PNAS
www.pnas.org

Supplemental Material for

A massively parallel barcoded sequencing pipeline to generate the first ORFeome and interactome map for rice

Shayne D. Wierbowski[1,2,*], Tommy V. Vo[2,*,α], Pascal Falter-Braun[3,4], Timothy O. Jobe[5], Lars H. Kruse[6], Xiaomu Wei[1], Jin Liang[2], Michael J. Meyer[1,2], Nurten Akturk[2], Christen A. Rivera-Erick[2], Nicolas A. Cordero[2], Mauricio I. Paramo[2,7], Elnur E. Shayhidin[2], Marta Bertolotti[2], Nathaniel D. Tippens[1,2], Kazi Akther[8], Rita Sharma[9], Yuichi Katayose[10], Kourosh Salehi-Ashtiani[11,12,13], Tong Hao[12,13], Pamela C. Ronald[14,15], Joseph R. Ecker[16,17], Peter A. Schweitzer[18], Shoshi Kikuchi[19], Hiroshi Mizuno[20], David E. Hill[12,13], Marc Vidal[12,13], Gaurav D. Moghe[6], Susan R. McCouch[8,†], Haiyuan Yu[1,2,†]

1. Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14853, USA
2. Weill Institute for Cell and Molecular Biology, Cornell University, Ithaca, NY 14853, USA
3. Institute of Network Biology (INET), Helmholtz Zentrum München, German Research Center for Environmental Health, Munich, Germany
4. Microbe-Host Interactions, Faculty of Biology, Ludwig-Maximilians-Universität (LMU) München, Munich, Germany
5. Botanical Institute, Cluster of Excellence on Plant Sciences (CEPLAS), University of Cologne, Zülpicher Str. 47b, 50674 Cologne, Germany
6. Plant Biology Section, School of Integrative Plant Sciences, Cornell University, Ithaca, NY 14853 USA
7. Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA
8. Section of Plant Breeding and Genetics, School of Integrated Plant Sciences, Cornell University, Ithaca, NY, 14853-1901, USA
9. School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India
10. Advanced Genomics Breeding Section, Institute of Crop Science, NARO, 1-2 Ohwashi, Tsukuba, Ibaraki, 305-8634, Japan
11. Laboratory of Algal, Systems, and Synthetic Biology, Division of Science and Math, and Center for Genomics and Systems Biology, New York University Abu Dhabi, P.O. Box 129188, Abu Dhabi, UAE.
12. Center for Cancer Systems Biology (CCSB), Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA
13. Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, Massachusetts 02115, USA
14. Department of Plant Pathology and the Genome Center, University of California, Davis, Davis, California 95616
15. Innovative Genomics Institute, Berkeley, California 94704
16. Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, CA 92037
17. Plant Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037
18. BRC Genomics Facility, Institute of Biotechnology, Cornell University, Ithaca, NY 14853, USA
19. Plant Genome Research Unit, Division of Genome and Biodiversity Research, Agrogenomics Research Center, National Institute of Agrobiological Sciences, Tsukuba, Japan
20. Molecular Biology, National Institute of Agrobiological Resources, Tsukuba 305, Japan

† Correspondence should be addressed to H.Y. (haiyuan.yu@cornell.edu) and S.M. (srm4@cornell.edu)

* These authors contributed equally to this work

α Present address: Laboratory of Biochemistry and Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

www.pnas.org/cgi/doi/10.1073/pnas.1918068117

Classification

BIOLOGICAL SCIENCES: Plant Biology

Keywords

Protein-protein interaction, rice, ORFeome, next-generation sequencing, networks

This PDF file includes:

Supplementary Materials
Supplementary Dataset Headings
Supplementary Figures S1 to S8

Supplementary Methods

PLATE-Seq

Plasmid(s) from individual wells of 96-well plates were amplified by PCR using a plasmid-specific forward primer and position-specific reverse primer, consisting of a position-specific barcode and TruSeq 3' sequencing adapter. The reverse primer for Gateway entry clones was comprised of the following: TruSeq 3' adapter (5' GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT 3'), two random bases (5' NN 3'), seven nucleotide long position-specific barcode, and entry-clone specific M13G reverse (5' CAGAGATTTTGAGACAC 3'). The reverse primer for Gateway AD clones was comprised of the following: TruSeq 3' adapter (same as above), three random bases (5' NNN 3'), seven nucleotide long position-specific barcode, barcode to denote the plasmid as an AD-construct (5' CACA 3'), and AD-clone specific reverse (5' CAGAGATTTTGAGACAC 3'). The reverse primer for Gateway DB clones was comprised of the following: TruSeq 3' adapter (same as above), three random bases (5' NNN 3'), seven nucleotide long position-specific barcode, barcode to denote the plasmid as a DB-construct (5' CGTC 3'), and DB-clone specific reverse (5' CAGAGATTTTGAGACAC 3'). All primers can be found in **Supplementary Dataset 1**.

Tn5 transposase was purified as described previously[1]. Double-stranded DNA to load into Tn5 enzyme was generated by annealing two oligos: /5Phos/CTGTCTCTTATACACATCT and a plate-specific oligo. Each plate-specific oligo consisted of the following sequences: TruSeq 5' adapter (5' TCTTTCCCTACACGACGCTCTTCCGATCT 3'), three random bases (5' NNN 3'), seven nucleotide long plate-specific barcode, Tn5 mosaic sequence (5' AGATGTGTATAAGAGACAG 3'). Annealing was performed by heating equimolar ratios of oligos in the presence 50mM NaCl at 95°C for 5 minutes, followed by slow-cooling of the mixture at room temperature. Purified Tn5 enzyme was mixed with the annealed product and kept at room temperature for 30 minutes to allow DNA loading.

Because efficiency of cluster formation and amplification during sequencing is restricted by size, the Tn5 tagmentation was optimized to yield fragments from 300 to 1000 bp appropriate for the Illumina MiSeq platform. Each tagmentation reaction was performed using pre-loaded Tn5 generated above in HEPES buffer (10mM HEPES-KOH, 5mM $MgCl_2$, 10% v/v DMF) at 55°C for 20 minutes. To stop the reaction, SDS (0.04% final) was then added and incubated at room temperature for 7 minutes. Reaction products were purified by PCR purification columns (Qiagen).

Finally, purified and tagmented products across multiple 96-well plates were pooled and subjected to low 7-cycle enrichment PCR. Primers used were forward 5' AATGATACGGCGACC ACCGAGATCTACACTCTTTCCCTACACGACGC 3' and reverse 5' CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTG 3'. Next, PCR products were purified using 0.6X AMPure XP beads (Beckman Coulter). Lastly, purified DNA was sequenced paired-end on Illumina MiSeq.

Plant material, stress treatments, sampling and RNA preparation

We used the rice cultivar Kitaake (*Oryza sativa L. ssp. japonica*) for tissue sampling and RNA preparation. To get the maximum coverage of the transcriptome, tissue samples were collected from different stages of development and in response to biotic and abiotic stress treatments. The developmental tissues including mature leaf, flag leaf, leaf sheath, stem nodes, stem internodes, 0-3 cm panicles, 3-15 cm panicles, mature panicles before anthesis, developing seeds at 0 days and 15 days after anthesis and, mature seeds were collected from greenhouse-grown plants. Two-week-old seedlings grown separately under light and dark conditions in a growth chamber were also sampled.

For cold stress treatment, two-week-old rice seedlings were exposed to decreasing temperatures from 15°C to 10°C and then 5°C for 24 hours. Leaf tissue was harvested after each treatment. For water deficit stress, two-week-old rice seedlings were gradually subjected to 75, 50 and 25% water deficit stress and samples were collected at each treatment. For salt stress treatment, two-week-old seedlings were subjected to increasing levels of salinity (50 mM, 100 mM and 150 mM) for 24 hours and leaf tissue was collected after

every treatment. For submergence stress, three-week-old seedlings in soil containing pots were completely submerged in plastic tanks filled with water and leaf tissue was harvested at 0-, 1- and 6- days post submergence. For pathogen inoculation, plants were grown in pots in a greenhouse for five weeks and then transferred to a growth chamber (14 h daytime period, 28/26°C temperature cycle and 90% humidity, light intensity 100 µmol m−2 s−1). Five to six-week-old plants were inoculated with bacterial suspension (OD600 0.5) of *Xanthomonas Oryzae pv. Oryzae* strain PXO99 (Philippine race 6) using scissors-dip method[2]. The leaf tissues were sampled 0-, 1- and 4-days post inoculation.

All tissue samples were flash frozen in liquid nitrogen and stored at -80°C for RNA extraction. Total RNA was extracted from each tissue sample using TRIzol reagent (Invitrogen, CA), treated with DNase I (Ambion) and purified using Macherey-Nagel Nucleospin RNA II kit (Macherey-Nagel, Duren, Germany) as per manufacturer's protocol. Purified RNA was quantified using a NanoDrop ND-100 spectrophotometer (Thermo Scientific) and equal quantity of RNA from different developmental stages and stress treatments was pooled in one tube for cDNA synthesis.

Construction of the *O. sativa* open reading frame library (ORFeome)

To construct a first draft ORFeome for rice that covered a manageable portion of the *O. sativa* genome, we first used RiceNet[3,4] to prioritize a subset of rice genes to clone. RiceNet leverages a combination of co-expression, domain co-occurrence, protein-protein interaction, genetic interaction, and phylogenetic profile similarity features to report a likelihood that pairs of genes share a functional association. Using a loose likelihood threshold, we identified 3,269 genes with predicted association with a seed set of 89 genes previously associated with biotic or abiotic stress tolerance[3,5]. The 89 seed genes and predicted associations are reported in **Supplementary Dataset 2**.

A single representative ORF was selected for each prioritized gene and the pairs of primers listed in **Supplementary Dataset 3** were designed to clone each ORF. To ensure maximum coverage of the prioritized *O. sativa* ORFs, RNA was isolated from a wide range of rice plant parts (e.g. leaf, stem, nodes, roots), at different developmental stages, and at various stress conditions (e.g. light, dark, cold-stress, salt-stress, drought-stress) as described above and detailed in **Supplementary Dataset 4**. RNAs were converted to cDNAs and used as templates to append Gateway *attB1* and *attB2* cloning sites to flank the start and stop codons, respectively, using ORF-specific PCR. Amplicons were cloned by Gateway BP reactions into entry vector pDONR223 and transformed into bacterial carrier *DH5α*.

As the BP cloning procedure might inadvertently introduce unwanted PCR artifacts into the entry vectors, we manually picked 2 bacterial transformants per entry clone and verified their identities by PLATE-seq. Only validated clones were retained in the *O. sativa* ORFeome. In cases where duplicate clones were detected, the clone with stronger sequencing evidence was retained.

Yeast two-hybrid (Y2H)

Y2H experiments were carried out as previously described by us and other groups[6-11]. In brief, *O. sativa* ORFs in entry vectors pDONR223 were first cloned into pDEST AD and DB destination vectors using Gateway LR reactions to generate N-terminal ORF fusions. We refer to these expression clones as AD-Y and DB-X. All AD-Y and DB-X expression clones were then transformed into Y2H *Saccharomyces cerevisiae* strains *MAT***a** Y8800 and *MAT*α Y8930 (genotype: *leu2-3, 112 trp1-901 his3Δ200 ura3-52 gal4Δ gal80Δ GAL2::ADE2 GAL1::HIS3@LYS2 GAL7::lacZ@MET2 cyh2R*), respectively. Next, we screened for autoactivators by individually mating each DB-X strain with a *MAT***a** Y8800 strain carrying the empty pDEST AD destination vector. To identify AD autoactivators, we mated each AD-Y strain with a *MAT*α Y8930 strain carrying empty pDEST DB destination vector. After allowing the yeast to mate on yeast extract peptone dextrose (YEPD) (1% yeast extract, 2% bactopeptone, 2% glucose, 0.45mM adenine sulfate) 2% agar plates at 30°C overnight, yeast were replica plated onto synthetic complete 2% agar plates without leucine and tryptophan (SC+Ade–Leu–Trp+His) and incubated at 30°C overnight to select for diploids with both pDEST AD and DB vector backbones. Finally, diploids were replica plated onto synthetic complete 2% agar

plates with 1mM 3-amino-1,2,4-triazole (3AT) and without leucine, tryptophan, and histidine (SC+Ade–Leu–Trp–His+1mM 3AT). Plates were incubated at 30°C for 3-5 days. Any AD-Y or DB-X that grew on SC+Ade–Leu–Trp–His+3AT were scored as autoactivators. We excluded autoactivators from all further screenings.

Thereafter, we performed the first round of testing (called phenotyping I) by mating each unique DB-X with individual mini-pools of 188 unique AD-Y on YEPD 2% agar plates. We selected for diploids by replica plating onto SC+Ade–Leu–Trp+His. To select for positive interactions, we performed the Y2H screening by replica plating the diploids onto SC+Ade–Leu–Trp–His+3AT and incubating at 30°C for 4 days. We used sterile toothpicks to pick and inoculate all positives into liquid SC+Ade–Leu–Trp+His to keep the yeast in the diploid state.

Next, all yeast colonies picked from phenotyping I were individually subjected to another round of Y2H testing called phenotyping II. Here, all picked colonies were spotted directly onto 2% agar plates of SC+Ade–Leu–Trp–His+3AT and SC–Ade–Leu–Trp+His. Plates were incubated at 30°C for 4 days. Positives from this round of screening were picked into liquid SC+Ade–Leu–Trp+His to keep the yeast in the diploid state.

Positive yeast picked from phenotyping II were subject to extraction of plasmid DNA by lysis using zymolyase enzyme (Seigakaku Corporation). Cell and enzyme were incubated at 37°C for 45 minutes and then at 95°C for 10 minutes. The identities of DB-X and AD-Y were determined by PLATE-seq.

Finally, for every AD-Y and DB-X interaction candidate identified by PLATE-seq, we performed pairwise Y2H testing of each identified pair. This was done by first mating each individual AD-Y with the DB-X putative interaction partner on YEPD plate at 30°C overnight. Then, diploids were selected by replica plating onto SC+Ade–Leu–Trp+His plate and incubating at 30°C overnight. Finally, diploids were selected for interaction-positive cells by replica plating onto SC+Ade–Leu–Trp–His+3AT and SC–Ade–Leu–Trp+His plates and incubating at 30°C for 4-7 days. To identify *de novo* autoactivators (autoactivators that likely arise from accumulation of random mutations during the screening process), we concurrently mated each DB-X with a *MAT***a** Y8800 strain carrying the empty pDEST AD destination vector. Afterwards, we followed the same procedure as during the first autoactivator-detection screen. All identified *de novo* autoactivators were removed from the screens. Thus, at the conclusion of the pairwise Y2H phase, we were able to definitively identify and verify all interacting AD-Y and DB-X while controlling for all *de novo* autoactivators.

PLATE-seq data analyses

Downstream analysis of PLATE-seq sequencing results was performed in order to determine the identity and positions of all clones sequenced. First, computational deconvolution of the sequencing data was performed to group reads according to their original location. For each paired read, the position- and plate-specific barcodes were identified from R2 and R1 respectively. For pooled sequencing of Y2H interaction candidates, an additional barcode for distinguishing AD-Y clones from DB-X clones was included on R2. After reads were grouped by location, the identifies of any ORFs present in each well were determined through BWA alignment (BWA version 0.7.12-r1039). The reference index was created from a list of predicted *O. sativa* ORF sequences using bwa index [reference]. Alignments were generated using bwa mem -a -t 12 [reference] [query] > [output]. In cases where multiple alignments were reported from one read, either the highest quality alignment was retained, or in cases of a tie, the read count was split equally among all alignments. The final output from the initial deconvolution provided read alignment counts for all ORFs detected per well.

These alignment counts were then processed uniquely depending on the sequencing application. For sequence verification and selection of clones to be included in the ORFeome, wells were first filtered to eliminate empty wells or wells containing multiple ORFs. Wells containing fewer than 50 reads in total or in which the most prevalent ORF represented fewer than 20% of the aligned reads in the well were removed. Additionally, any wells where the most prevalent ORF was detected in the reverse orientation were removed. The identities of each well were then called based on the most prevalent ORF. In order to retain

only the highest quality single-isolates in the final ORFeome, in cases where multiple sequenced clones matched the same ORF, the isolate with the highest quality was retained.

For detection of candidate Y2H interactions in the pooled sequencing setup, criteria were loosened in order to minimize false negatives. ORF alignment counts per-well were obtained as described above and used to define two sets of potentially present ORFs; one for AD-Y ORFs and one DB-X ORFs for. Any ORF that had at least 50 aligned reads in total and represented at least 20% of the well was retained in these sets. To avoid dropping "empty" wells the majority ORF for each AD-Y and DB-X was retained regardless of the total number of aligned reads. All pairwise combinations of these detected AD-Y and DB-X ORFs were reported as putative interactions to be verified independently by follow-up Y2H.

Bimolecular fluorescence complementation (BiFC)

BiFC assays were performed using onion infiltration. Vectors used were Gateway-compatible constructs pSAT4-DEST-N(1-174)EYFP-C1 (CD3-1089), pSAT4A-DEST-N(1-174)EYFP-N1 (CD3-1080), pSAT5-DEST-C(175-end)EYFP-C1 (CD3-1097) and pSAT5A-DEST-C(175-end)EYFP-N1 (CD3-1096) and were acquired from the *Arabidopsis* Biological Resource Center (ABRC) and described previously[12]. To allow these plasmids to replicate in *Agrobacterium*, the pSa origin-of-replication from pGREENII-0179 was inserted adjacent to the *E. coli* origin-of-replication.

We selected seven high confidence protein pairs from our yeast two-hybrid screen and cloned each ORF into the four modified pSAT vectors using an LR recombination reaction. After, expression clones were transformed individually into *Agrobacterium* strain GV3101 carrying the pSOUP helper plasmid. Liquid cultures were grown from selected agrobacteria colonies and used to infiltrate onion as described by Xu *et al*.[13]. In brief, the agrobacteria were pelleted by centrifugation and resuspended into the complete resuspension buffer recommended by Xu *et al*. Each culture was then diluted to an OD_{600} of 0.1. For each interaction pair, we then prepared eight mixtures representing all possible interaction orientations by pipetting equal volumes of the appropriate *Agrobacterium* strains into new tubes. Next, we infiltrated approximately 100-200 µL of the *Agrobacterium* mixtures as described by Xu *et al*. We then incubated the samples in the dark at 28^{o}C for 3-4 days before performing confocal microscopy on epidermal peels taken from the onion samples.

Confocal images were collected on a Leica TCS-SP5 microscope (Leica Microsystems, Exton, PA USA) using a 20X water immersion objective. YFP was excited with the blue argon ion laser (488 nm), and emitted light was collected between 525nm and 595nm. Three images of each sample were taken at random locations on each sample with no changes made to the instrument settings between images or samples except for adjusting the focus. Bright field images were collected simultaneously with the fluorescence images using the transmitted light detector. Images were processed using Leica LAS-AF software (version 3.3.0) and fluorescence was quantified using ImageJ (version 1.51n). Reported values are a ratio of the absolute fluorescence measured for each sample relative to the positive control. The positive control used for comparison was the AtABI1-AtOST1 strong interaction and the negative control used was AtOST1-AtOIP26 as previously reported[14].

Rice Expression Analysis

In order to determine whether interacting rice genes exhibited higher co-expression than non-interacting rice genes, we first obtained rice expression data from the Michigan State University Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/expression.shtml)[15]. Only the first 16 expression libraries corresponding to RNA sequencing data derived from tissue at various developmental stages under physiological conditions were used. Gene co-expression was reported as the Spearman correlation coefficient between the expression vectors for two genes. All homo-dimer interactions were removed from our rice interactome for the analysis. An equal number of hetero-dimer pairs were randomly sampled from either the ORFeome or entire *O. sativa* genome for comparison.

Rice GO Term Semantic Similarity Analysis
Analysis of the semantic similarities between GO annotations among rice gene pairs was performed using GOssTo[16,17]. GOSlim assignments for ~30,000 *O. sativa* ORFs were obtained using the batch download feature from the Michigan State University Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/downloads_gad.shtml)[15]. The directed acyclic graph (DAG) used to represent the ontology relationship was the core ontology available through The Gene Ontology Resource (http://purl.obolibrary.org/obo/go.obo)[18,19]. In order to increase coverage across all ontology relationship and thus improve GOssTo performance, the MSU annotations were supplemented with a secondary set of *O. sativa* proteome annotations downloaded through the European Bioinformatics Institute (EMBL-EBI) ftp server (ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/proteomes/2610640.O_sativa_subsp_japonica_Rice.goa)[20]. GOssTo was run with the following command…

```
java -Xms64G -jar Gossto.jar -calculationdata genewise -calculationtype ism -evidencecodes EXP,IDA,IPI,IMP,IGI,IEP,TAS,IC -goapath [annotations] -hsm Resnik -hsmoutput [out1] -ismoutput [out2] -matrixStyle m -obopath [ontologyDAG] -ontology all -relations part_of,is_a -weightedJaccard True
```

This command generated pairwise sematic similarity scores between all rice proteins for each the molecular function, biological process, and cellular component GOSlim terms. We retained the final "integrated similarity measure" (ISM) outputs for all further analyses. We compared the distributions of these semantic similarities between our reported rice interactome, and non-interacting protein pairs sampled from our ORFeome. A cutoff of 0.75 was selected to distinguish pairs that were functionally similar from pairs that were not functionally similar.

Conservation analysis
In order to analyze how conserved the captured interactions are across the tree of life we used the OrthoMCL pipeline to identify orthologs of the identified interacting genes from *Sorghum bicolor, Solanum lycopersicum, Arabidopsis thaliana, Physcomitrella patens, Saccharomyces cereviseae, Homo sapiens, and Escherichia coli*[21]. To test whether interactions are more or less conserved than random expectation, we randomly sampled genes from the rice genome and determined their breadth of conservation across the sampled species. To test for statistical significance, we performed a bootstrap analysis by 1000 sampling replicates and tested if the actual number of captured interacting genes conserved at each node was significantly different from the random expectation using the z-test function in R.

Literature Interactome Set
In order to compare our novel rice interactions to those previously reported in the literature, we used a curated set of high-quality protein-protein interactions[22] compiled from seven primary interaction databases; BioGRID[23], MINT[24], iRefWeb[25], DIP[26], IntAct[27], HPRD[28,29], MIPS[30], and the PDB[31,32]. The interaction set used was *O. sativa* binary high-quality containing 237 interactions downloaded on May 15, 2019. This interaction set was supplemented with 372 additional interactions from a high-throughput Y2H rice-kinase interactome screen[33] for a total of 609 previously reported interactions.

Literature Support for Detected Interaction
A total of 20 highly co-expressed (SCC ≥ 0.8) interacting gene pairs were manually searched for previous literature evidence supporting the existence of an interaction in other species. As detailed in **Supplementary Dataset 8**, in 12 out of 20 interactions searched (60%), previous studied had detected the interaction among homologous genes in other species[34-53].

***De novo* Sequence Assembly for the Human *BIRC7* ORF**

To explore applications of PLATE-seq beyond clone identification, we used both a custom sequence assembly script and the online version of CAP3[54]. For the custom script a random set of seed reads was selected to begin assembly. These starting contigs were iteratively expanded by aligning all remaining reads to them and incorporating overhanging alignments into the contigs. Intermediate contigs were merged together when sufficient overlap was detected between them. This process was repeated until only one candidate contig remained or until no additional reads could be incorporated into the existing contigs. Final assemblies between our custom script and CAP3 agreed with each other. This method was only applied for demonstrative purposes on the *BIRC7* test case shown in **Fig. S2a** and has not be extended to our whole ORFeome.

Identification of Variants in Human *BIRC7* ORF

To explore applications of PLATE-seq beyond clone identification, we applied the sequence analysis scripts from our established CLONE-seq pipeline[55,56] to identify variants relative to the reference sequence. In brief, all reads were aligned to the *BIRC7* reference sequence and all possible mutations were scored based on the ratio of non-reference reads to reference reads at each position, normalized by the sequencing error rate estimated from neighboring positions. This method was only applied for demonstrative purposes to identify the previously reported C882T variant in the *BIRC7* clone shown in **Fig. S2d**.

Supplementary References

1 Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res* **24**, 2033-2040, doi:10.1101/gr.177881.114 (2014).

2 Kauffman, H. E., Reddy, A.P.K., Hsieh, S.P.Y. and Merca, S.D. An improved technique for evaluating resistance of rice varieties to Xanthomonas oryzae. *Plant Disease Reporter* **57**, 537-541 (1973).

3 Lee, I. *et al.* Genetic dissection of the biotic stress response using a genome-scale gene network for rice. *Proc Natl Acad Sci U S A* **108**, 18548-18553, doi:10.1073/pnas.1110384108 (2011).

4 Lee, T. *et al.* RiceNet v2: an improved network prioritization server for rice genes. *Nucleic Acids Res* **43**, W122-127, doi:10.1093/nar/gkv253 (2015).

5 Seo, Y. S. *et al.* Towards establishment of a rice stress response interactome. *PLoS Genet* **7**, e1002020, doi:10.1371/journal.pgen.1002020 (2011).

6 Arabidopsis Interactome Mapping, C. Evidence for network evolution in an Arabidopsis interactome map. *Science* **333**, 601-607, doi:10.1126/science.1203877 (2011).

7 Vo, T. V. *et al.* A Proteome-wide Fission Yeast Interactome Reveals Network Evolution Principles from Yeasts to Human. *Cell* **164**, 310-323, doi:10.1016/j.cell.2015.11.037 (2016).

8 Das, J. *et al.* Cross-species protein interactome mapping reveals species-specific wiring of stress response pathways. *Sci Signal* **6**, ra38, doi:10.1126/scisignal.2003350 (2013).

9 Yu, H. *et al.* High-quality binary protein interaction map of the yeast interactome network. *Science* **322**, 104-110, doi:10.1126/science.1158684 (2008).

10 Rual, J. F. *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173-1178, doi:10.1038/nature04209 (2005).

11 Li, S. *et al.* A map of the interactome network of the metazoan C. elegans. *Science* **303**, 540-543, doi:10.1126/science.1091403 (2004).

12 Williams, B., Kabbage, M., Britt, R. & Dickman, M. B. AtBAG7, an Arabidopsis Bcl-2-associated athanogene, resides in the endoplasmic reticulum and is involved in the unfolded protein response. *Proc Natl Acad Sci U S A* **107**, 6088-6093, doi:10.1073/pnas.0912670107 (2010).

13 Xu, K. *et al.* A rapid, highly efficient and economical method of Agrobacterium-mediated in planta transient transformation in living onion epidermis. *PLoS One* **9**, e83556, doi:10.1371/journal.pone.0083556 (2014).

14 Waadt, R. *et al.* Identification of Open Stomata1-Interacting Proteins Reveals Interactions with Sucrose Non-fermenting1-Related Protein Kinases2 and with Type 2A Protein Phosphatases That Function in Abscisic Acid Responses. *Plant Physiol* **169**, 760-779, doi:10.1104/pp.15.00575 (2015).

15 Kawahara, Y. *et al.* Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. *Rice (N Y)* **6**, 4, doi:10.1186/1939-8433-6-4 (2013).

16 Caniza, H. *et al.* GOssTo: a stand-alone application and a web tool for calculating semantic similarities on the Gene Ontology. *Bioinformatics* **30**, 2235-2236, doi:10.1093/bioinformatics/btu144 (2014).

17 Yang, H., Nepusz, T. & Paccanaro, A. Improving GO semantic similarity measures by exploring the ontology beneath the terms and modelling uncertainty. *Bioinformatics* **28**, 1383-1389, doi:10.1093/bioinformatics/bts129 (2012).

18 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29, doi:10.1038/75556 (2000).

19 The Gene Ontology, C. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res* **47**, D330-D338, doi:10.1093/nar/gky1055 (2019).
20 Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* **47**, W636-W641, doi:10.1093/nar/gkz268 (2019).
21 Li, L., Stoeckert, C. J., Jr. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* **13**, 2178-2189, doi:10.1101/gr.1224503 (2003).
22 Das, J. & Yu, H. HINT: High-quality protein interactomes and their applications in understanding human disease. *BMC Syst Biol* **6**, 92, doi:10.1186/1752-0509-6-92 (2012).
23 Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* **34**, D535-539, doi:10.1093/nar/gkj109 (2006).
24 Licata, L. *et al.* MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res* **40**, D857-861, doi:10.1093/nar/gkr930 (2012).
25 Turner, B. *et al.* iRefWeb: interactive analysis of consolidated protein interaction data and their supporting evidence. *Database (Oxford)* **2010**, baq023, doi:10.1093/database/baq023 (2010).
26 Xenarios, I. *et al.* DIP: the database of interacting proteins. *Nucleic Acids Res* **28**, 289-291, doi:10.1093/nar/28.1.289 (2000).
27 Orchard, S. *et al.* The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res* **42**, D358-363, doi:10.1093/nar/gkt1115 (2014).
28 Peri, S. *et al.* Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res* **13**, 2363-2371, doi:10.1101/gr.1680803 (2003).
29 Keshava Prasad, T. S. *et al.* Human Protein Reference Database--2009 update. *Nucleic Acids Res* **37**, D767-772, doi:10.1093/nar/gkn892 (2009).
30 Pagel, P. *et al.* The MIPS mammalian protein-protein interaction database. *Bioinformatics* **21**, 832-834, doi:10.1093/bioinformatics/bti115 (2005).
31 Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235-242, doi:10.1093/nar/28.1.235 (2000).
32 Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat Struct Biol* **10**, 980, doi:10.1038/nsb1203-980 (2003).
33 Ding, X. *et al.* A rice kinase-protein interaction map. *Plant Physiol* **149**, 1478-1492, doi:10.1104/pp.108.128298 (2009).
34 Sanders, J., Brandsma, M., Janssen, G. M. C., Dijk, J. & Moller, W. Immunofluorescence studies of human fibroblasts demonstrate the presence of the complex of elongation factor-1 beta gamma delta in the endoplasmic reticulum. *Journal of Cell Science* **109**, 1113-1117 (1996).
35 Kim, K. K., Kim, R. & Kim, S. H. Crystal structure of a small heat-shock protein. *Nature* **394**, 595-599, doi:10.1038/29106 (1998).
36 Tanabe, N., Yoshimura, K., Kimura, A., Yabuta, Y. & Shigeoka, S. Differential expression of alternatively spliced mRNAs of Arabidopsis SR protein homologs, atSR30 and atSR45a, in response to environmental stress. *Plant Cell Physiol* **48**, 1036-1049, doi:10.1093/pcp/pcm069 (2007).
37 Golovkin, M. & Reddy, A. S. An SC35-like protein and a novel serine/arginine-rich protein interact with Arabidopsis U1-70K protein. *J Biol Chem* **274**, 36428-36438, doi:10.1074/jbc.274.51.36428 (1999).
38 Riegler, H. *et al.* Crystal structure and functional characterization of a glucosamine-6-phosphate N-acetyltransferase from Arabidopsis thaliana. *Biochem J* **443**, 427-437, doi:10.1042/BJ20112071 (2012).

39 Lee, S. J. & Baserga, S. J. Imp3p and Imp4p, two specific components of the U3 small nucleolar ribonucleoprotein that are essential for pre-18S rRNA processing. *Mol Cell Biol* **19**, 5441-5452, doi:10.1128/mcb.19.8.5441 (1999).
40 Sa-Moura, B. *et al.* Mpp10 represents a platform for the interaction of multiple factors within the 90S pre-ribosome. *PLoS One* **12**, e0183272, doi:10.1371/journal.pone.0183272 (2017).
41 Huang, X. Y. *et al.* CYCLIN-DEPENDENT KINASE G1 is associated with the spliceosome to regulate CALLOSE SYNTHASE5 splicing and pollen wall formation in Arabidopsis. *Plant Cell* **25**, 637-648, doi:10.1105/tpc.112.107896 (2013).
42 Lorkovic, Z. J., Lehner, R., Forstner, C. & Barta, A. Evolutionary conservation of minor U12-type spliceosome between plants and humans. *RNA* **11**, 1095-1107, doi:10.1261/rna.2440305 (2005).
43 Anderson, L. E., Yousefzai, R., Ringenberg, M. R. & Carol, A. A. Both chloroplastic and cytosolic fructose bisphosphatase isozymes are present in the pea leaf nucleus. *Plant Science* **166**, 721-730, doi:10.1016/j.plantsci.2003.11.008 (2004).
44 Beatrix, B., Sakai, H. & Wiedmann, M. The alpha and beta subunit of the nascent polypeptide-associated complex have distinct functions. *J Biol Chem* **275**, 37838-37845, doi:10.1074/jbc.M006368200 (2000).
45 Couvreur, B. *et al.* Eubacterial HslV and HslU subunits homologs in primordial eukaryotes. *Mol Biol Evol* **19**, 2110-2117, doi:10.1093/oxfordjournals.molbev.a004036 (2002).
46 Song, H. K. *et al.* Isolation and characterization of the prokaryotic proteasome homolog HslVU (ClpQY) from Thermotoga maritima and the crystal structure of HslV. *Biophysical Chemistry* **100**, 437-452, doi:10.1016/s0301-4622(02)00297-1 (2002).
47 Kwon, A.-R., Kessler, B. M., Overkleeft, H. S. & McKay, D. B. Structure and Reactivity of an Asymmetric Complex between HslV and I-domain Deleted HslU, a Prokaryotic Homolog of the Eukaryotic Proteasome. *Journal of Molecular Biology* **330**, 185-195, doi:10.1016/s0022-2836(03)00580-1 (2003).
48 Schauber, C. *et al.* Rad23 links DNA repair to the ubiquitin/proteasome pathway. *Nature* **391**, 715-718, doi:10.1038/35661 (1998).
49 Dantuma, N. P., Heinen, C. & Hoogstraten, D. The ubiquitin receptor Rad23: at the crossroads of nucleotide excision repair and proteasomal degradation. *DNA Repair (Amst)* **8**, 449-460, doi:10.1016/j.dnarep.2009.01.005 (2009).
50 Dantuma, N. P. & Lindsten, K. Stressing the ubiquitin-proteasome system. *Cardiovasc Res* **85**, 263-271, doi:10.1093/cvr/cvp255 (2010).
51 Dantuma, N. P. & Bott, L. C. The ubiquitin-proteasome system in neurodegenerative diseases: precipitating factor, yet part of the solution. *Front Mol Neurosci* **7**, 70, doi:10.3389/fnmol.2014.00070 (2014).
52 Silva, G. M. *et al.* Role of glutaredoxin 2 and cytosolic thioredoxins in cysteinyl-based redox modification of the 20S proteasome. *FEBS J* **275**, 2942-2955, doi:10.1111/j.1742-4658.2008.06441.x (2008).
53 Demasi, M. *et al.* Redox regulation of the proteasome via S-glutathionylation. *Redox Biol* **2**, 44-51, doi:10.1016/j.redox.2013.12.003 (2013).
54 Huang, X. & Madan, A. CAP3: A DNA sequence assembly program. *Genome Res* **9**, 868-877, doi:10.1101/gr.9.9.868 (1999).
55 Wei, X. *et al.* A massively parallel pipeline to clone DNA variants and examine molecular phenotypes of human disease mutations. *PLoS Genet* **10**, e1004819, doi:10.1371/journal.pgen.1004819 (2014).

56 Fragoza, R. *et al.* Extensive disruption of protein interactions by genetic variants across the allele frequency spectrum in human populations. *Nat Commun* **10**, 4141, doi:10.1038/s41467-019-11959-3 (2019).

Supplementary Dataset Headings

Supplementary Dataset 1. List of primers used for PLATE-seq pipeline in this study.

Supplementary Dataset 2. List of 89 genes used to seed RiceNet prioritization and predicted gene interactions.

Supplementary Dataset 3. List of primers used to amplify rice cDNA.

Supplementary Dataset 4. Developmental stage, environmental exposures, and parts of rice plant used for RNA isolations.

Supplementary Dataset 5. O. sativa ORFeome.

Supplementary Dataset 6. O. sativa protein interactome.

Supplementary Dataset 7. Quantification and statistical analyses of BiFC experiments.

Supplementary Dataset 8. Manual literature search on top 20 interacting genes.

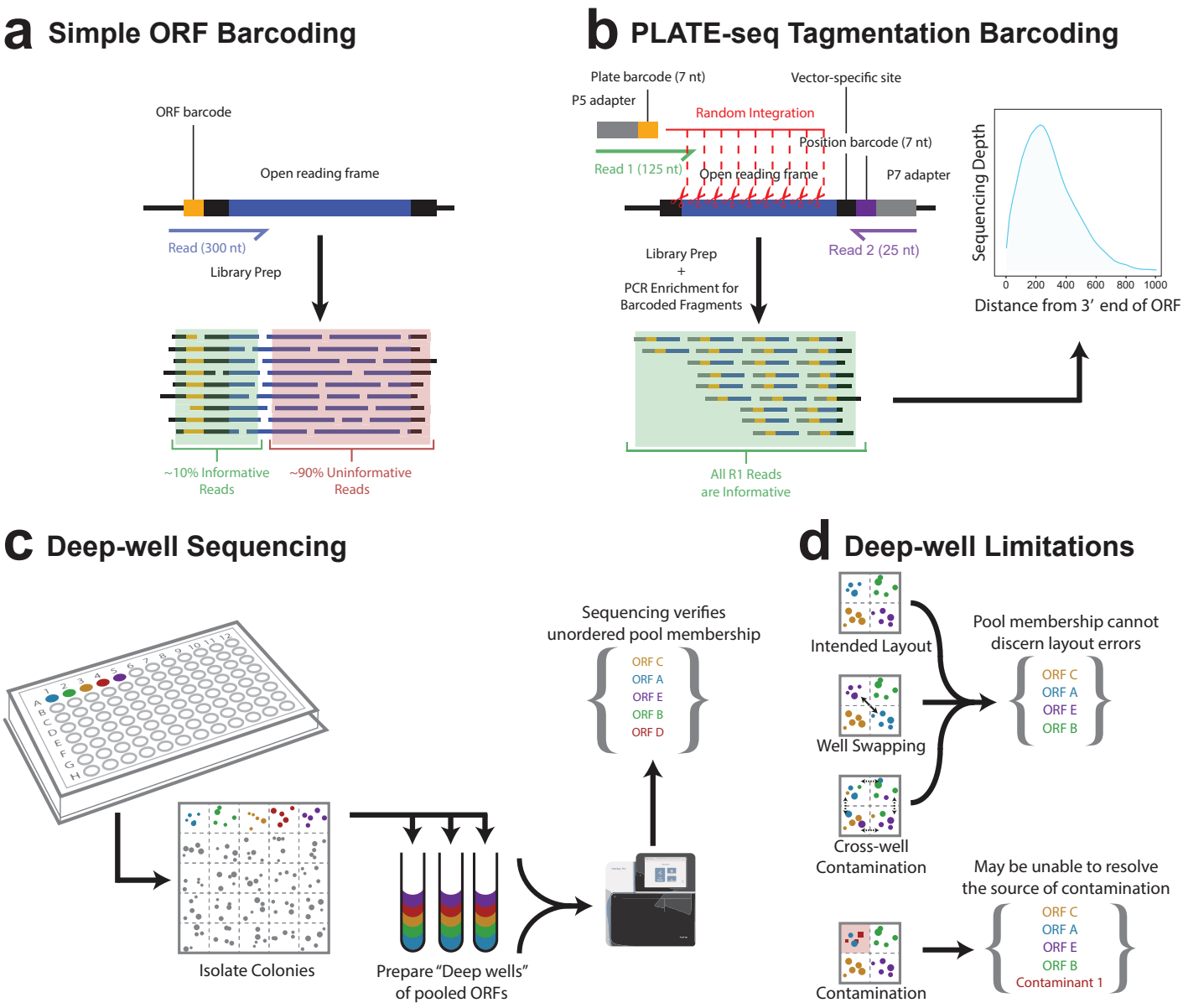


Figure S1. Schematic comparison of other high-throughput sequencing strategies to PLATE-seq. **a,** Simple barcoding strategies typically append individual barcodes at the beginning of each ORF. A lengthy vector sequence generally separates the barcode from the beginning of the ORF, so even a long 300 bp read only provides coverage limited to a small portion from the 5' end of the ORF. Moreover, the library prep generates a substantial fraction of uninformative reads with no barcode. **b,** By contrast, the PLATE-seq approach employs Tn5 tagmentation to randomly insert the P5 adapter and plate barcode within the ORF. Thus, even employing a much shorter 150 bp total read, nearly 1,000 bp from the 3' end of the ORF can be sequenced reliably. The kernel density estimate shows the average sequencing depth derived from PLATE-seq results from 94 ORFs included in our human positive control plate. Moreover, all fragments generated through the PLATE-seq library prep contain both barcodes and thus are all informative. **c,** Schematic representation of a deep-well sequencing strategy which pools many wells together to report the overall ORF membership across an entire plate but cannot verify the location of each ORF. **d**, The deep-well sequencing strategy is consequentially unable to differentiate permutations on the same set of ORFs, or verify the source of any contaminants.

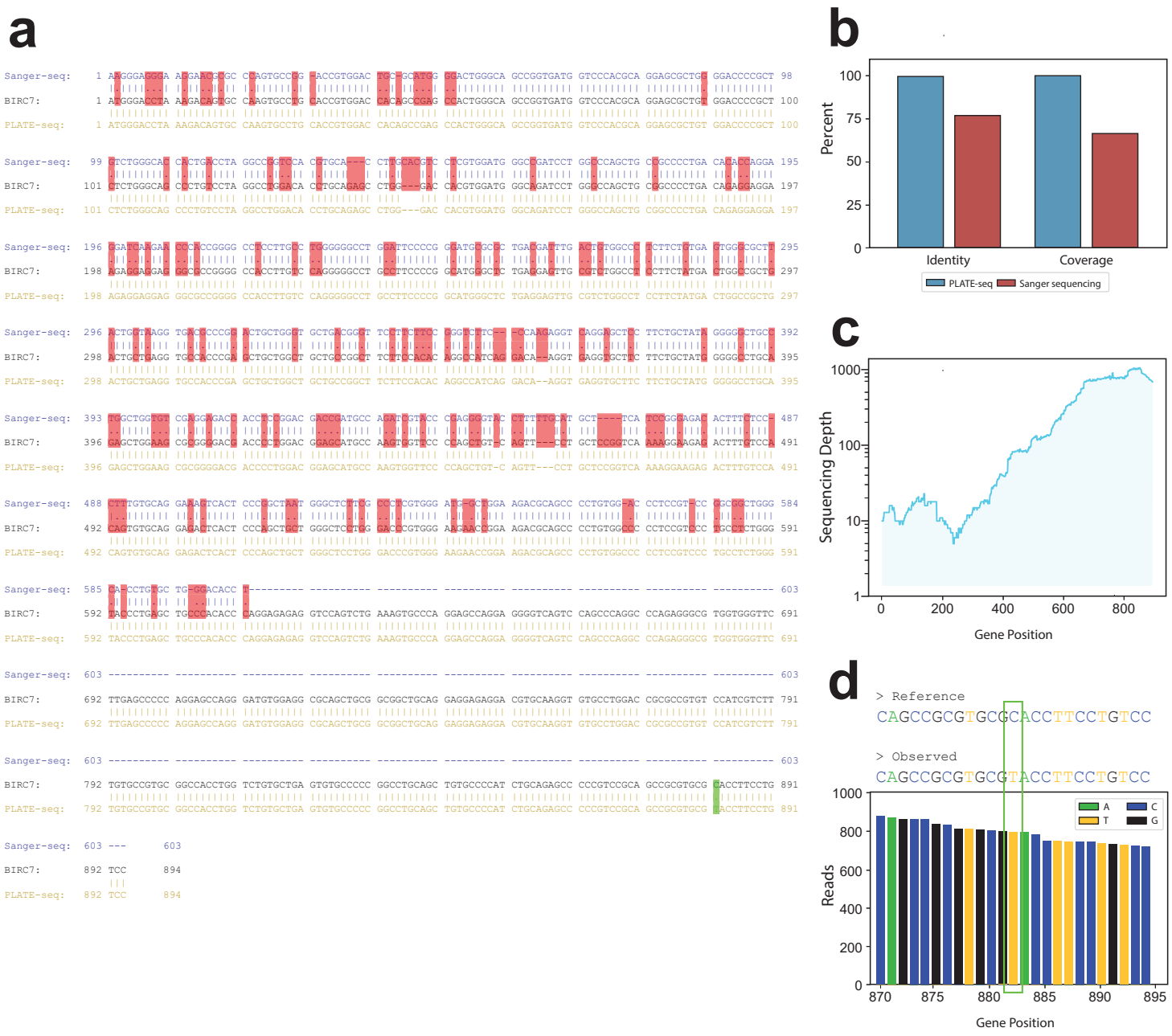


Figure S2. Demonstration of sequence reconstruction and variant calling as alternative applications for PLATE-seq.

a, A contig assembly script was applied to reads from one of the wells of our human control plate containing an ORF for the human gene *BIRC7* (ORF ID 7606 in the human ORFeome 8.1 library). The alignment between the *BIRC7* reference (center), Saner Sequence obtained for this clone (top, blue), and sequence reconstructed from PLATE-seq reads (bottom, yellow) is shown. Disagreements in the alignment are highlighted in red. **b,** Bar plots summarizing the quality of each alignment. For convenience vector backbone sequence from the Sanger and PLATE-seq sequences are trimmed so the alignment only spans coding sequence of *BIRC7*. The Sanger Sequencing result sequenced positions 1-613 (68.6% coverage) with 76.2% identity to the reference sequence. The PLATE-seq result sequenced achieved perfect coverage and percent identity (ignoring one mismatch from a known C882T variant in the clone). **c,** A distribution showing the number of PLATE-seq reads aligned to each position of the *BIRC7* gene. **d,** PLATE-seq reads can be used to identify variants between a reference alignment and the sequenced clone. The stacked read counts for each possible allele observed from all reads aligned to positions 870-894 of *BIRC7* are shown. Highlighted is a C882T variant supported by 784 reads (sufficient coverage to make a definitive call) and which was previously reported for this ORF when the 8.1 library was initially sequenced. This same variant is also highlighted green in panel **a**.

a

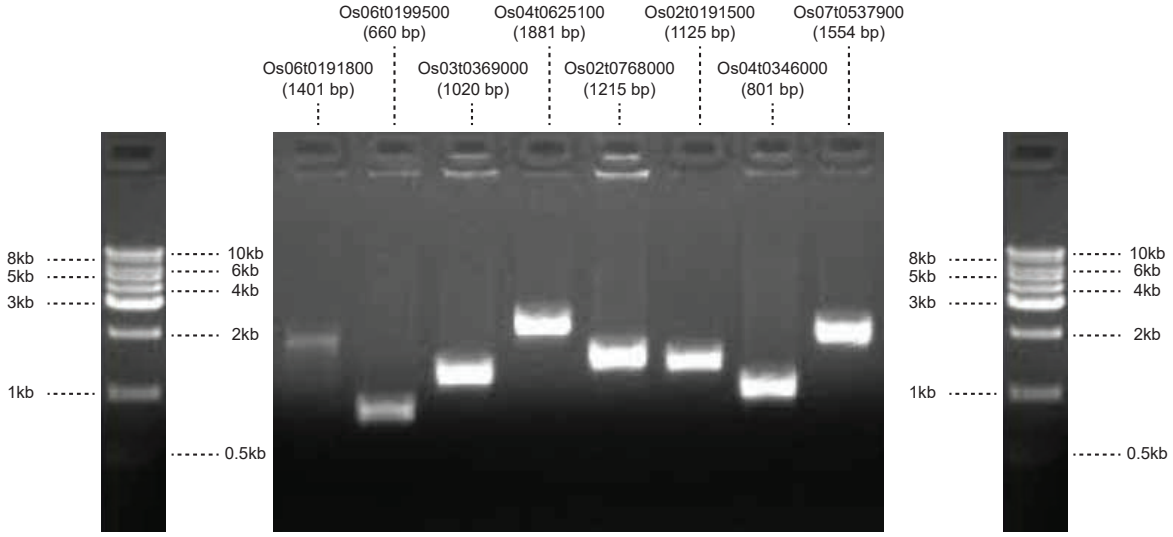


Figure S3. Image of PCR amplicons from a subset of verified *O. sativa* ORFs.
A representative gel in which one row of clones from the final *O. sativa* ORFeome was amplified using primers described in **Supplementary Dataset 3**. These gels were used as a sanity check to confirm the presence and approximate lengths of the desired ORFs. The DNA molecular weight standard lane has been cropped from the same gel image and reproduced on both sides for easier comparison.

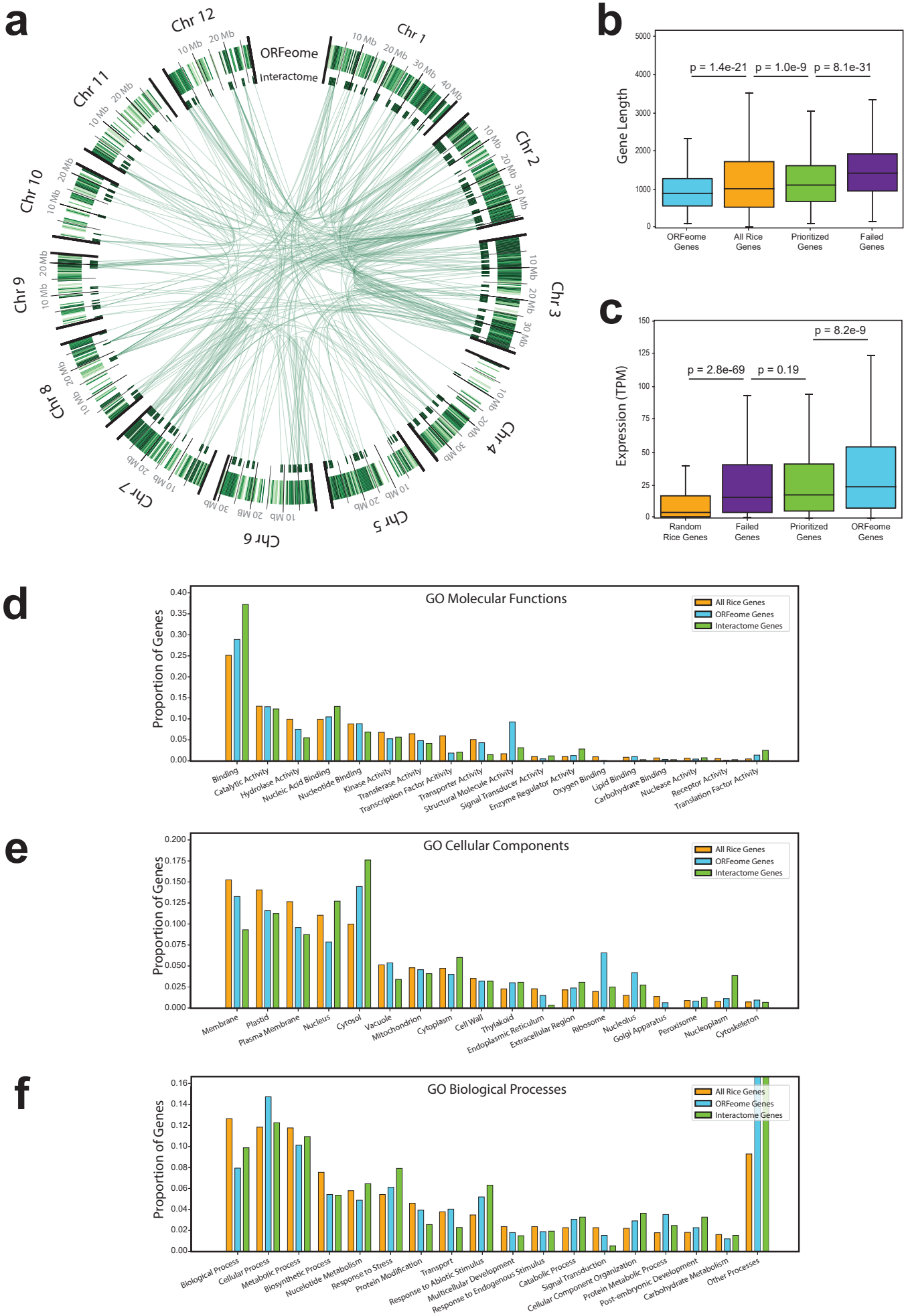
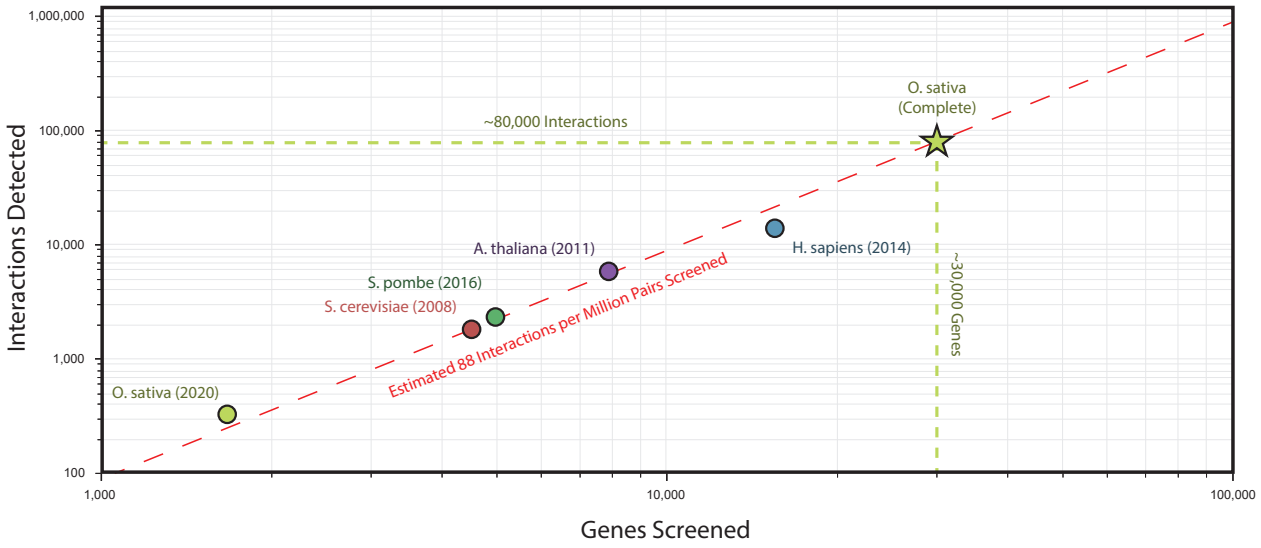

a
ORFeome
Interactome
Chr 1
Chr 2
Chr 3
Chr 4
Chr 5
Chr 6
Chr 7
Chr 8
Chr 9
Chr 10
Chr 11
Chr 12
b
Gene Length
p = 1.4e-21
p = 1.0e-9
p = 8.1e-31
ORFeome Genes
All Rice Genes
Prioritized Genes
Failed Genes
c
Expression (TPM)
p = 2.8e-69
p = 0.19
p = 8.2e-9
Random Rice Genes
Failed Genes
Prioritized Genes
ORFeome Genes
d
GO Molecular Functions
Proportion of Genes
All Rice Genes
ORFeome Genes
Interactome Genes
Binding
Catalytic Activity
Hydrolase Activity
Nucleic Acid Binding
Nucleotide Binding
Kinase Activity
Transferase Activity
Transcription Factor Activity
Transporter Activity
Structural Molecule Activity
Signal Transducer Activity
Enzyme Regulator Activity
Oxygen Binding
Lipid Binding
Carbohydrate Binding
Nuclease Activity
Receptor Activity
Translation Factor Activity
e
GO Cellular Components
Proportion of Genes
All Rice Genes
ORFeome Genes
Interactome Genes
Membrane
Plastid
Plasma Membrane
Nucleus
Cytosol
Vacuole
Mitochondrion
Cytoplasm
Cell Wall
Thylakoid
Endoplasmic Reticulum
Extracellular Region
Ribosome
Nucleolus
Golgi Apparatus
Peroxisome
Nucleoplasm
Cytoskeleton
f
GO Biological Processes
Proportion of Genes
All Rice Genes
ORFeome Genes
Interactome Genes
Biological Process
Cellular Process
Metabolic Process
Biosynthetic Process
Nucleotide Metabolism
Response to Stress
Protein Modification
Transport
Response to Abiotic Stimulus
Multicellular Development
Response to Endogenous Stimulus
Catabolic Process
Signal Transduction
Cellular Component Organization
Protein Metabolic Process
Post-embryonic Development
Carbohydrate Metabolism
Other Processes

Figure S4. Summary statistics on the *O. sativa* ORFeome.
a, Circle plot depicting the density of 2,300 ORFeome genes (outer circle) and 289 interactome genes (inner circle) across the *O. sativa* genome. Internal arcs represent interacting genes. In general, genes were evenly sampled across the genome. **b,** Comparison of the gene lengths of all annotated rice genes, 2,300 rice ORFeome genes, the initial prioritized set of 3,269 rice genes, and the set of 969 genes for which we failed to obtain a successful clone. Significant differences were observed between all groups (after down sampling to match the smallest group) as ascertained by two-sided Kolmogorov–Smirnov test. **c.** Comparison of the average expression values among the same four groups. The initial prioritized gene set and failed set were significantly more highly expressed compared to the random subset of genes from the whole genome. To a lesser degree, ORFeome genes were significantly more highly expressed than the prioritized gene set and failed set. All p-values were ascertained by Kruskal-Wallis multiple comparison after down-sampling all groups to match the smallest group. **d-f** Fraction of genes in the rice genome, rice ORFeome, and rice interactome that are represented in GO molecular function, cellular component, or biological process categories, respectively.

a



b

Saccharomyces cerevisiae (Yu *et al.* 2008)	*Arabidopsis thaliana* (Arabidopsis Interactome Mapping Consortium 2011)	*Homo sapiens* (Rolland *et al.* 2014)	*Saccharomyces pombe* (Vo *et al.* 2016)	*Oryza sativa* (Wierbowski *et al.* this study)
3,917x5,246 Search Space (20.5 Million Pairs)	8,044x7,771 Search Space (62.5 Million Pairs)	15,517x15,517 Search Space (240.7 Million Pairs)	4,989x4,989 Search Space (24.9 Million Pairs)	1,671x1,671 Search Space (2.80 Million Pairs)
1,809 Interactions Detected	5,664 Interactions Detected	13,944 Interactions Detected	2,278 Interactions Detected	321 Interactions Detected
88.0 Interactions per Million Pairs Screened	90.6 Interactions per Million Pairs Screened	57.9 Interactions per Million Pairs Screened	91.5 Interactions per Million Pairs Screened	115.0 Interactions per Million Pairs Screened

Figure S5. Expanded summary of interaction detection rates among previous high-throughput Y2H interactome screens.
a, Visualization of the number of interactions detected per number of pairwise genes screened across the *S. cerevisiae*, *A. thaliana*, *H. sapiens*, *S. pombe*, and *O. sativa* interactome networks. The red dotted line represents the average detection rate across the five interactomes (88 interactions per million pairs screened). The *O. sativa* star depicts an estimated total number of ~80,000 interactions that could be detected from a completed ORFeome of ~30,000 rice genes. **b,** A complete table comparing the raw size of the search space, number of interactions detected, and interactions detected per million pairs screened across the five interactomes.

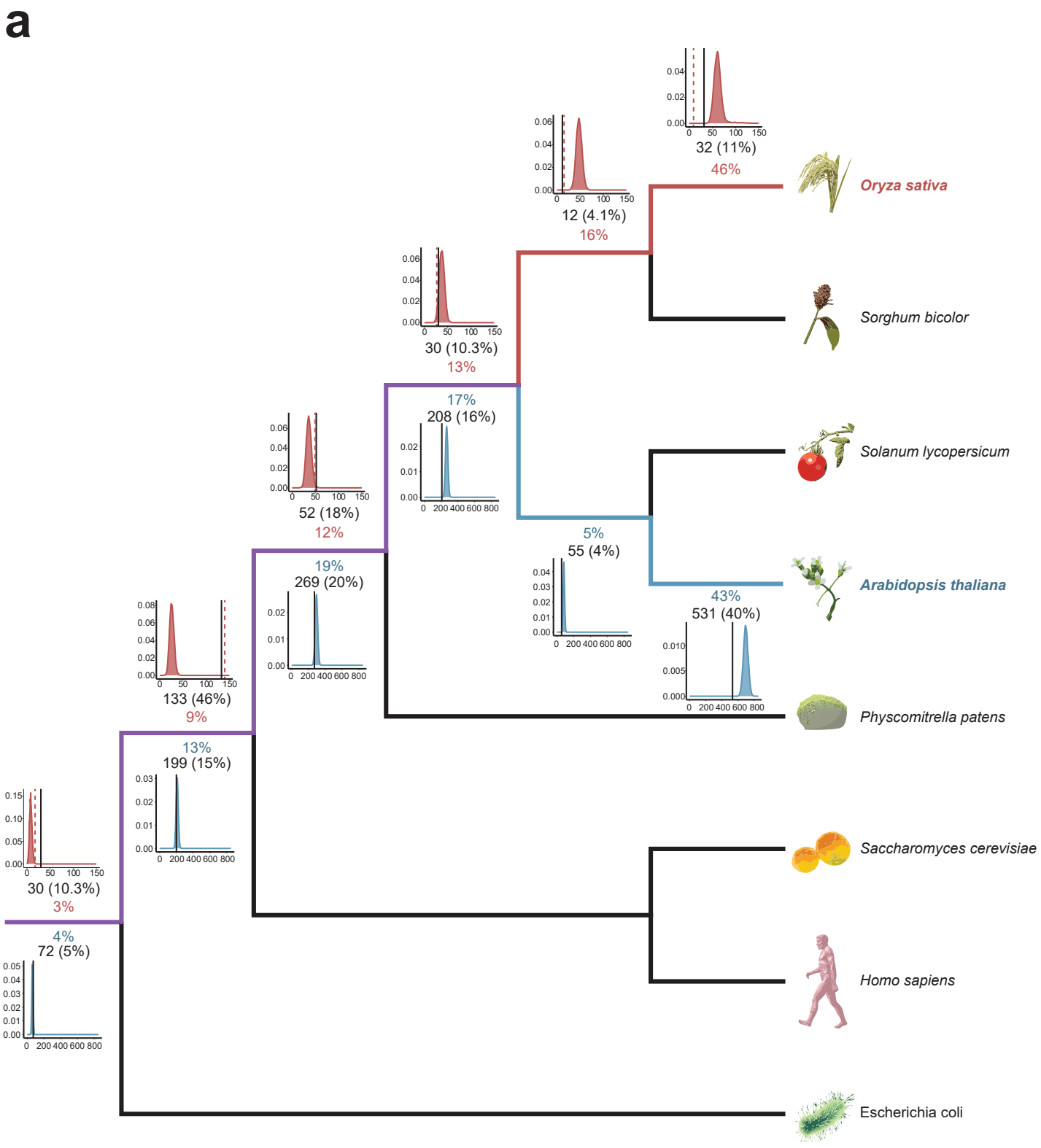

a
30 (10.3%)
3%
4%
72 (5%)
133 (46%)
9%
13%
199 (15%)
52 (18%)
12%
19%
269 (20%)
30 (10.3%)
13%
17%
208 (16%)
12 (4.1%)
16%
5%
55 (4%)
32 (11%)
46%
43%
531 (40%)
Oryza sativa
Sorghum bicolor
Solanum lycopersicum
Arabidopsis thaliana
Physcomitrella patens
Saccharomyces cerevisiae
Homo sapiens
Escherichia coli

Figure S6. Conservation analysis of interacting genes in rice and *Arabidopsis*.
Phylogenetic tree for eight species included in the conservation analysis of interacting genes reported in *Oryza sativa* (top, red) and *Arabidopsis thaliana* (bottom, blue). All interacting genes (n=289 and n=1,334 for *O. sativa* and *A. thaliana* respectively) were binned into the most ancestral node wherein an orthologous gene could be detected among related species. The histograms at each node represent the distributions of 1,000 bootstrap replicates using randomly sampled non-interacting genes. The mean of each distribution is reported by the colored percentage above or below each distribution. The black lines mark the number of interacting genes conserved at each node in the tree. The exact count of interacting genes is reported in black above or below each distribution). The dotted red line (*O. sativa* only) marks the expected number of genes conserved at each node derived from our ORFeome. A black/red line to the left of the distribution indicates under-representation of interacting genes compared to background expectation, while a black/red line to the right indicates over-representation. While a statistically significant difference between the actual count of interacting genes and the mean of random expectation was detected at all nodes ($p < 2.2e-16$ as determined by z-test), the effect size of the difference (absolute [% observed - % expected]) was low in *A. thaliana*, compared to *O. sativa*. This finding indicates that the interacting genes in *A. thaliana* interactome show little to no evolutionary bias, while the *O. sativa* interactome and ORFeome are biased towards more conserved, and hence, more widely distributed genes.

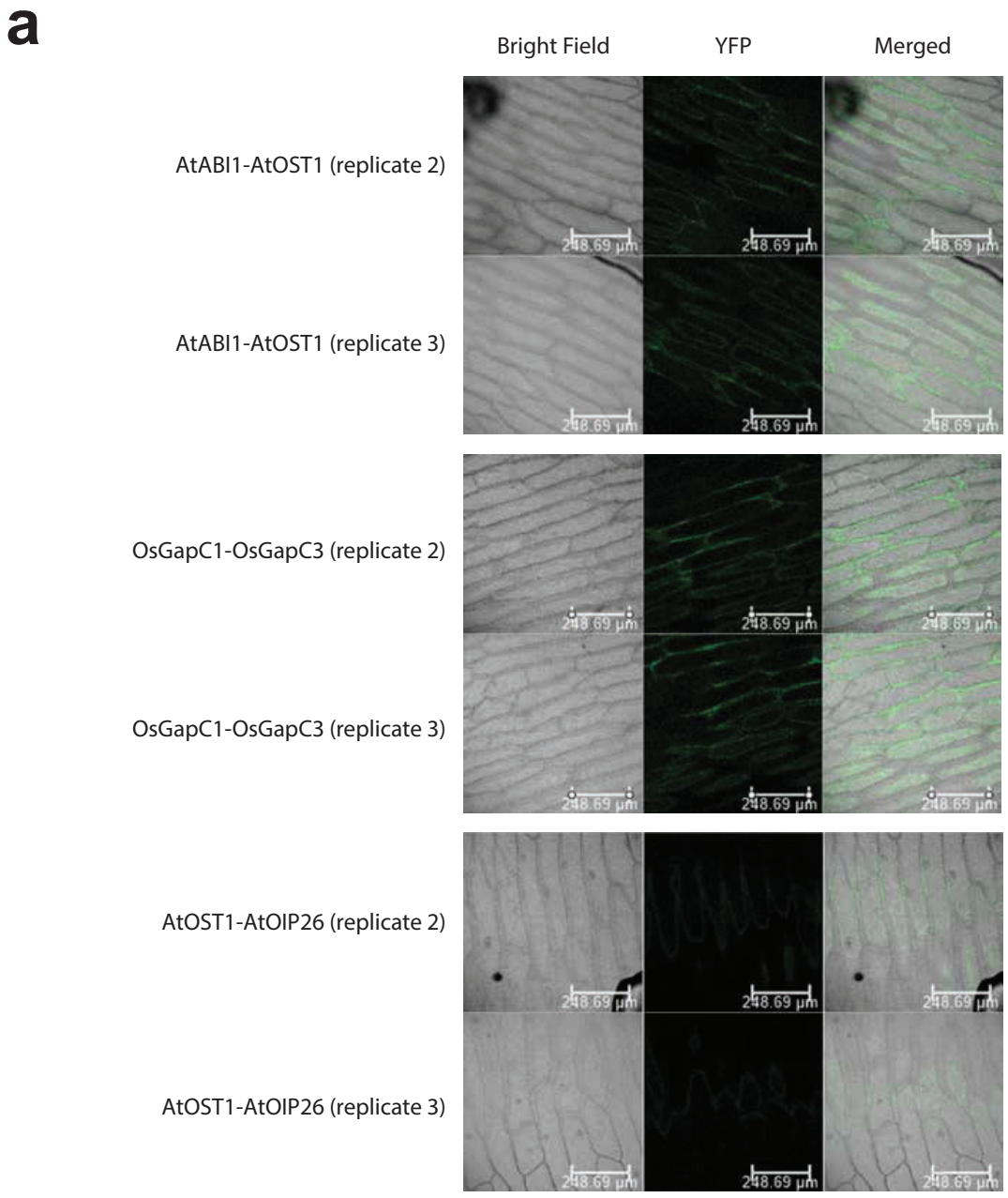


Figure S7. Images of BiFC biological replicates of Fig. 2i.
Confocal fluorescence images of biological replicates for the positive control (AtABI1-AtOST1), for the rice protein pair encoded by LOC_Os08g03290 (OsGapC1) and LOC_Os02g38920 (OsGapC3), and for negative control (AtOST1-AtOIP26).

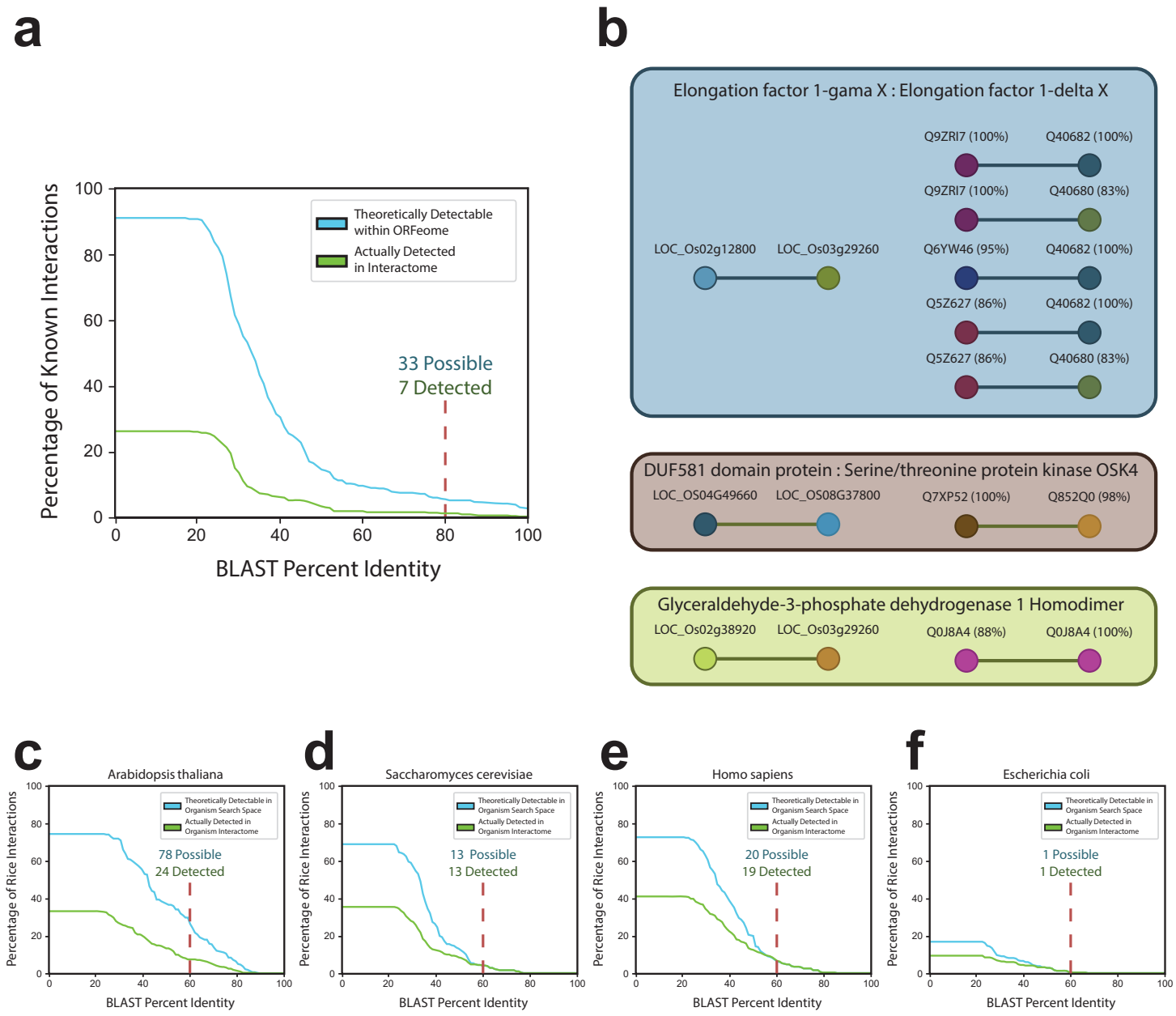


Figure S8. Recapitulation rate of previously reported *O. sativa* interactions by Y2H screen. **a,** All 609 previously reported rice interactions were compared against our ORFeome genes using BLAST to determine the fraction of interactions that could theoretically be recapitulated within our Y2H screen at varying percent identity cutoffs. These BLAST results were intersected with our interactome to determine the fraction of interactions that were actually detected. Using a lenient 80% percent identity 7 out of 33 interactions (21%) were recapitulated by our Y2H screen. **b,** Three of our detected interactions (left) recapitulated seven previously reported interactions (right). Our interactions are labeled using MSU IDs whereas the previously reported interactions are labeled using UniProt IDs. Only the top interaction (Q9ZRI7 and Q40682) is a perfect sequence match against the ORFs used in our interactome, whereas the others are loosely inferred to be close homologs or alternate isoforms. The second interaction (Q7XP52 and Q852Q0) was nearly identical but was originally reported between a related kinase OSK3 rather than OSK4. **c-f,** We repeat this analysis comparing the 322 rice interactions reported here to the binary-high quality interactomes of four additional species (*Arabidopsis thaliana*, *Saccharomyces cerevisiae*, *Homo sapiens*, and *Escherichia coli* respectively). Shown are the percentages of rice interactions with homologous interactions reported in the comparison interactome map at various percent identity cutoffs (green lines) and the percentage that could have theoretically been recapitulated—i.e. rice interactions with homologs for both genes in the comparison interactome regardless of if the interaction was reported (blue lines).