nature research

Corresponding author(s): Christina D. Smolke

Last updated by author(s): Jul 16, 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	Agilent MassHunter Workstation LC/MS Data Acquisition for 6400 Series Triple Quadrupole software (ver. B.08.02) was used for collection of LC-MS/MS data. Agilent MassHunter Workstation Optimizer for 6400 Series Triple Quadrupole software (ver. B.08.02) was used for development of multiple reaction monitoring parameters for chemical standards. Zen Pro software (Blue edition) was used for acquisition of microscopy data. NCBI BLAST online software (ver. 2.8.1) was used for alignment of DNA and amino acid sequences. RaptorX online software (ver. 2018) was used for construction of template-based homology models of protein structures.
Data analysis	Agilent MassHunter Workstation Qualitative Analysis Navigator software (ver. B.08.00) was used for analysis of all LC-MS/MS data. Microsoft Excel 2013 and Graphpad Prism 7 were used for analysis and statistical evaluation of all quantitative data, and for preparation of figures. NIH ImageJ software (ver. 1.52) was used for analysis of microscopy data, along with the "Diffraction PSF 3D" (ver. 6 June 2005) and "Parallel Spectral Deconvolution" (ver. 1.9) plugins. PyMOL software (ver. 2.2.3) was used for visualization and analysis of enzyme structures. De novo transcriptome assembly was performed using the publicly available Trinity software package (ver. 2.7.0) and the following plugins: SRA Toolkit (ver. 2.9.1), FastQC (ver. 0.11.7), BBTools/BBDuk.sh (ver. 38.16), Trinotate (ver. 3.1.1), HMMER (ver. 3.2.1). Inkscape (ver. 0.91) was used for preparation of figure panels. Small molecule geometry optimization was performed using the Gaussian software package (ver. 16). Molecular docking simulations were conducted using the Schrodinger Maestro software suite (ver. 12.0) and the Glide XP plugin (ver. 8.3). Phylogenetic sequence analyses were performed using ClustalX2 (ver. 2.1) and visualized using FigTree (ver. 1.4.3). The custom R script used for coexpression analysis of plant RNA sequencing data has been deposited in, and is available from, a public GitHub repository at github.com/smolkelab/Oxidoreductase_coexpression_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

Data supporting the findings of this work are available within the paper and its Supplementary Information files. The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request. This article includes raw source data files associated with Figs. 1-4 and Extended Data Figs. 3, 4, 6 (Supplementary Figure 1), 9, and 10. Novel genetic sequences identified and characterized in this study are available via the following accession codes from public databases. 1000Plants (1KP) database: scaffold-AIOU-2012986-Brugmansia_sanguinea (BsUGT); scaffold-JNVS-2051323-Datura_metel (DmUGT). Medicinal Plant RNAseq database: medp_datin_20101112|6354 (DiHDH); medp_datst_20101112|10433 (DsHDH). MSU Medicinal Plant Genomics Resource: full amino acid sequences and database accession numbers (IDs) for all tested HDH candidates are provided in Supplementary Table 1. Accession numbers for previously reported gene and protein sequences in the GenBank/UniProt databases are provided in Supplementary Table 2. Protein crystal structures used for homology modeling are available from the RCSB protein data bank: Arabidopsis thaliana salicylate UDP-glucosyltransferase UGT74F2 with bound UDP, PDB: 5V2K; Populus tremuloides sinapyl alcohol dehydrogenase with bound NADPH, PDB: 1YQD. The custom R script used for identification of HDH candidates via coexpression analysis of RNA sequencing data is available from the Smolke Laboratory GitHub: github.com/smolkelab/Oxidoreductase_coexpression_analysis.

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	All presented titer data represent measurements from sample sizes of n = 3 biological replicates, where biological replicates (as defined in 'Replication' below) represent independently grown microbial cultures. Since (i) engineering was performed at the cellular level, (ii) metabolite titers are a bulk measure of a cellular population, and (iii) each assayed microbial culture represents a large population of individual cells, n = 3 biological replicates were chosen as they are sufficient for reliable measurement of changes in metabolite production at the population level and for statistical analyses, based on previous metabolic engineering papers.
Data exclusions	No data were excluded from the analyses.
Replication	All measurements of metabolite titers were performed on n = 3 biologically independent samples (biological triplicates). All replicates performed in this study were biological replicates, rather than technical replicates, and represent independent data points. For example, replicate microbial cultures were grown in separate containers/wells and assayed independently from one another. Fluorescence microscopy and Western blot analyses shown in the paper are representative of two or three independent experiments; all attempts at replication of results presented in this study were successful.
Randomization	Randomization of individual cells was not relevant to this study, as the biological subjects for assays were bulk cultures of microbial cells. Microbial strains used for all experiments were inoculated from randomly selected colonies on agar plates. Representative fields of view were selected randomly for fluorescence microscopy.
Blinding	Complete blinding was not feasible for this study, as samples were prepared and measured by the same researcher. However, all samples were labeled with numeric codes, rather than names or descriptions, throughout the sample preparation and data collection work-flow to simulate blinding.

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

Antibodies

Antibodies used: Rabbit IgGκ anti-HA HRP-conjugated antibody (Absolute Antibody, 16.43/Ab00828-23.0)

Validation: As stated on the datasheet for this product on the manufacturer's website, the rabbit IgGκ anti-HA HRP-conjugated antibody was validated for use in Western blot (as well as other applications). A rat IgG2a version of this antibody clone (16.43), which shares the same variable domain sequences as the rabbit IgGκ chimeric antibody used here, has been used in a previous study for immunohistochemical staining of rat YB2/0 hybridoma cells (see manufacturer's website for reference).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s): Wild-type Saccharomyces cerevisiae strain CEN.PK2-1D was obtained from EuroSCARF (30000B).

Authentication: The commercial cell line used in this study (i.e., Saccharomyces cerevisiae CEN.PK2-1D) was not independently authenticated by the authors of this study. Engineered yeast strains were authenticated by DNA sequencing of relevant genomic modifications, as described in the Methods section of the study.

Mycoplasma contamination: Not applicable.

Commonly misidentified lines (See ICLAC register): Not applicable.