtimes	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
\boxtimes	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
\boxtimes	Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), 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

Aglient Masshunter Quantitative for QQQ (v B.09.00/Build 9.0.647.0) was used for ACSOs measurements.

Data analysis

CANU (v1.9); NextPolish (v1.2.4); Pilon (v1.2.2); Purge_dups (v1.2.3); Bionano Solve (v3.5.1); ALLHiC (v0.9.8); BUSCO (v3.0.2); CEGMA (v2.5); BWA (v0.7.8); LTR_retriever (v1.0.7); Mequery (v1.3); Chromonomer (V1.07); RepeatMasker (v4.07); RepeatModeler (v1.05); TRF (v4.09); Augustus (v3.2.3); Geneid (v1.4); Genescan (v1.0); GlimmerHMM (v3.0.2); SNAP (v2013.11.29); blast (v2.2.26); Tophat (v2.0.13); Cufflinks (v2.1.1); Trinity (v2.1.1); PASA (v2.2.0); EVidenceModeler (v1.1.1); InterProScan (v4.7); InterPro (v29.0); tRNAscan-SE (v1.4); INFERNAL (v1.1.2); FASTKs (https://github.com/mrmckain/FASTKs); MCLUST (https://sites.stat.washington.edu/mclust/); OrthoMCL (v1.4); MUSCLE (v3.8.31); RaxML (v8.2.12); CAFÉ (v4.2); TreeBest (v1.9.2); DupGen_finder (https://github.com/qiao-xin/DupGen_finder); HMMER (v3.3.1); samtools (v0.1.19); ANNOVAR (v20191024, https://annovar.openbioinformatics.org/en/latest/user-guide/download/); SNPhylo (v20180901); admixture (v1.23); VCFtools (v0.1.12b); PopLDdecay (v3.41); XP-CLR (v1.0); Python(v3.7.0); Perl (v5.30.0);

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 raw sequencing data generated in this study have been deposited to CNSA (https://db.cngb.org/cnsa/) under accession number "CNP0002276". The expression data of A. sativum was downloaded from NCBI GEO database under accession number "GEO:GSE145455".

Human research participants

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

Reporting on sex and gender	Not suitable for our study.
Population characteristics	Not suitable for our study
Recruitment	Not suitable for our study.
Ethics oversight	Not suitable for our study.

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 tha	at is the best fit for your research. I	lf you	u are not sure, read the appropriate sections before making your selection.
X Life sciences	Behavioural & social sciences		Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see <u>nature.com/documents/nr-reporting-summary-flat.pdf</u>

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We re-sequenced 135 diverse accessions, including A. fistulosum, shallots, Chinese red onions and one Allium altaicum. Ninety -six affirmed A. fistulosum accessions collected worldwide, including China, Japan, European, American and Central Asia countries, which covered the main bunching/welsh onion producing areas.
Data exclusions	A wild Allium accession was excluded because it's low illumina short-reads mapping rate (28.16%).
Replication	There were three or four biological replications for RNA sequencing and all the replications were successful and used in this study. And there are 3 biological replicates when quantified the isoalliin contents.
Randomization	For quantifying ACSOs contents, three samples was randomly selected in a single accession.
Blinding	Data collection and analysis were performed in a blinded manner.

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.

ומנמוכ	ייין דער איני
701010	7
5	5
61111	racortico

⋜	
_	

IVIa	teriais & experimental systems	IVIe	tnods
n/a	Involved in the study	n/a	Involved in the study
\boxtimes	Antibodies	\boxtimes	ChIP-seq
\boxtimes	Eukaryotic cell lines	\boxtimes	Flow cytometry
\boxtimes	Palaeontology and archaeology	\boxtimes	MRI-based neuroimaging
\boxtimes	Animals and other organisms		
\boxtimes	Clinical data		
\boxtimes	Dual use research of concern		