nature research

Corresponding author(s): Asaph Aharoni

Last updated by author(s): Jul 27, 2020

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	All presented data have been acquired using existing and routinely used software. Sequencing and data processing were described in detail in Tzfadia et al. (2018). The software used included: Cutadapt v1.16.6 (Martin 2011; https://cutadapt.readthedocs.io/en/stable/index.html) - read trimming, STAR v2.4.2a (Dobin and Gingeras 2015; https://github.com/alexdobin/STAR/releases) - read mapping, HTSeq-count v0.6.0 (Anders et al. 2014; https://www-huber.embl.de/users/anders/ HTSeq/doc/overview.html) - read count. Metabolomic data have been acquired using and analyzed: MassLynx software v4.1 (Waters Inc.; https://www.waters.com) - exact mass calculation and elemental composition, Databridge v4.1 (Waters Inc.; https://www.waters.com) - file conversion, R package XCMS v3.6.1 (Smith et al. 2006; https://www.bioconductor.org/packages/release/bioc/html/xcms.html) - peak picking and alignment processing, R package CAMERA v1.40.0 (Kuhl et al. 2012 https://www.bioconductor.org/packages/2.4/bioc/html/CAMERA.html) - clustering of masses, Weizmass v1 (Shahaf et al. 2016; https://github.com/AharoniLab/MatchWeiz) - annotation of metabolites.
Data analysis	Software packages used for data analysis included: R v3.6 (R Development Core Team, 2017; https://www.r-project.org/), Bioconductor v3.9 (Gentleman et al. 2004; https://www.bioconductor.org/), R package sva v3.32.1 (Leek et al. 2019; https://bioconductor.org/packages/release/bioc/html/sva.html) - batch correction, R package edgeR v3.26.8 (Robinson et al. 2010; https://bioconductor.org/packages/release/bioc/html/edgeR.html) - RNAseq data normalization and differential gene expression analysis, R package ecp v3.0.0 (James and Matteson 2013; https://cran.r-project.org/web/packages/ecp/index.html) - multiple change-point analysis for introgression mapping, R package limma v3.40.5 (Ritchie et al. 2015; https://bioconductor.org/packages/release/bioc/html/limma.html) - differential gene expression and metabolite accumulation, R package igraph v1.2.4.1 (Csardi et al. 2006; https://igraph.org/r/) - network analysis

R package glmnet v3.0-1 (Friedman et al. 2010; https://cran.r-project.org/web/packages/glmnet/index.html) - regularized regression analysis
R package cluster v2.0.5 (Maechler et al. 2014; https://cran.r-project.org/web/packages/cluster/index.html) - GAP statistics
R package weights v0.85 (Pasek et al. 2018; https://cran.r-project.org/web/packages/weights/index.html) - weighted correlation analysis

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

RNAseq data deposited in Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo) with the GEO accession number GSE151451. Metabolic data available in e!DAL repository (https://www.ebi.ac.uk/metabolights/) with the study ID code XXX (data in submission). Custom code is accessible on the GitHub repository: https://github.com/NAMlab/kILBIL. All results of the QTL analysis can be browsed and visualized interactively using a dedicated online data browser under: https://szymanskilab.shinyapps.io/kilbil/ . Source data for figures presented in the study are provided as Supplemental Data 3. Any additional information will be available upon request from corresponding authors of 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	Statistical based sample size calculation was not applicable to our study. In the QTL experiments sample size was determined by the structure of the population and its size, previous observations concerning the effect size of phenotypic and metabolic traits in the same (Ofner et al. 2016 Plant J. 87, 151–160; Eshed & Zamir 1995 Genetics 141, 1147–1162) and similar tomato pupulations (Ballester et al. 2016 Front. Plant Sci. 7, 1428; Schauer et al. 2006 Nature Biotechnology, 24(4), 447–454; Brog et al. 2018 Plant Journal, 391–403) indicated the sample size sufficient for functional characterization of tomato genes. Sample sizes for VIGS experiments and enzymatic assays were based on previous published results on the same genotypes (Cárdenas, et al. 2016 Nat. Commun. 7, 10654; Cárdenas et al. 2019 Nat. Commun. 10, 5169.). Details on technical/biological replicates used in various experiments presented in the study are provided in Methods section as well as in Main and Supplemental Figures and Supplemental Notes provided with the manuscript.
Data exclusions	No data were excluded from the analysis.
Replication	Reproducibility of the QTL analysis is based on the independent introgression lines that share specific introgressed genomic regions has been shown for individual cases, including independent experiments for the steroidal alkaloid genes (GAME31 and GAME5), for which multiple independent VIGS experiments as well as enzymatic assays (GAME5) have been performed. Results of the Botrytis susceptibility assay have been successfully confirmed in an independent experiment for selected lines exhibiting most contrasting phenotype. B. cinerea resistance assay for the ACD2, ACO5 and 4CL-Like-silenced plants included independent biological experiments as well as succesfull replication of the experiment. Details on technical/biological replicates used in various experiments presented in the study are provided in Methods section as well as in Main and Supplemental Figures and Supplemental Notes provided with the manuscript.
Randomization	Plants were grown in a randomized design. Sample analysis order has been further independently randomized for the RNAseq analysis and metabolic profiling.
Blinding	For the fruit harvest, RNAseq and metabolic profiling, and B. cinerea susceptibility assays, samples were labeled with IDs denoting their randomized run order. Sample to genotype-matching was recovered using ID to genotype mapping files on the data analysis stage. For the remaining experiments blinding was not relevant.

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