

A whole-genome shotgun assembly for genome characterization of the common ice plant (*Mesembryanthemum crystallinum* L.) – Supplementary Information

Ryoma Sato¹, Yuri Kondo¹, Sakae Agarie^{2*}

Affiliations:

¹Graduate school of Bioresource and Bioenvironmental Sciences, Kyushu University, 744 Motooka Nishi-ku Fukuoka, 819-0395, Japan

²Faculty of Agriculture, Kyushu University, 744 Motooka Nishi-ku Fukuoka, 819-0395, Japan

*Correspondence: agarie@agr.kyushu-u.ac.jp

Supplementary Information

Supplementary Notes 1 (Address to genome information) and Note 2 (Address to web sites), Supplementary Tables 1-5, Supplementary Figs. 1–8, and Supplementary references.

Supplementary Dataset

Supplementary Excel formatted datasets 1-3.

Github repository

All supplementary data are archived in the github site

(https://github.com/Shokusei/Sato_et_al_2022_supplementary_data).

Supplementary Notes

Supplementary Note 1. Addresses to genome information.

Kewa caespitosa:

https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900/322/205/GCA_900322205.1_ASM90032220v1/GCA_900322205.1_ASM90032220v1_genomic.fna.gz

Macarthuria australis:

https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900/322/265/GCA_900322265.1_ASM90032226v1/GCA_900322265.1_ASM90032226v1_genomic.fna.gz

Pharnaceum exiguum:

https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900/322/385/GCA_900322385.1_ASM90032238v1/GCA_900322385.1_ASM90032238v1_genomic.fna.gz

Populus trichocarpa:

https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/775/GCF_000002775.4_Pop_tri_v3/GCF_000002775.4_Pop_tri_v3_genomic.fna.gz

Solanum chaucha:

https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002/312/845/GCA_002312845.1_K_fedtschenkoi_M2_v1/GCA_002312845.1_K_fedtschenkoi_M2_v1_genomic.fna.gz

Arabidopsis thaliana:

https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/001/735/GCF_000001735.4_TAIR10.1/GCF_000001735.4_TAIR10.1_genomic.fna.gz

Oryza sativa:

https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/433/935/GCF_001433935.1_IRGSP-1.0/GCF_001433935.1_IRGSP-1.0_genomic.fna.gz

Supplementary Note 2. Addresses to web sites.

eHALOPH database: <https://www.sussex.ac.uk/affiliates/halophytes/>

REAPR: <https://www.sanger.ac.uk/tool/reapr/>

Jellyfish: <https://github.com/gmarcais/Jellyfish>

GenomeScope2.0: <https://github.com/tbenavi1/genomescope2.0>

Musket: <http://musket.sourceforge.net/homepage.htm>

ALGA: <https://github.com/swacisko/ALGA>

Redundans: <https://github.com/lpryszcz/redundans.git>

Fastqc: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

gVolante: <https://gvolante.riken.jp/>

BUSCOv5: <https://gitlab.com/ezlab/busco/-/releases#5.2.1>

RepeatModeler2: <https://github.com/Dfam-consortium/RepeatModeler>

Repbase: <https://www.girinst.org/>

TEclass: <http://www.compgen.uni-muenster.de/teclass>

RepeatMasker: <https://github.com/rmhubble/RepeatMasker>

tRNAscan-SE2.0: <http://trna.ucsc.edu/tRNAscan-SE/>

PlantRNA: <https://seve.ibmp.unistra.fr/plantrna/>

infernai: <https://github.com/EddyRivasLab/infernai>

BRAKER2: <https://github.com/Gaius-Augustus/BRAKER#braker-with-rna-seq-data>

DIAMOND: <https://github.com/bbuchfink/diamond>

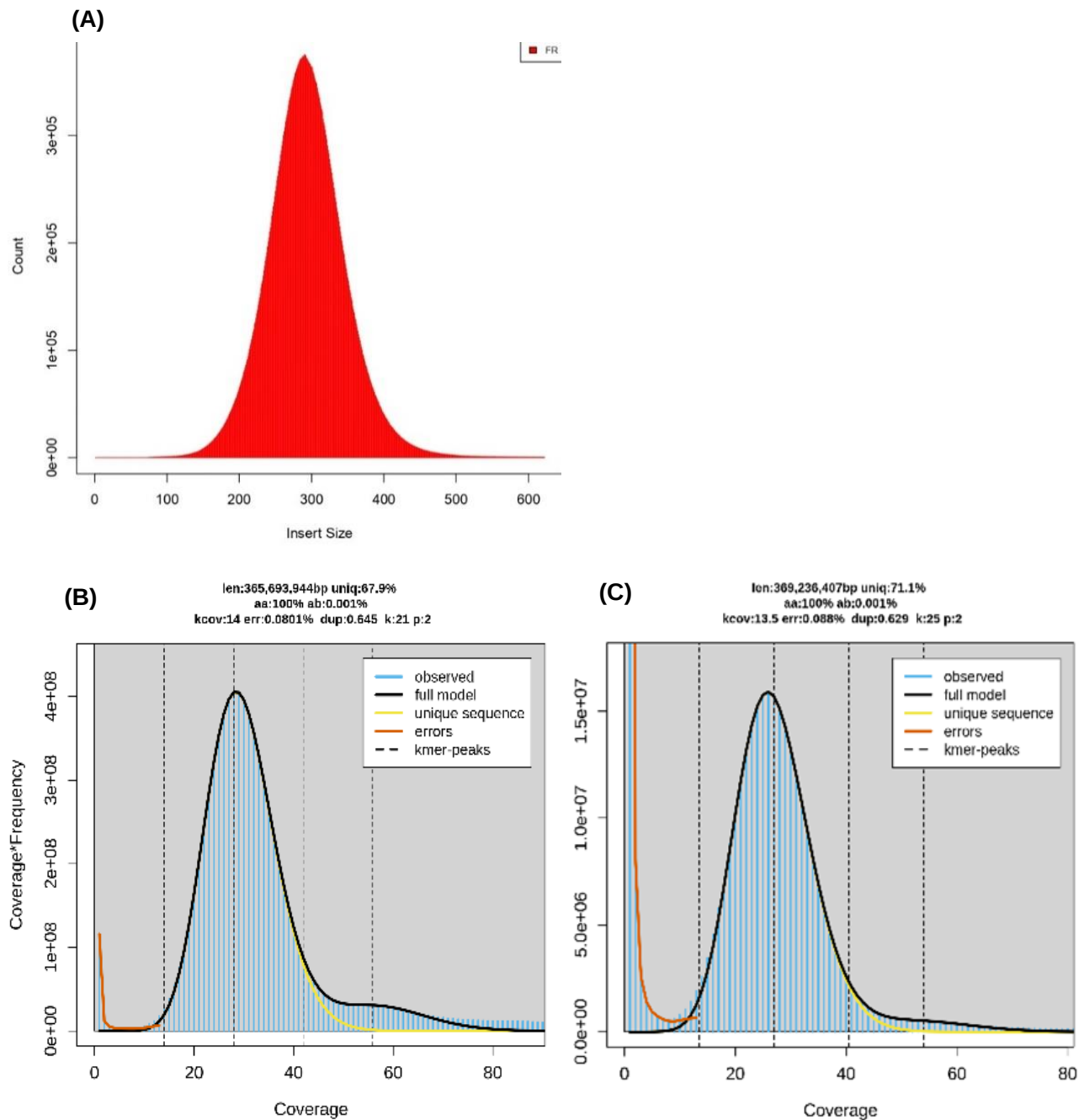
HMMER: <https://github.com/EddyRivasLab/hmmer>

iTAK: <http://itak.feilab.net/cgi-bin/itak/index.cgi>

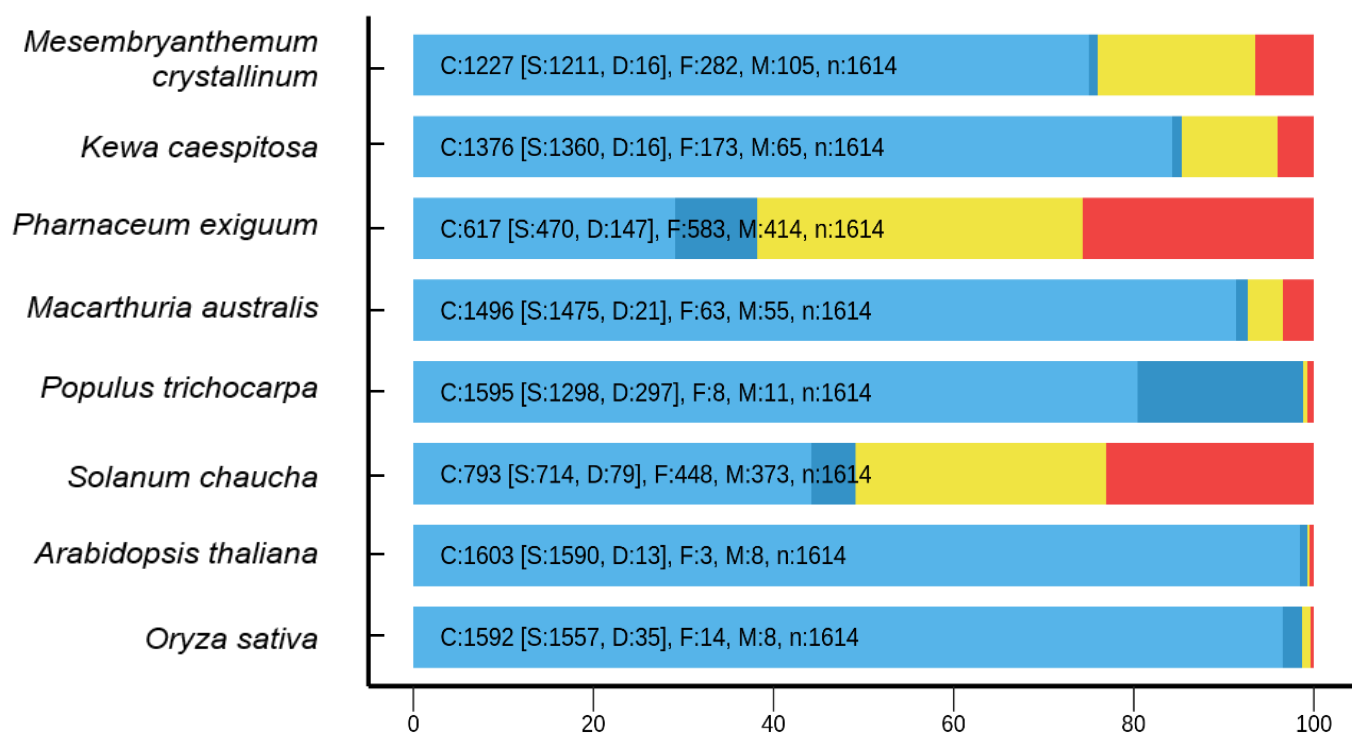
DAVID: <https://david.ncifcrf.gov/>

PGDBj: <http://pgdbj.jp/plantdb/plantgenome.html?ln=ja>

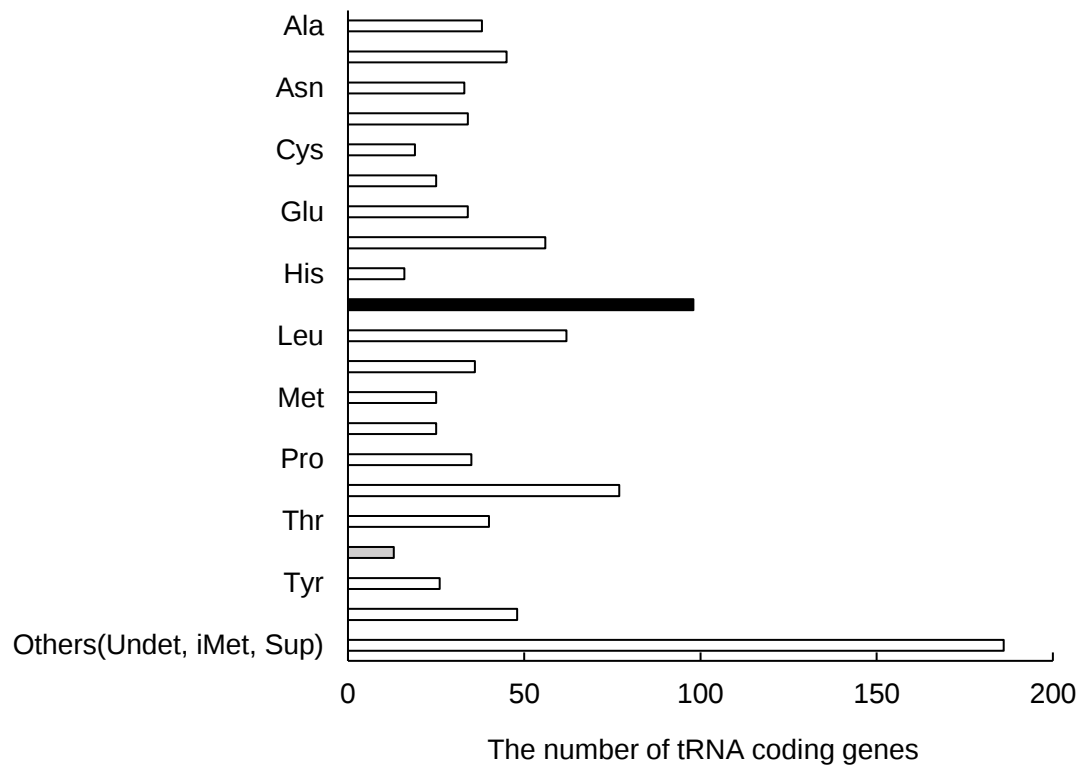
Supplementary Figures



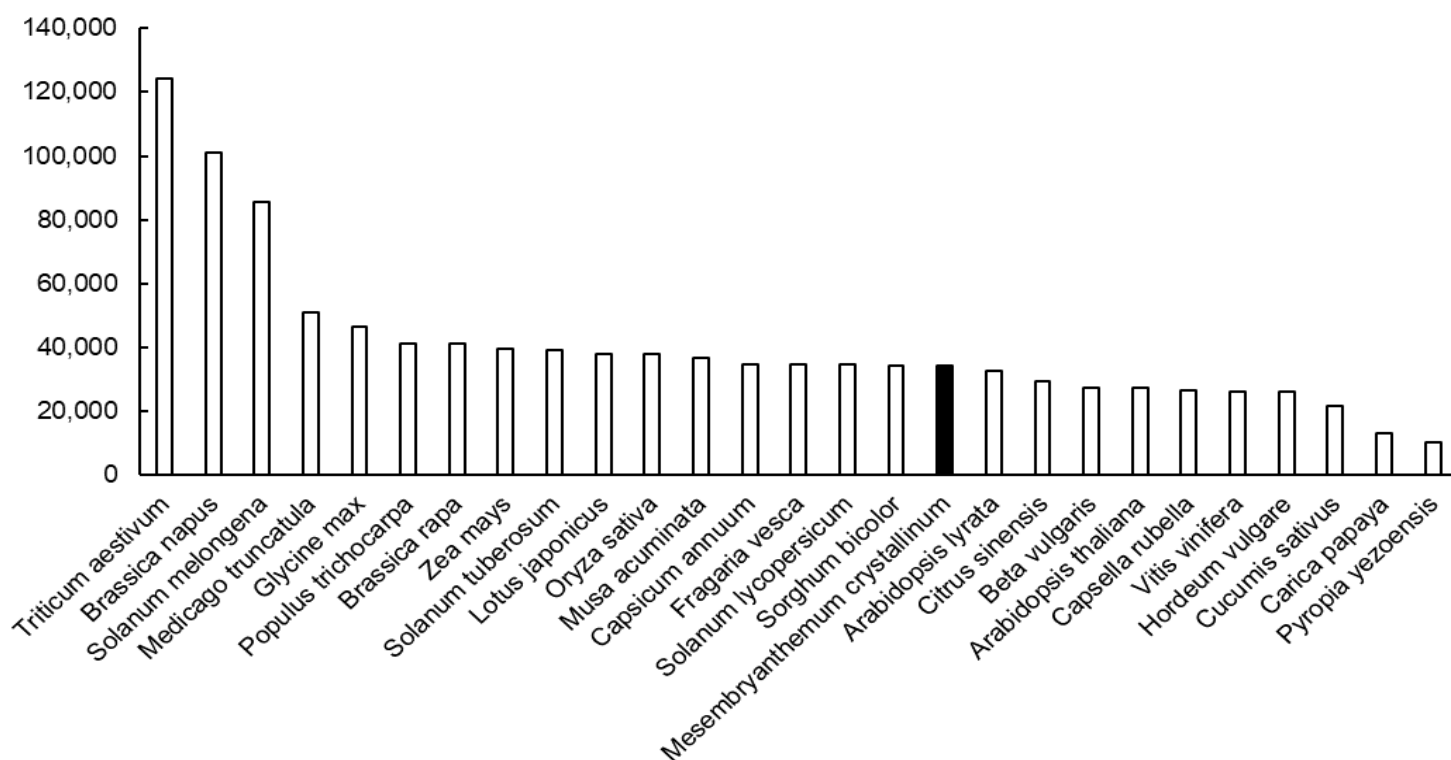
Supplementary Figure S1. (A): Insert size distribution in the Illumina pair-end raw reads of *Mesembryanthemum crystallinum* genome using REAPR (key parameter: smaltmap). (B) and (C): GenomeScope2 k -mer [21 (B) and 25 (C)] distribution showing estimation of genome size (len), homozygosity (aa), heterozygosity (ab), mean k -mer coverage for heterozygous bases (kcov), read error rate (err), the average rate of read duplications (dup), k -mer size used on the run (k), and ploidy (p).



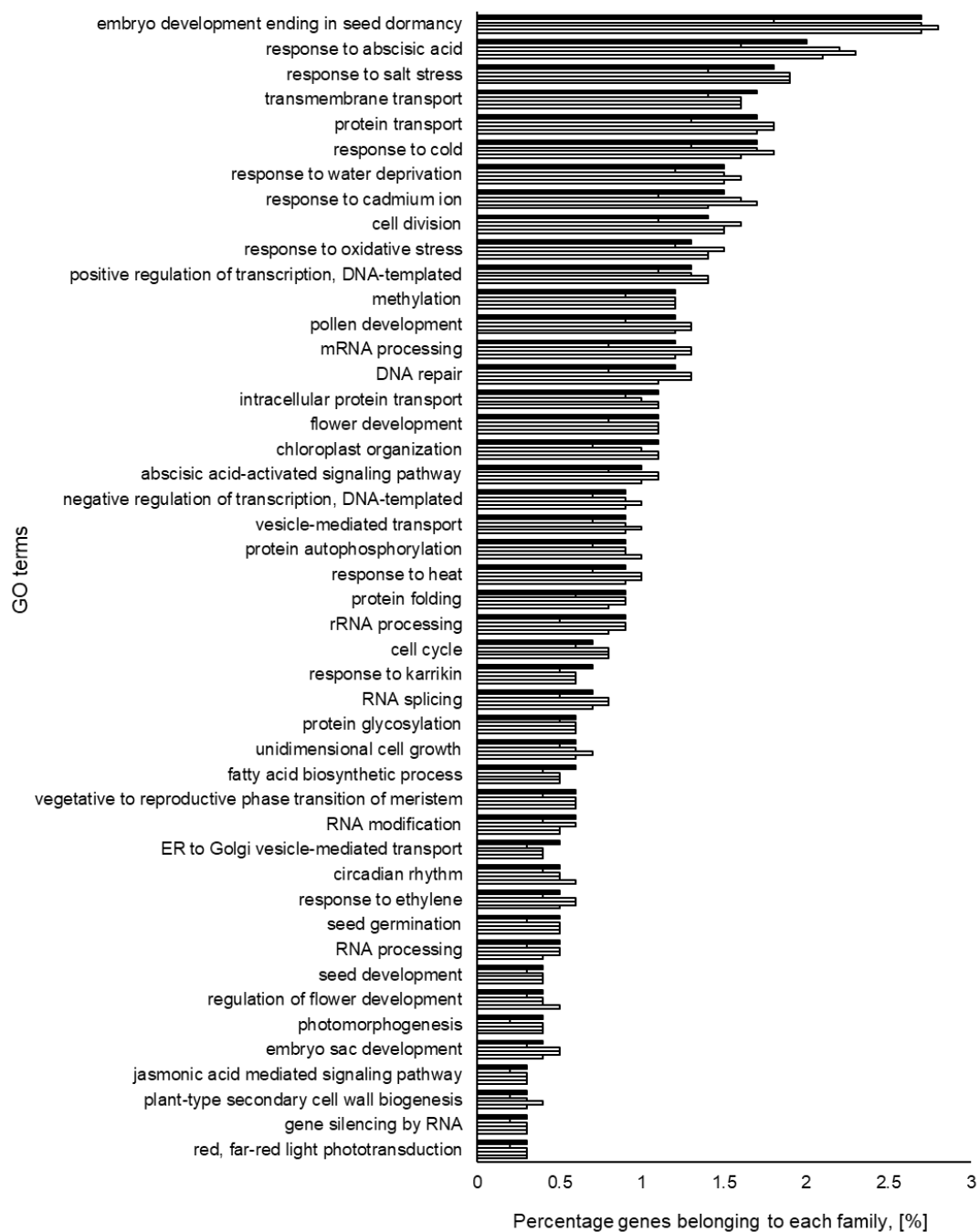
Supplementary Figure S2. Genome sequence integrity verification results by the BUSCO program. This figure was generated by the following command (generate_plot.py -wd BUSCO_summary/). Light blue: Complete (Single-Copy), blue: Complete (Duplicated), yellow: Fragmented, red: Missing.



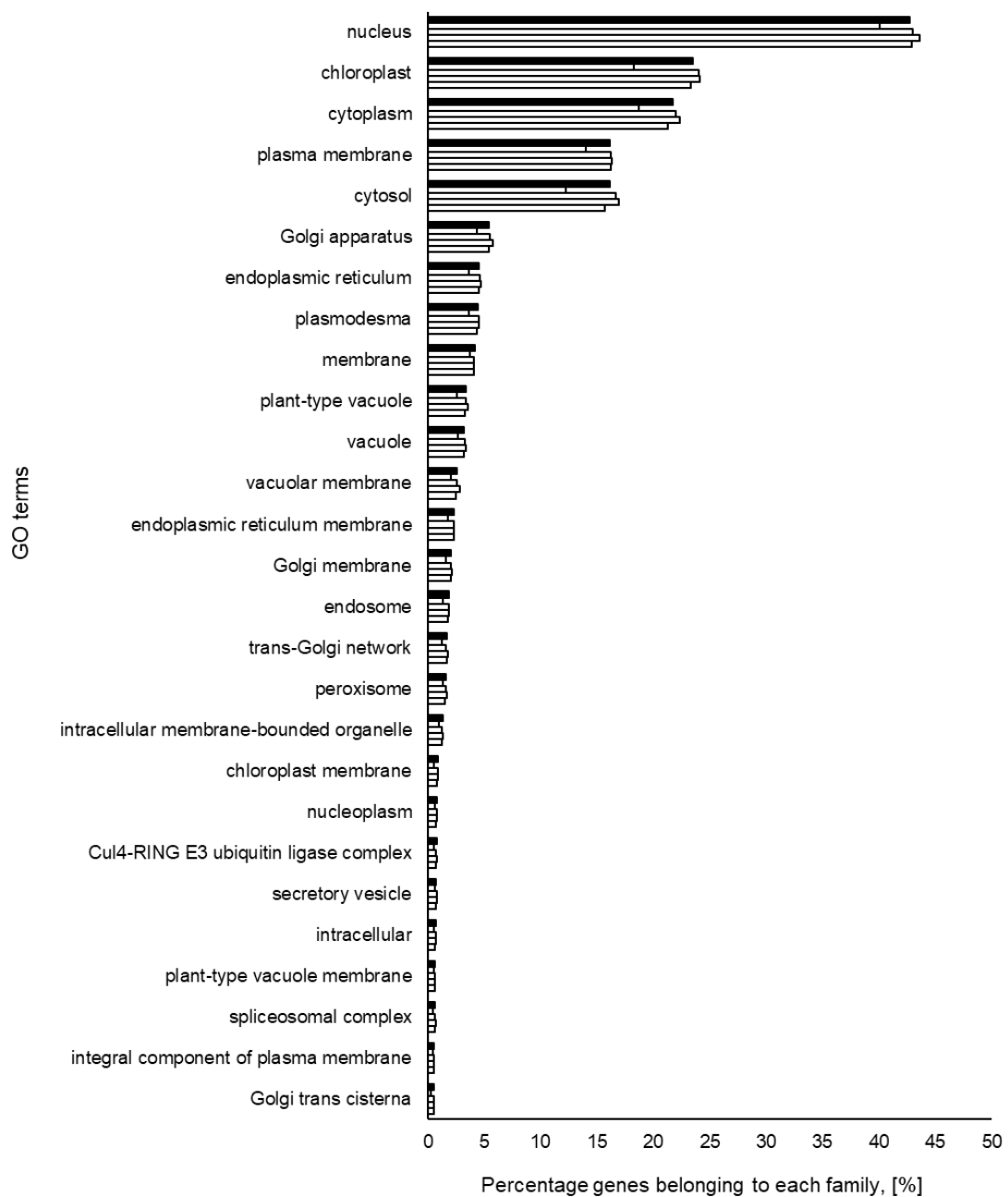
Supplementary Figure S3. The ice plant tRNA distribution. The tRNA sequences were extracted from the ice plant genome sequence using tRNA-SE2.0 and classified according to the codons specified. The tRNA-isoleucine (Ile) with the highest number is shown in black, tRNA-tryptophan (Trp) with the lowest number is shown in gray, and others are shown in white.



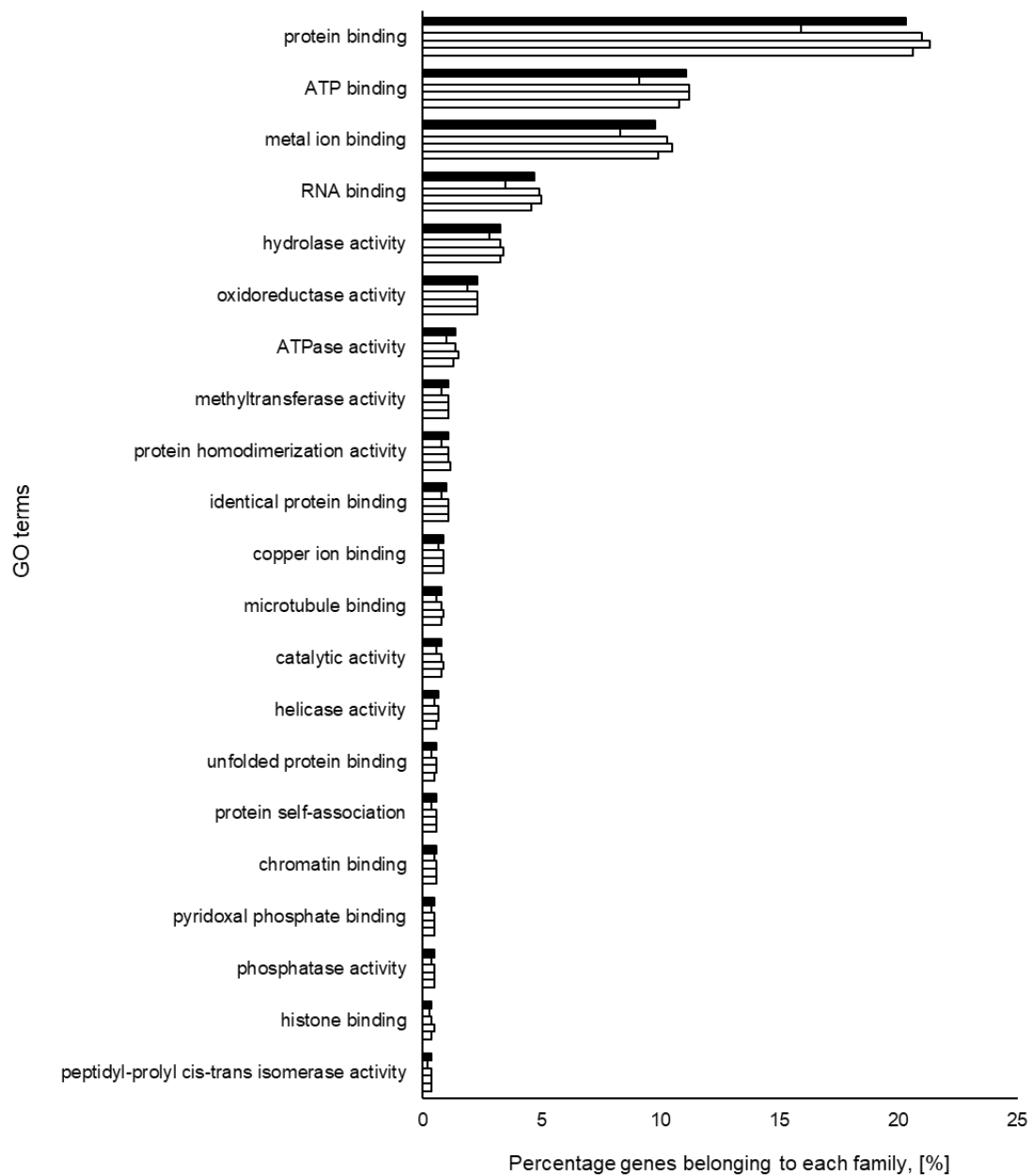
Supplementary Figure S4. Predicted gene counts for 27 plant species. Ice plant genes number are shown as a black bar.



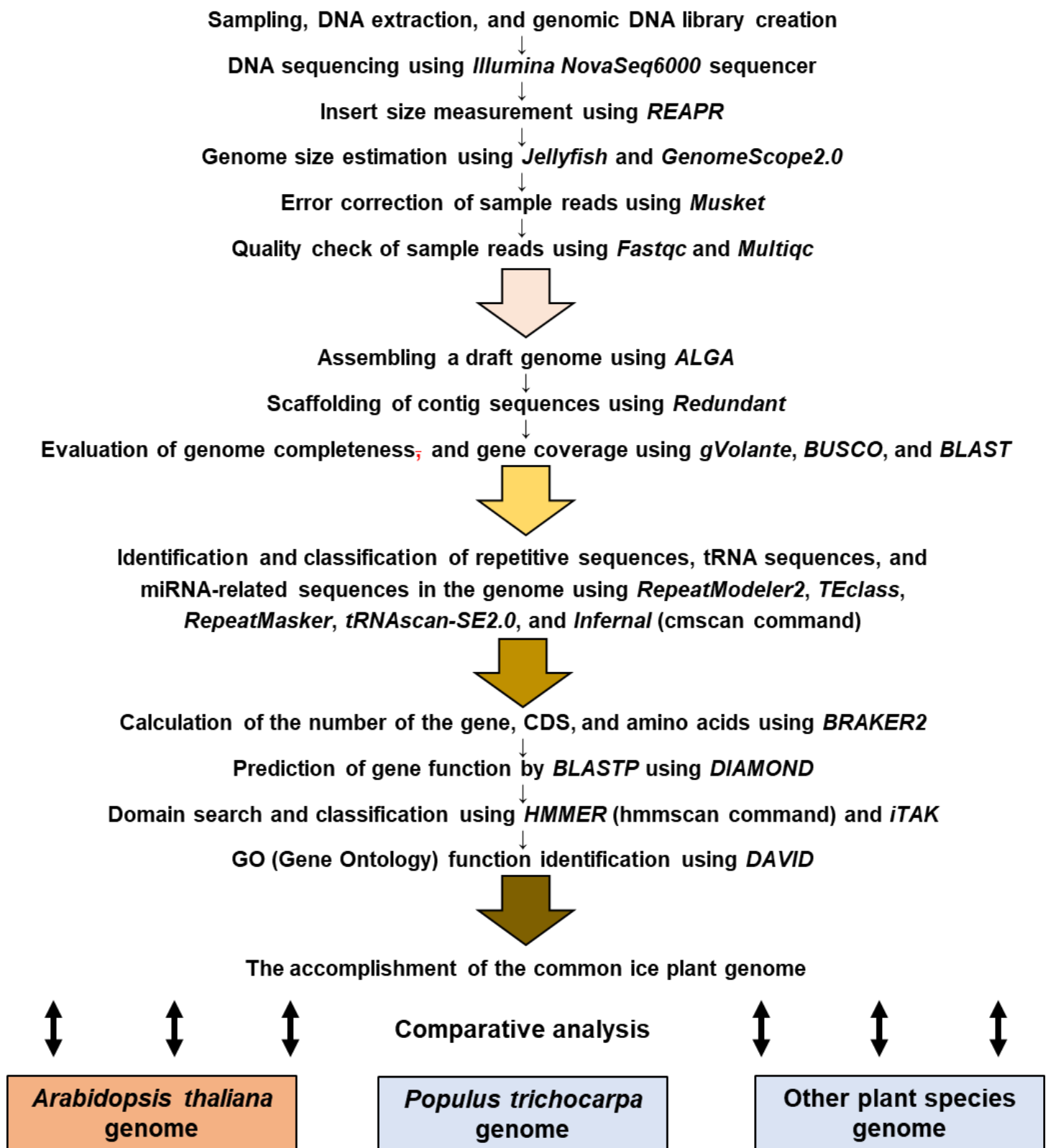
Supplementary Figure S5. Ratio of gene numbers belonging to the biological process (BP) to the total number of genes in five plant species including ice plant. gene ontology (GO) terms present in all plants were selected. The GO terms are, from the top, ice plant, *Arabidopsis*, rice, maize, and poplar. The ice plant is shown in black, and other GO terms are shown in white. Vertical axis indicates GO function name and horizontal axis indicates the percentage.



Supplementary Figure S6. Ratio of gene numbers belonging to the cellular component (CC) to the total number of genes in five plant species including ice plant. GO terms present in all plants were selected. The GO terms are, from the top, ice plant, *Arabidopsis*, rice, maize, and poplar. The ice plant is shown in black, and other GO terms are shown in white. Vertical axis indicates GO function name and horizontal axis indicates the percentage.



Supplementary Figure S7. Ratio of gene numbers belonging to the molecular function (MF) to the total number of genes in five plant species including ice plant. GO terms present in all plants were selected. The GO terms are, from the top, ice plant, *Arabidopsis*, rice, maize, and poplar. The ice plant is shown in black, and other GO terms are shown in white. Vertical axis indicates GO function name and horizontal axis indicates the percentage.



Supplementary Figure S8. Flowchart of *de novo* genome assembly and subsequent analysis. The names of software and web tools used, in this study, are indicated in italics.

Supplementary Tables

Supplementary Table S1. Genome sequencing results.

Reads	Read from 5' end (Read1)	Read from 3' end (Read2)	Total bases
Unique reads (bp)	32,999,186	33,536,575	66,535,761
Duplicate reads (bp)	20,661,762	20,124,373	40,786,135
Total bases	53,660,948	53,660,948	107,321,896

Paired-end sequencing was performed and trimmed with musket. Unique reads, duplicate reads, and total number of reads are shown.

Supplementary Table S2. Assembled and scaffolded genome statistics.

Process	Assembling	Scaffolding
Software programs	ALGA (v1.0.3)	Redundans (v0.14a)
Total length [nt]	341,841,913	286,008,514
Total sequences	161,183	49,782
GC contents [%]	37.03	36.59
Number of sequences > 1 K (nt)	66,037	40,499
Number of sequences > 10 K (nt)	6,248	8,602
N50 [nt]	5,736	10,562
Number of gaps (>=5 N's)	0	17,358
Longest sequence [nt]	59,958	90,422

Supplementary Table S3. Comparison of the percentage of tRNAs among the ice plant and the other 8 plant species by Smirnov-Grabs outlier test.

Amino acids	Percentage of tRNAs in 9 plant species (%)									Average (%)	T-value	P-value
	<i>M. crystallinum</i>	<i>A. thaliana</i>	<i>O. sativa</i>	<i>S. lycopersicum</i>	<i>P. trichocarpa</i>	<i>E. japonicum</i>	<i>S. tuberosum</i>	<i>Vitis spp.</i>	<i>G. max</i>			
Ala	3.91	5.64	7.92	6.22	7.33	4.80	6.25	9.97	7.00	6.56	-1.57	0.11
Arg	4.63	5.98	7.92	6.78	7.72	5.80	6.25	6.91	8.14	6.68	-1.90	0.07
Asn	3.40	2.56	2.97	3.04	3.56	3.00	2.58	3.07	3.14	3.04	1.17	0.19
Cys	1.96	2.22	1.78	1.94	1.98	2.00	1.90	2.05	2.00	1.98	-0.21	0.37
Gln	2.57	2.74	4.55	3.32	3.56	2.80	3.13	3.84	3.57	3.34	-1.30	0.16
Glu	3.50	3.93	7.13	6.09	6.34	4.40	6.39	5.88	5.71	5.49	-1.69	0.10
Gly	5.77	6.32	7.92	8.85	7.72	6.20	9.92	7.93	8.00	7.63	-1.48	0.13
Ile	10.09	3.76	4.55	4.56	4.75	3.60	4.35	4.35	4.57	4.95	2.77	< 0.05
Leu	6.39	6.67	8.32	8.58	8.12	6.80	7.88	8.18	7.71	7.63	-1.64	0.10
Lys	3.71	5.13	5.35	5.67	5.94	4.80	5.57	4.35	5.86	5.15	-2.05	0.06
Phe	2.57	2.56	2.77	3.46	3.37	2.60	3.40	2.81	2.86	2.93	-1.03	0.22
Ser	7.93	9.40	6.34	7.05	6.93	8.40	6.79	7.16	7.57	7.51	0.47	0.34
Thr	4.12	3.76	4.36	4.43	2.77	3.80	3.80	4.09	3.29	3.82	0.60	0.32
Trp	1.34	2.22	2.57	2.21	2.57	1.80	2.45	2.56	2.43	2.24	-2.28	< 0.05
Tyr	2.68	11.97	2.57	3.60	3.17	7.60	3.67	2.05	2.86	4.46	-0.58	0.32

The tRNA database of eight plant species, *A. thaliana*, *O. sativa*, *S. lycopersicum*, *P. trichocarpa*, *E. japonicum*, *S. tuberosum*, *Vitis spp.*, *G. max* were obtained from the PlantRNA database¹.

Supplementary Table S4. Statistics on coding sequences (CDS) and amino acid sequences predicted from the ice plant genome.

Regions	Whole genome	Genes	CDS	Amino acids
Sequences	49,782	34,223	35,702	35,702
Total bases, [nt] or amino acids, [aa]	286,008,514	91,283,320	30,364,506	9,958,616
Average length, [nt/aa]	5,745.20	2,667.30	1,185	411.1
Median length, [nt/aa]	3,357	1,761	950	330
N50, [nt/aa]	10,681	4,346	1,511	509

Supplementary Table S5. MiRNA loci derived from the ice plant genome sequence and the target mRNA functions.

MiRNA loci	Gene or precursor	Number of loci	Target mRNA and its function	References
MIR169	gene	25	Nuclear factor γ subunit A (NF-YA)	2
mir-399	precursor	16	PHO2: ubiquitin-conjugating enzyme	3
MIR159	gene	8	MYB33, MYB65	4
mir-166	precursor	7	Homeobox-leucine zipper family protein (HD-ZIP)	2
mir-598	precursor	4	#N/A	#N/A
mir-395	precursor	4	WRKY transcription factor	5
mir-172	precursor	4	APETALA2 (AP2) and AP2-like transcription factors	6
MIR164	gene	4	NAC domain-containing proteins	2
MIR398	gene	3	Blue copper-binding protein (BCBP)	7
MIR167	gene	3	Auxin response factors (ARFs) and IAA-Ala resistant3 (IAR3)	8
mir-160	precursor	3	Auxin response factor (ARF)16	2
MIR811	gene	2	HLH DNA-binding domain-containing proteins	9
MIR408	gene	2	Blue copper (Cu) proteins	10
MIR397	gene	2	Laccase (LAC) genes involved in lignin synthesis	11
mir-393	precursor	2	L-type lectin receptor kinase V.7	12
MIR390	gene	2	SiRNA transacting against auxin response factor gene	13
MIR162	gene	2	Dicerlike-1	2
MIR535	gene	1	SQUAMOSA promoter binding protein-like 14 (OsSPL14)	14
MIR403	gene	1	AGO(Argonaute)2 and AGO3	15
MIR396	gene	1	Squamosa promoter-binding protein-like (SPL)	2
MIR394	gene	1	Growth-regulating factor (GRF)2	2
MIR319	gene	1	MYB domain protein 33	2
MIR168	gene	1	Argonaute1	2
mir-156	precursor	1	Squamosa promoter-binding protein-like (SPL)	2

Supplementary Table S6. Gene ontology (GO) terms and number of genes belonging to ice plant considered to be more beneficial in ice plant than in *Arabidopsis*, rice, maize, and poplar.

GO	GO terms	Number of Gene Reference ⁽¹⁾	
Biological Process	Response to virus	33	16
	Pollen tube development	27	17
	Triglyceride biosynthetic process	21	18
	Wax biosynthetic process	21	19
	Auxin transport	14	20
	DNA topological change	12	21,22
	Nucleic acid phosphodiester bond hydrolysis	12	23
	Production of siRNA involved in chromatin silencing by small RNA	10	24,25
	Photoinhibition	10	26,27
	Methylation-dependent chromatin silencing	8	28
	Mitotic chromosome condensation	8	29,30
Cellular Component	Cytoplasmic vesicle	38	31
	SNARE complex	35	32
	Protein storage vacuole	14	33
	O-acetyltransferase activity	34	34
Molecular Function	SNAP receptor activity	34	32,35
	Nucleotidyl transferase activity	29	36
	Enoyl-CoA hydratase activity	9	37

⁽¹⁾ Results of investigation on whether these gene functions are related to halophilism and salt tolerance (photosynthetic conversion) are stated.

Supplementary References

1. Cognat, V. *et al.* PlantRNA, a database for tRNAs of photosynthetic eukaryotes. *Nucleic Acids Res* **41**, 273–279 (2013).
2. Chiang, C. P. *et al.* Identification of ice plant (*Mesembryanthemum crystallinum* L.) microRNAs using RNA-seq and their putative roles in high salinity responses in seedlings. *Front Plant Sci* **7**, 1143 (2016).
3. Chiou, T. J. *et al.* Regulation of phosphate homeostasis by microRNA in *Arabidopsis*. *Plant Cell* **18**, 412–421 (2006).
4. Zheng, Z. *et al.* Target RNA secondary structure is a major determinant of miR159 efficacy. *Plant Physiol* **174**, 1764–1778 (2017).
5. Zhang, Q. *et al.* Md-MiR156ab and Md-mir395 target WRKY transcription factors to influence apple resistance to leaf spot disease. *Front Plant Sci* **8**, 1–14 (2017).
6. Jung, J. H., Lee, S., Yun, J., Lee, M. & Park, C. M. The miR172 target *TOE3* represses *AGAMOUS* expression during *Arabidopsis* floral patterning. *Plant Science* **215–216**, 29–38 (2014).
7. Brousse, C. *et al.* A non-canonical plant microRNA target site. *Nucleic Acids Res* **42**, 5270–5279 (2014).
8. Liu, X., Huang, S. & Xie, H. Advances in the regulation of plant development and stress response by miR167. *Frontiers in Bioscience - Landmark* **26**, 655–665 (2021).
9. Wu, F., Shu, J. & Jin, W. Identification and validation of miRNAs associated with the resistance of maize (*Zea mays* L.) to *Exserohilum turcicum*. *PLoS One* **9**, e87251 (2014).
10. Gao, Y. *et al.* The evolution and functional roles of miR408 and its targets in plants. *Int J Mol Sci* **23**, 530 (2022).

11. Huang, S., Zhou, J., Gao, L. & Tang, Y. Plant miR397 and its functions. *Functional Plant Biology* **48**, 361–370 (2021).
12. Djami-Tchatchou, A. T. & Dubery, I. A. MiR393 regulation of lectin receptor-like kinases associated with LPS perception in *Arabidopsis thaliana*. *Biochem Biophys Res Commun* **513**, 88–92 (2019).
13. Wen, F. L. *et al.* Identification of miR390-TAS3-ARF pathway in response to salt stress in *Helianthus tuberosus* L. *Gene* **738**, 144460 (2020).
14. Zhang, L. L. *et al.* Osa-miR535 targets SQUAMOSA promoter binding protein-like 4 to regulate blast disease resistance in rice. *Plant Journal* **110**, 166–178 (2022).
15. Jagtap, S. & Shivaprasad, P. V. Diversity, expression and mRNA targeting abilities of Argonaute-targeting miRNAs among selected vascular plants. *BMC Genomics* **15**, 1049 (2014).
16. Aguilar, E. & Lozano-Duran, R. Plant viruses as probes to engineer tolerance to abiotic stress in crops. *Stress Biology* letter to the editor (2022)
17. Guo, J. & Wang, B. Blocked pollen release prevents fruit formation in the halophyte *Elaeagnus angustifolia* in non-saline habitats. *Environ Exp Bot* **189**, 104566 (2021).
18. Lu, J., Xu, Y., Wang, J., Singer, S. D. & Chen, G. The role of triacylglycerol in plant stress response. *Plants* **9**, 472 (2020).
19. Xue, D., Zhang, X., Lu, X., Chen, G. & Chen, Z. H. Molecular and evolutionary mechanisms of cuticular wax for plant drought tolerance. *Front Plant Sci* **8**, 621 (2017).
20. Korver, R. A., Koevoets, I. T. & Testerink, C. Out of shape during stress: a key role for auxin. *Trends Plant Sci* **23**, 783–793 (2018).
21. Hettiarachchi, G. H. C. M., Reddy, M. K., Sopory, S. K. & Chattopadhyay, S. Regulation of TOP2 by various abiotic stresses including cold and salinity in pea and

- transgenic tobacco plants. *Plant Cell Physiol* **46**, 1154–1160 (2005).
22. John, R. *et al.* Over-expression of *topoisomerase II* enhances salt stress tolerance in tobacco. *Front Plant Sci* **7**, 1280 (2016).
 23. Diaz-Baena, M., Galvez-Valdivieso, G., Delgado-Garcia, E., Pineda, M. & Piedras, P. Nuclease and ribonuclease activities in response to salt stress: Identification of *PvRNS3*, a T2/S-like ribonuclease induced in common bean radicles by salt stress. *Plant Physiology and Biochemistry* **147**, 235–241 (2020).
 24. Borsani, O., Zhu, J., Verslues, P. E., Sunkar, R. & Zhu, J. K. Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell* **123**, 1279–1291 (2005).
 25. Swiezewski, S. *et al.* Small RNA-mediated chromatin silencing directed to the 3' region of the *Arabidopsis* gene encoding the developmental regulator, *FLC*. *Proc Natl Acad Sci U S A* **104**, 3633–3638 (2007).
 26. Yan, K., Wu, C., Zhang, L. & Chen, X. Contrasting photosynthesis and photoinhibition in tetraploid and its autodiploid honeysuckle (*Lonicera japonica* thunb.) under salt stress. *Front Plant Sci* **6**, 227 (2015).
 27. Yan, K. *et al.* Salt adaptability in a halophytic soybean (*Glycine soja*) involves photosystems coordination. *BMC Plant Biol* **20**, 155 (2020).
 28. Singroha, G., Kumar, S., Gupta, O. P., Singh, G. P. & Sharma, P. Uncovering the epigenetic marks involved in mediating salt stress tolerance in plants. *Front Genet* **13**, 811732 (2022).
 29. Younesi-Melerdi, E., Nematzadeh, G. A., Pakdin-Parizi, A., Bakhtiarizadeh, M. R. & Motahari, S. A. *De novo* RNA sequencing analysis of *Aeluropus litoralis* halophyte plant under salinity stress. *Sci Rep* **10**, 9148 (2020).

30. Baranova, E. N. & Gulevich, A. A. Asymmetry of plant cell divisions under salt stress. *Symmetry (Basel)* **13**, 1811 (2021).
31. Baral, A., Shruthi, K. S. & Mathew, M. K. Vesicular trafficking and salinity responses in plants. *IUBMB Life* **67**, 677–686 (2015).
32. Kwon, C., Lee, J. H. & Yun, H. S. SNAREs in plant biotic and abiotic stress responses. *Mol Cells* **43**, 501–508 (2020).
33. Neves, J. *et al.* Abiotic stress triggers the expression of genes involved in protein storage vacuole and exocyst-mediated routes. *Int J Mol Sci* **22**, 10644 (2021).
34. Mulet, J. M., Alemany, B., Ros, R., Calvete, J. J. & Serrano, R. Expression of a plant serine O-acetyltransferase in *Saccharomyces cerevisiae* confers osmotic tolerance and creates an alternative pathway for cysteine biosynthesis. *Yeast* **21**, 303–312 (2004).
35. Hossain, M. R., Bassel, G. W., Pritchard, J., Sharma, G. P. & Ford-Lloyd, B. V. Trait specific expression profiling of salt stress responsive genes in diverse rice genotypes as determined by modified significance analysis of microarrays. *Front Plant Sci* **7**, 567 (2016).
36. Yang, H. *et al.* Identification and expression profiling of *Oryza sativa* nucleotidyl transferase protein (*NTP*) genes under various stress conditions. *Gene* **628**, 93–102 (2017).
37. Fukuda, N. *et al.* *ECERIFERUM 10* encoding an enoyl-CoA reductase plays a crucial role in osmotolerance and cuticular wax loading in *Arabidopsis*. *Front Plant Sci* **13**, 898317 (2022).