Supporting Information Appendix

Insights into Salt Tolerance from the Genome of *Thellungiella salsuginea*

Hua-Jun Wu[1]*, Zhonghui Zhang[1]*, Jun-Yi Wang[2]*, Dong-Ha Oh[4]*, Maheshi Dassanayake[4]*, Binghang Liu[2]*, Quanfei Huang[2]*, Hai-Xi Sun[1], Ran Xia[1], Yaorong Wu[1], Yinan Wang[1], Zhao Yang[1], Yang Liu[1], Wanke Zhang[1], Huawei Zhang[1], Jinfang Chu[1], Cunyu Yan[1], Shuang Fang[1], Jinsong Zhang[1], Yiqin Wang[1], Fengxia Zhang[1], Guodong Wang[1], Sang Yeol Lee[5], John M Cheeseman[4], Bicheng Yang[2], Bo Li[2], Jiumeng Min[2], Linfeng Yang[2], Jun Wang[2,†], Chengcai Chu[1,†], Shou-Yi Chen[1,†], Hans J Bohnert[4,5], Jian-kang Zhu[3,†], Xiu-Jie Wang[1,†] and Qi Xie[1,†]

[1] State Key Laboratory of Plant Genomics, National Center for Plant Gene Research (Beijing), Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Datun Road, Beijing 100101, China

[2] BGI-Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen 518083, China

[3] Department of Horticulture & Landscape Architecture, Purdue University, West Lafayette, Indiana 47907, USA

[4]Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[5]Division of Applied Life Sciences, Gyeongsang National University, Jinju 660-701, Korea

[6]College of Science, King Abdulaziz University, Jeddah 21589, Saudi Arabia

* These authors contributed equally to this work.

[†]Corresponding authors

To whom correspondence should be addressed: E-mail: qxie@genetics.ac.cn or xjwang@genetics.ac.cn or zhu132@purdue.edu

FAX: 86-10-64889351; 86-10-64860941

Table of Contents

Supporting Materials and Methods

Assembly accuracy

The accuracy of the assembled genome was confirmed using available ESTs and BAC sequences. Nearly 98% of all ESTs showed exact sequence matches with the assembled genome over at least 50% of their entire length. Four BAC sequences from NCBI and two from BGI showed 95% coverage and greater than 99.9% accuracy of low repeat regions.

Repeat annotation

Known TEs were identified using RepeatMasker (version 3.3.0) to search against the Repbase TE library (version 15.11) (1). TEdenovo pipeline included in the REPET (2) package was used for identifying novel repetitive sequences. Default parameters were used except for "minNbSeqPerGroup: 5". The resultant *de novo* output identified consensus TEs, excluding sequences classified as "NoCat", was used as the reference repeats library in a second RepeatMasker run to identify and mask novel repetitive sequences in the *T. salsuginea* genome.

Gene prediction and annotation

Protein coding gene models were identified by FGENESH++ pipeline (Softberry Inc., Mount Kisco, NY) with parameters trained with *A. thaliana* gene models. Genome sequences masked by RepeatMasker using RepBase and the *de novo* reference TE library as described in Repeat annotation section were used as input. To facilitate the gene prediction with transcriptome evidence, a *T. salsuginea* reference transcriptome was assembled from Illumina RNA-seq reads using Abyss and Vmatch (http://www.vmatch.de/). Known *T. salsuginea* ESTs and full-length cDNA sequences from NCBI database were added to the reference transcriptome. *De novo* predicted gene models were corrected based on comparison to all known plant protein sequences from the NCBI NR database. The reference transcriptome was aligned to the genome sequence and used to identify the borders of exons and untranslated regions (UTRs) for gene models with transcriptome evidence. Open reading frame (ORF) sequences less than 150 nucleotides were filtered out. The nucleotide ORF and protein sequences were annotated based on sequence homology to known sequences, using BlastN and

BlastP (E-value ≤ 1e-5) to search against the NCBI nt and nr databases (ftp://ftp.ncbi.nih.gov/blast/db/), respectively. The Blast2GO pipeline was used for Gene Ontology annotation, with the incorporation of InterProScan and KEGG pathway search results (3).

Gene family analysis

We used a best hit strategy for systemic identification of gene copy number variations in gene families in *T. salsuginea*. All *T. salsuginea* genes were subjected to BlastP search (E-value ≤ 1e-5) against all *A. thaliana* genes. The best hit to each *T. salsuginea* gene were picked up and considered as its most close orthologous gene in *A. thaliana*. A gene relationship table was generated based on the best hit strategy and was then used to calculate the gene copy number variations in each collected family. Transcription factor gene families in *A. thaliana* were downloaded from PlantTFDB (4), and stress related gene families in *A. thaliana* were manually collected from published records. Gene family member variations in other species were performed similarly. For comparison of gene models with *A. thaliana* and *T. parvula*, protein-coding gene models in TAIR10 (www.arabidopsis.org) and the version 2.0 annotation of *T. parvula* (www.thellungiella.org) were used. Gene models were clustered using OrthoMCL. Orthologous gene pairs were defined as sharing deduced amino acid sequence homology (BlastP, E-value < 1e-5) over 50% of the total length of the shorter gene being compared.

Identification of segmental and tandem duplications

To identify segmental duplications, we first performed self BlastP (-v 5 -b 5 -e 1e-10) using the deduced protein sequences of the *T. salsuginea* and *A. thaliana* genomes. A Perl script provided by DAGchainer was used to remove the repetitive matches (5). This was done by clustering all groups of matched genes that fall within 50 kb of each other and reporting only the single highest scoring match in each region. Segmental duplicated blocks were then identified using DAGchainer with optimized parameters (-s -I -D 200000 -g 10000 for *A. thaliana*; -s -I -D 500000 -g 25000 for *T. salsuginea* because of the large number of transposon insertions). To identify tandem duplications, we performed self BlastP using protein sequences with the parameters -v 100 -b 100 -e 1e-5. All genes were grouped with the following parameters: identity ≥ 70%; coverage ≥ 30%. Homologous genes within the same group and with fewer than five genes in between were identified

as tandem duplicated gene pairs.

LTR retrotransposon carrying genes and retrogenes

We used a similar method to that described by Jiang *et al.* (6) to perform systemic identification of LTR retrotransposons carrying genes and retrogenes. Full-length LTR retrotransposons were identified by using LTR_FINDER (7) with parameters -S 5 -C, which will contain at least 5 of 11 typical structural or sequence features of LTR retrotransposons. Protein coding genes entirely located within these LTR retrotransposons were considered as LTR (retrotransposon) carrying genes. To find retrogenes, we performed BlastP using the single-exon protein sequences as query, multiple-exon protein sequences as database and used the cutoff of identity ≥ 70%, query coverage ≥ 70 % and E value < 1e-8 to select retrogenes.

Phylogenetic tree construction and species divergent time estimation

The phylogenetic tree of the *T. salsuginea* and the other plant genomes was constructed using the 2226 single-copy orthologuous genes and 4-fold degenerate sites (4dTv) method. The divergence time between *T. salsuginea* and *A. thaliana* was estimated by the MULTIDIVTIME program.

Quantification of TsHKT1 transcripts with real time reverse transcription polymerase chain reaction (RT-PCR)

RNA samples from *A. thaliana* and *T. salsuginea* seedlings with and without salt stress were prepared essentially as described by Oh et al (8). To deduce absolute copy numbers of transcripts per μg total RNA samples, calibration curves were generated by performing real time PCR using 7900 HT Fast Real-Time PCR system (Applied Biosystems, Carlsbad, CA) with serial dilutions of known amount of recombinant plasmid DNA molecules that contain the template sequences (9). The recombinant plasmids were prepared by cloning RT-PCR products amplified by the following primers into the pGemTeasy vector (Promega, Madison, WI):

AtHKT1 223F GAAGTCTTCTCCAACACCCAACTT

AtHKT1 823R TACTTGAGGGATTAGGAGCCAGA

TsHKT1;1 44F TTGCTAAAAATCCTTCCGTCCTCT

TsHKT1;1 770R CCCGAAACGAGAAACAATAAAAAGC

TsHKT1;2 409F AATCATGTCAAGCTTTCTAGTCAG

TsHKT1;2 1152R TCCTTTAATTTCATCTCCGGAATCGTGT

TsHKT1;3 424F GATCATGTCAAGATTTCTAGTCAGA

TsHKT1;3 1181R AAATCCACTTTTCTTTCCCTTCTTTTCATTTC

Real time RT-PCR was performed using primers that are specific to each of the *A. thaliana* and *T. salsuginea HKT1* gene homologs. From the real time RT-PCR results and the calibration curves, the absolute transcript copy numbers were calculated as described by Pfaffl (9). Primer sequences are listed below:

AtHKT1 476F CGGTGGTTCTTAGTTACCATCTT

AtHKT1 594R GAGAGGTGAGATTTCTTTGGAACT

TsHKT1;1 195F GTCTCCTCCATGTCCACCATCG

TsHKT1;1 305R AGAGTGTGAGGAATGAAGTAAAGACCTCG

TsHKT1;2 782F CAAATCGAGAAGAATTGGGTTACATTCT

TsHKT1;2 903R GCAGAATAGAAGAAACTGTATCATCACAAGC

TsHKT1;3 785F CAAAGCGCGACGAATTTGGTTATATTC

TsHKT1;3 928R GCAGAAGAGAAGAAACTGTATCATCACAAAC

References

1. Jurka J, *et al.* (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110:462-467.

2. Flutre T, Duprat E, Feuillet C, & Quesneville H (2011) Considering transposable element diversification in de novo annotation approaches. *PLoS One* 6:e16526.

3. Gotz S, *et al.* (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* 36:3420-3435.

4. Zhang H, *et al.* (2011) PlantTFDB 2.0: update and improvement of the comprehensive plant transcription factor database. *Nucleic Acids Res* 39:D1114-1117.

5. Haas BJ, Delcher AL, Wortman JR, & Salzberg SL (2004) DAGchainer: a tool for mining segmental genome duplications and synteny. *Bioinformatics* 20:3643-3646.

6. Jiang SY, Christoffels A, Ramamoorthy R, & Ramachandran S (2009) Expansion mechanisms and functional annotations of hypothetical genes in the rice genome. *Plant Physiol* 150:1997-2008.

7. Xu Z & Wang H (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* 35:W265-268.

8. Oh DH, *et al.* (2009) Loss of halophytism by interference with SOS1 expression. *Plant Physiol* 151:210-222.

9. Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29:e45.

Supporting Figures

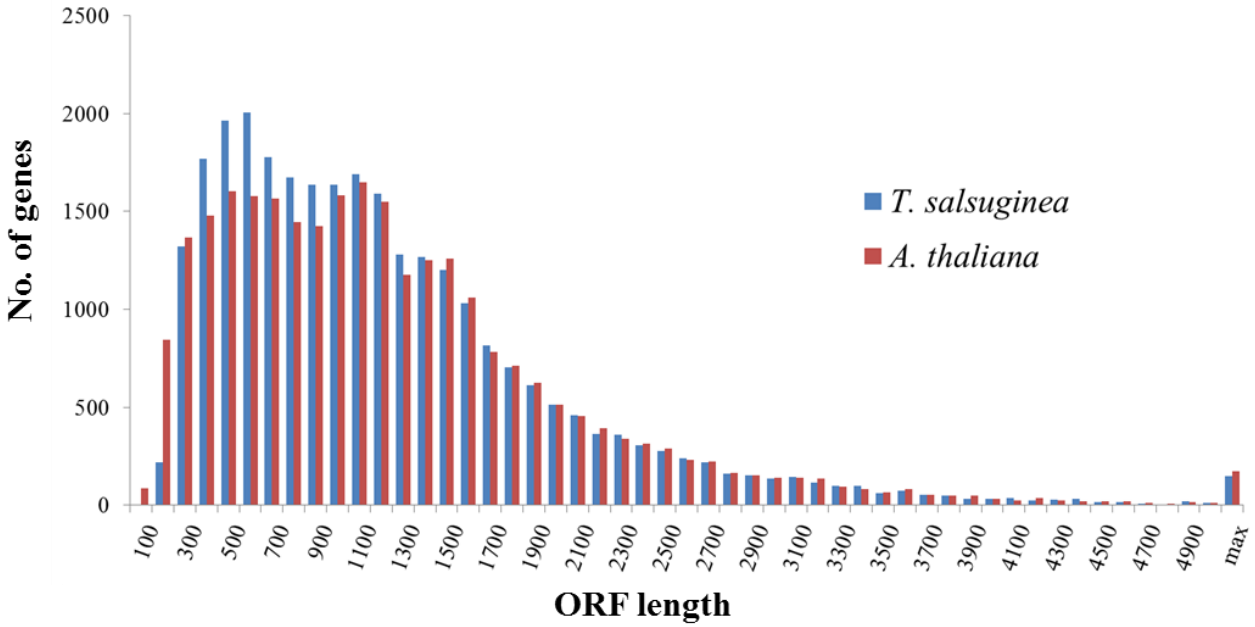


Fig. S1. ORF length distribution comparison between *T. salsuginea* and *A. thaliana*.

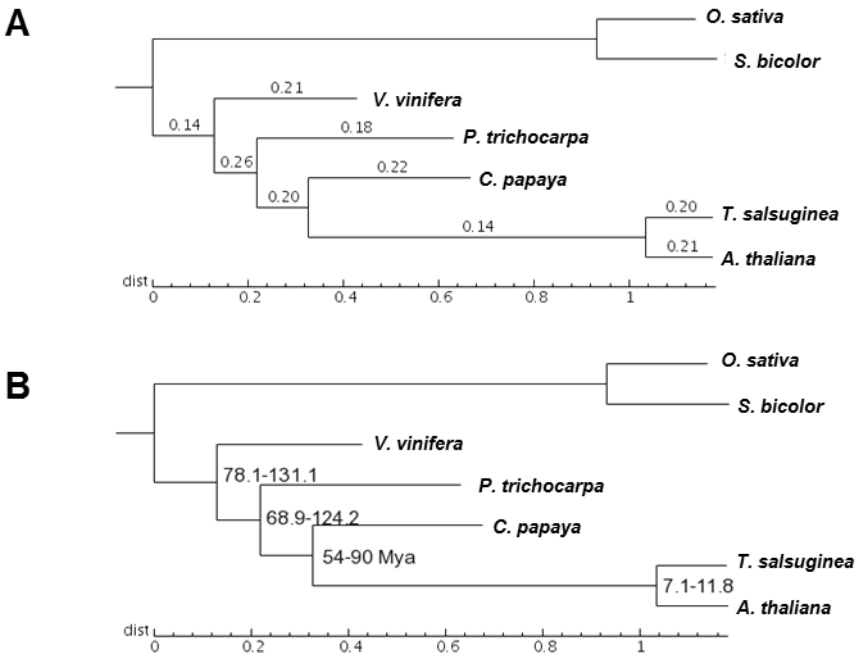


Fig. S2. Phylogenetic tree and estimation of species divergent time.

A. Phylogenetic tree of selected plant species constructed with 2226 single-copy gene families on 4-fold degenerate sites. The branch length represents the neutral divergence rate. Numbers shown on the branches represent the dN/dS rate of each branch. The posterior probabilities (credibility of the topology) for inner nodes are all 100%.

B. Estimation of divergent time. The numbers on the nodes identify the divergent time from the present (million years ago, Mya). The calibration time (fossil record time) interval (54-90 Mya) for Capparales was taken from published reports (Wikström, 2001; Crepet, 2004).

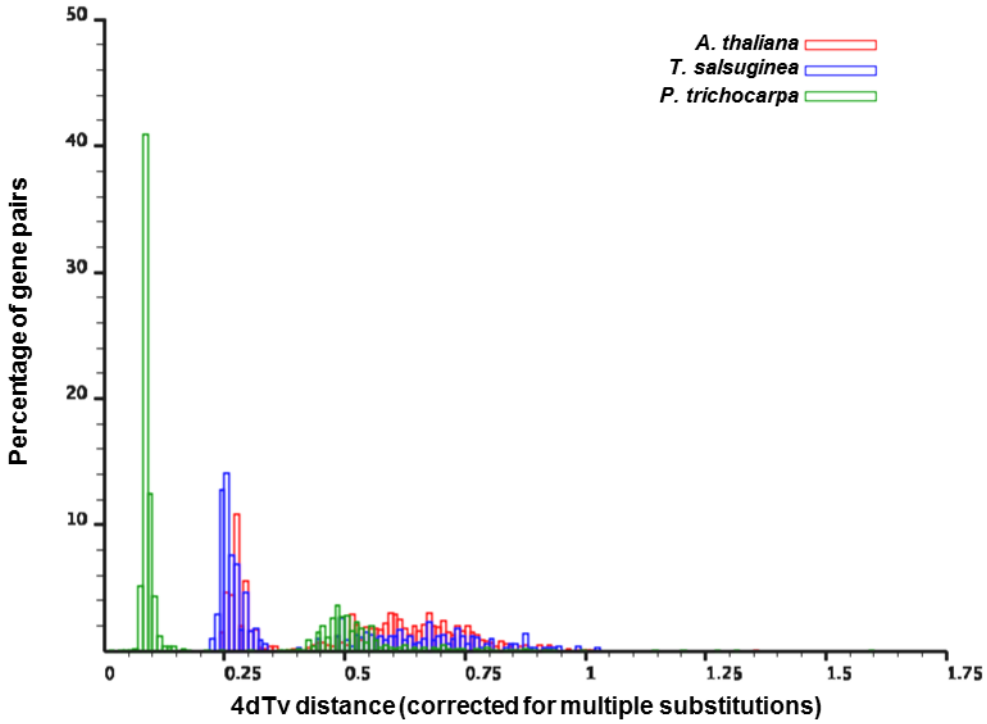
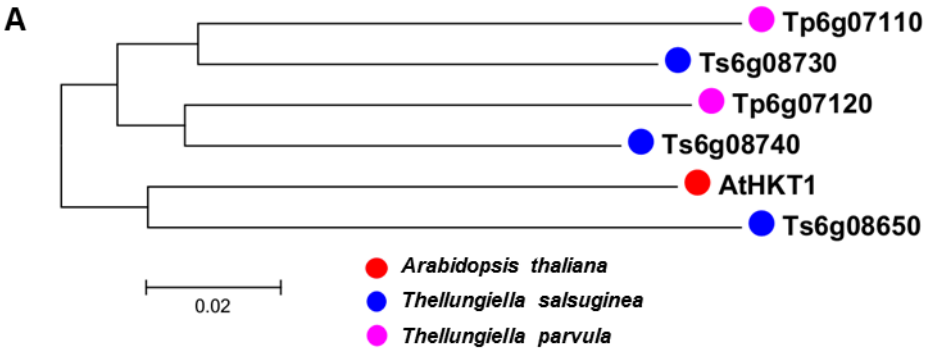
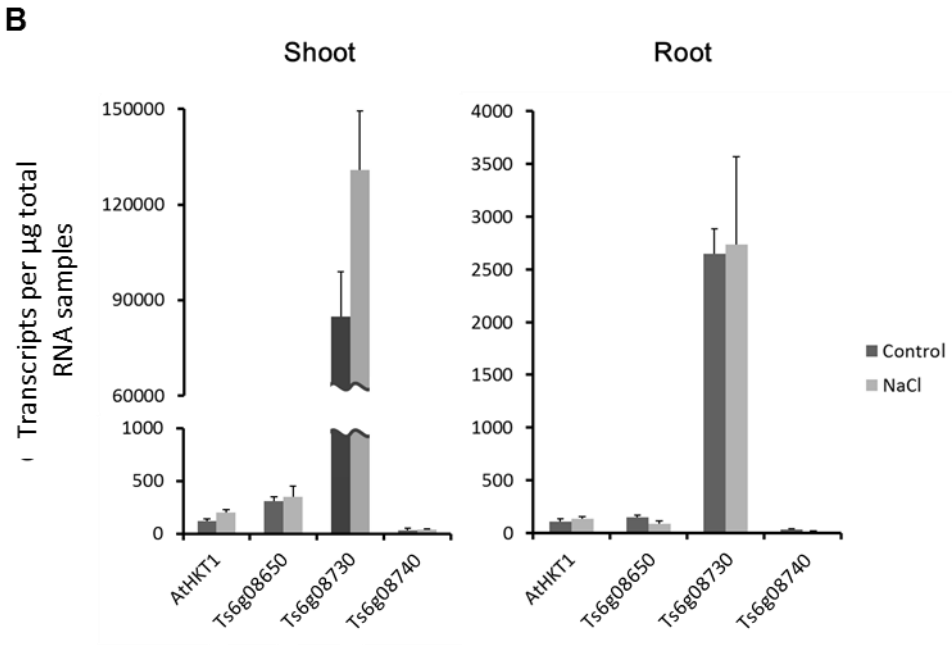


Fig. S3. 4dtv distance distribution for *T. salsuginea*, *A. thaliana* and *P. trichocarpa*.
The intra-genomic syntenic blocks among *T. salsuginea*, *A. thaliana*, and *P. trichocarpa* were detected using Mcscan program. The intervening gene number cutoffs in each block are 10 for *T. salsuginea* and *A. thaliana*, and 8 for *P. trichocarpa*, respectively. The 4dtv distances are calculated based on 4-fold degenerate sites following the HKY substitution model.


A
Tp6g07110
Ts6g08730
Tp6g07120
Ts6g08740
AtHKT1
Ts6g08650
0.02
Arabidopsis thaliana
Thellungiella salsuginea
Thellungiella parvula


B
Shoot
Root
Transcripts per µg total RNA samples
150000
120000
90000
60000
1000
500
0
4000
3500
3000
2500
2000
1500
1000
500
0
AtHKT1
Ts6g08650
Ts6g08730
Ts6g08740
Control
NaCl

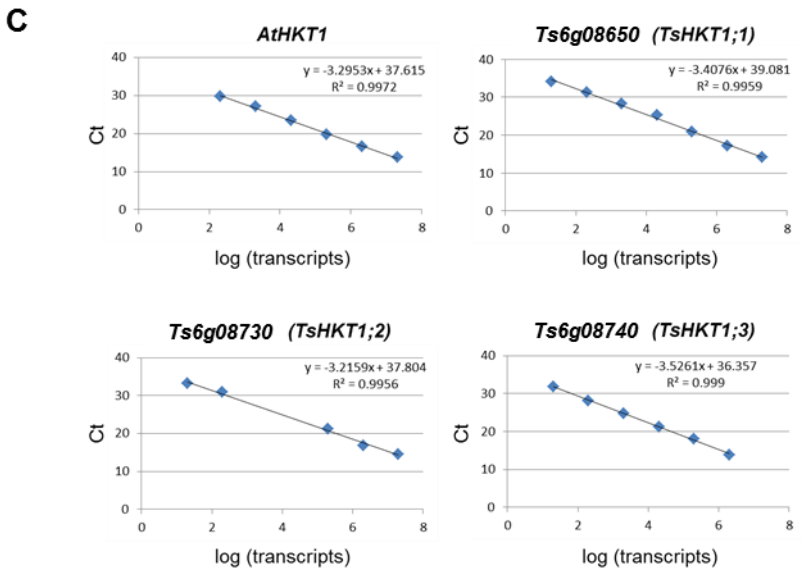

C
AtHKT1
y = -3.2953x + 37.615
R² = 0.9972
Ts6g08650 (TsHKT1;1)
y = -3.4076x + 39.081
R² = 0.9959
Ts6g08730 (TsHKT1;2)
y = -3.2159x + 37.804
R² = 0.9956
Ts6g08740 (TsHKT1;3)
y = -3.5261x + 36.357
R² = 0.999
Ct
log (transcripts)

Fig S4 Phylogenetic and expression analysis of *HKT1* genes.

A. Phylogenetic analysis of plant *HKT1* genes identifies three gene groups (Class I, II and III).

B. Quantification of transcripts of *HKT1* homologs from *A. thaliana* and *T. salsuginea*. RNA samples from 2 week-old *A. thaliana* and 3 week-old *T. salsuginea* plants treated with 200 mM NaCl for 12 hours were subjected to quantitative real-time RT-PCR as described in *SI Appendix*.

C. Standard calibration curves used for deducing the absolute transcript copy numbers from the real-time RT-PCR results. For detailed methods, see the *SI Appendix* and references therein.

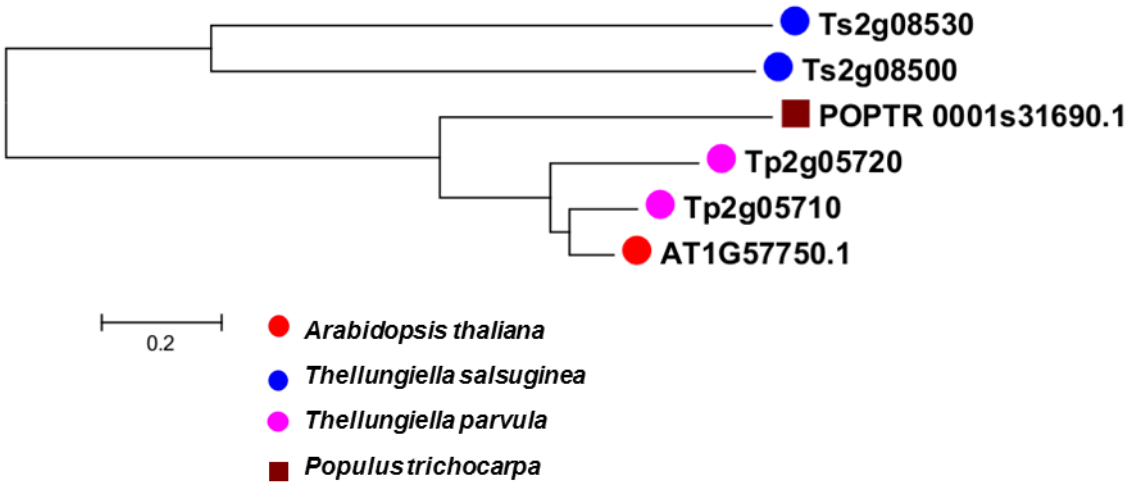


Fig. S5. Phylogenetic analysis of *MAH1/CYP96A15* genes in *T. salsuginea, A. thaliana, T. parvula, P. trichocarpa.*

The phylogenetic tree was constructed using the Neighbor Joining Method with the Mega 5.0 software. The MAH1/CYP96A15 gene, which belongs to the P450 gene family and functions as a key enzyme in the alkane-forming pathway, is tandem duplicated in both *T. salsuginea* and *T. parvula*. We failed to find the corresponding MAH1 genes in *V. vinifera*, *C. papaya and O. sativa*.

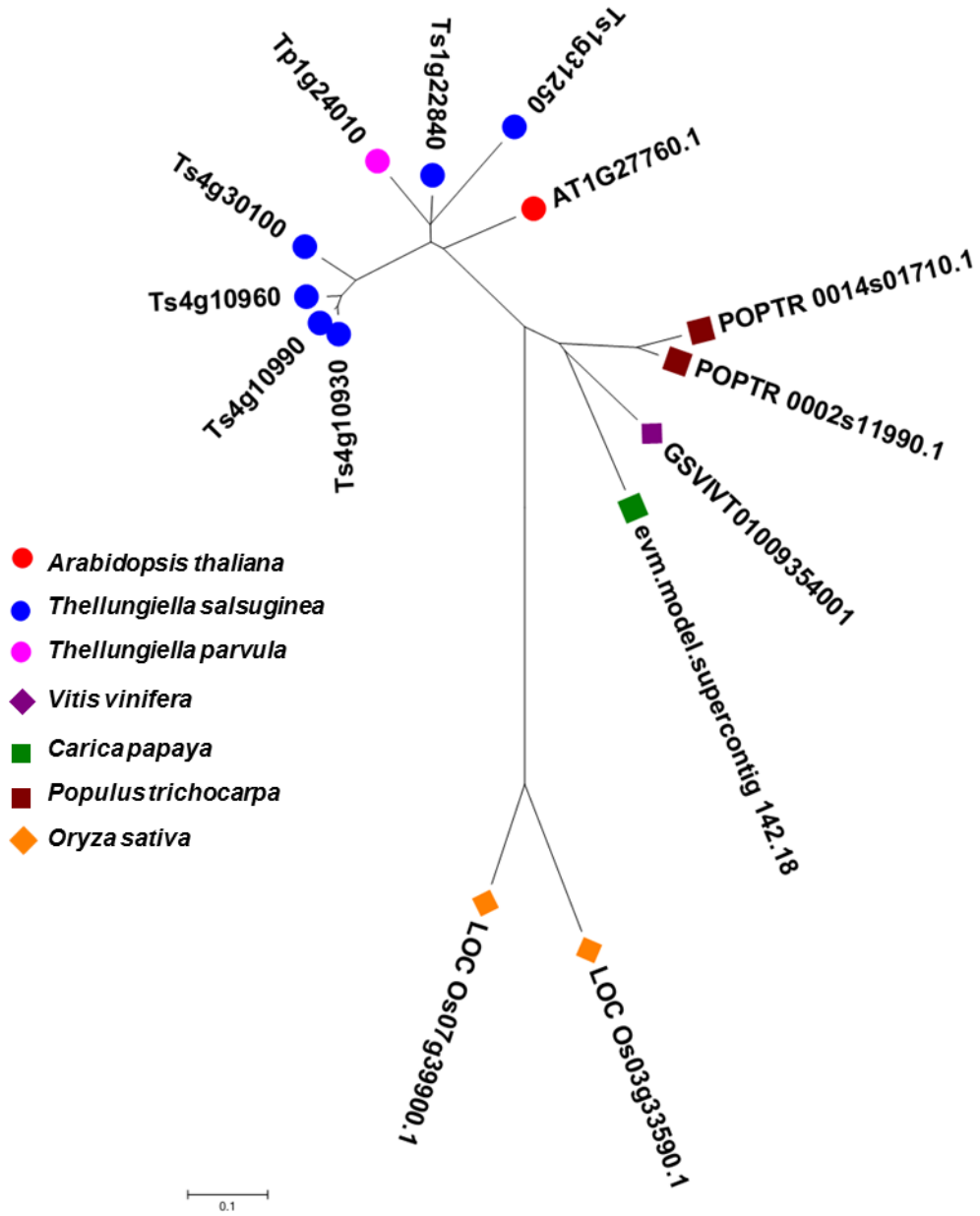


Fig. S6. Phylogenetic analysis of *SAT32* genes in *T. salsuginea, A.thaliana, T. parvula, V. vinifera, P. trichocarpa, C. papaya, O. sativa*.

The phylogenetic tree was constructed using the Neighbor Joining Method with the Mega 5.0 software.

Supporting Tables

Table S1. Features of the *T. salsuginea* genome.

Feature	Value
Estimated genome size	260 Mb
Assembled genome sequence	233,653,061 bp
Length of scaffolds in seven chromosomes	186,126,548 bp
Number of scaffolds anchored to chromosomes	515
Number of unplaced scaffolds	2167
Length of unplaced scaffolds	47,526,513 bp
Total number of scaffolds	2682
N50	403,516 bp
Number of scaffolds at least N50	119
Transposable elements (percentage)	121,046,173 bp (51.81%)
DNA transposons	20,160,164 bp (8.63%)
Retrotransposon	90,570,024 bp (38.76%)
Other	10,315,985 bp (4.42%)
Number of genes	28,457
Length of coding regions (percentage)	58,138,525 bp (24.88%)
Gene density	122 genes per Mb
Average gene length	2,041 bp
Average protein length	398 aa
Number of exons (per gene)	149,079 (5.23)
Average exon length	228 bp
Average intron length	200 bp
Gene annotation (percentage)	
InterPro	19,920 (69.9%)
GO	21,859 (76.8%)
With NCBI NR blast hit	26,016 (91.3%)
With ATH blast hit	25,288 (88.8%)
Unannotated	1,836 (6.45%)
Non-coding RNAs	
miRNA	162
tRNA	447
rRNA	11
snRNA	432

Table S2. Summary of the *T. salsuginea* genome sequencing data. The estimated genome size of 260 Mb is used to calculate the sequencing depth.

Insert size (bp)	Average read size (bp)	No. of sequencing lanes	No. of usable reads (Million)	No. of usable bases (Mb)	Sequencing depth (fold)
180	90	1	47.90	4311.56	16.58
200	41	3	98.75	4048.91	15.57
340	60	1	51.01	3060.76	11.77
374	75	2	84.65	6349.08	24.42
682	75	2	65.25	4893.7	18.82
2000	44	3	93.48	4113.16	15.82
2000	44	1	20.19	888.21	3.42
5000	44	1	25.12	1105.31	4.25
5000	44	3	90.76	3993.57	15.36
10000	44	2	14.07	619.24	2.38
10000	44	1	16.83	740.55	2.85
10000	44	1	15.70	690.76	2.66
Total		21	623.73	34814.81	133.90

Table S3. Statistics of repeat sequences in the *T. salsuginea* genome.

Identification method	Type of repeats	On seven chromosomes	Unanchored	All
RepBase	Retroposon	19,017,116	9,762,654	28,779,770
	DNA transposon	4,721,481	1,065,990	5,787,471
	Other	230,210	32,886	263,096
TEdenovo	Retroposon	36,353,568	25,436,686	61,790,254
	DNA transposon	11,721,545	2,651,148	14,372,693
	Other	7,256,637	2,796,252	10,052,889
Total repeats		79,300,557 (43%)	41,745,616 (88%)	121,046,173 (52%)

Table S4. Non-coding RNA genes in the assembled genome.

Type		Copy number	Average length(bp)	Total length(bp)
tRNA		447	74	33,154
rRNA		11	508	5,588
snRNA	CD-box snoRNA	323	99	31,919
	HACA-box snoRNA	37	124	4,589
	splicing	72	141	10,163
miRNA	Conversed	126	152	19,111
	Novel	36	118	4,252

Table S5. Functional comparison on different types of duplicated genes between *T. salsuginea* and *A. thaliana*. Blast2GO results of protein coding regions from *T. salsuginea* and *A. thaliana* were mapped to categories in the second level of GO terms. Fisher's exact test was performed to identify the significantly differed GO terms. P-values less than 0.05 and 0.01 are shown with light and dark grey circles, respectively. TD: tandem duplicated genes; SD: segmental duplicated genes; LTR: LTR retrotransposon carrying genes; RETRO: retrogenes.

Gene category	Total (ATH/TSA)	TD (ATH/TSA)	SD (ATH/TSA)	LTR (ATH/TSA)	RETRO (ATH/TSA)
biological regulation	5655/5368	480/495	2373/2253	65/87	74/96
carbon utilization	84/91	4/4	38/41	0/2	2/1
cell killing	11/17	6/9	2/2	0/1	0/0
cell proliferation	41/49	0/4	19/18	0/1	1/0
cellular component organization or biogenesis	1999/2248	154/202	819/893	21/34	27/32
cellular process	10594/11452	1125/1234	3885/4093	105/196	149/196
death	187/215	22/31	51/51	3/3	1/4
developmental process	2258/2724	175/227	987/1149	16/28	20/32
establishment of localization	2278/2394	224/237	898/921	15/21	18/25
growth	406/470	36/45	212/243	0/5	8/8
immune system process	357/398	55/62	132/107	10/11	2/5
localization	2364/2490	231/251	935/960	16/21	18/27
locomotion	17/28	3/6	6/13	2/0	0/0
metabolic process	9670/10308	1308/1269	3230/3436	92/190	138/175
multi-organism process	1135/1274	193/188	454/484	24/21	9/12
multicellular organismal process	2186/2677	175/242	926/1094	18/31	17/33
negative regulation of biological process	431/522	18/39	198/191	3/7	5/13
pigmentation	7/7	0/0	5/5	0/0	0/0
positive regulation of biological process	461/480	32/42	213/200	4/13	3/8
regulation of biological process	4778/5065	386/464	2034/2130	55/84	64/95
reproduction	1230/1490	91/155	497/577	8/15	8/16
reproductive process	1202/1463	88/149	489/567	7/15	8/16
response to stimulus	5412/6049	737/813	2208/2417	63/91	93/134
rhythmic process	61/101	0/6	39/56	0/0	0/0
signaling	1713/1737	201/208	694/747	33/32	36/38
viral reproduction	11/22	3/0	5/11	2/0	0/0
cell	17451/16118	1960/1624	6020/5822	232/242	252/258
cell junction	24/28	0/0	16/16	0/0	0/0
extracellular region	654/637	141/108	247/277	7/8	14/14
extracellular region part	53/73	9/5	25/43	1/0	0/1
macromolecular complex	4393/4382	278/312	1900/1826	41/47	64/70
membrane-enclosed lumen	2579/2556	175/191	1239/1160	27/33	36/36
organelle	10496/10750	833/931	3823/3837	104/160	123/172
symplast	17/18	0/0	11/10	0/0	0/0
virion	15/3	0/0	4/3	0/0	0/0
antioxidant activity	162/155	28/20	55/56	1/1	1/1
binding	12951/13785	1359/1478	4396/4718	130/312	171/246
catalytic activity	9228/9703	1273/1289	3052/3221	95/193	108/173
channel regulator activity	7/6	2/4	0/0	0/0	0/0
electron carrier activity	547/535	146/124	175/156	4/17	8/24
enzyme regulator activity	380/375	28/35	171/157	3/7	1/1
metallochaperone activity	4/3	0/0	2/0	0/0	0/0
molecular transducer activity	419/429	50/52	163/161	3/5	3/3
nucleic acid binding transcription factor activity	1734/1669	108/118	871/840	14/19	17/22
nutrient reservoir activity	67/56	28/20	19/24	3/1	0/0
protein binding transcription factor activity	61/66	2/4	18/22	0/0	1/1
protein tag	5/4	1/2	2/3	0/0	0/0
receptor activity	294/344	30/47	116/136	3/1	3/2
structural molecule activity	567/545	43/35	245/222	9/5	2/8
translation regulator activity	3/3	0/0	0/0	0/0	0/0
transporter activity	1347/1405	167/167	512/524	9/11	11/22
Total genes	**27416/28457**	**2708/2723**	**8429/8178**	**444/842**	**353/535**

Table S6. Comparison of transcription factor gene families between *T. salsuginea*, *T. parvula* and *A. thaliana*.

Gene Family	No. of genes		
	T. salsuginea	***T. parvula***	***A. thaliana***
RAV	9	6	6
NF-X1	3	2	2
EIL	9	8	6
LSD	4	4	3
ARR-B	18	21	14
G2-like	53	47	42
Nin-like	17	14	14
GRAS	40	35	33
HSF	28	23	24
CAMTA	7	6	6
E2F/DP	9	9	8
CPP	9	6	8
GRF	10	10	9
AP2	20	16	18
B3	69	59	64
Trihelix	31	29	29
M-type	70	52	66
MIKC	44	42	42
GATA	31	31	30
HD-ZIP	49	55	48
bZIP	75	76	74

Note: the TF data were downloaded from: http://planttfdb.cbi.pku.edu.cn/index.php?sp=At.

RAV Family: RAV transcription factor were strongly induced after pathogen infection and salt (PMID: 16927203) & RAV transcription factor were induced by cold stress (PMID: 15728337).

NF-X1 Family: The AtNFXL1 gene encodes a NF-X1 type zinc finger protein required for growth under salt stress (PMID: 16905136).

GRAS Family: involves in plant development regulation. RGL3 transcript levels were transiently increased by cold (PMID: 18757556).

HSF Family: heat stress factors. Salt and osmotic stress induced *HsfA2* gene expression, and *HSFA2* overexpression mutant showed enhanced osmotic stress (PMID: 17890230).

Trihelix Family: The transcript level of *OsGTγ-1* was strongly induced by salt stress, and overexpression of *OsGT γ-1* in rice enhanced salt tolerance at the seedling stage (PMID: 20039179).

EIL : ethylene. **LSD:** PCD. **ARR-B:** cytokinin. **G2-like:** chloroplast development. **Nin-like:** root nodules. **CAMTA:** calmodulin binding TF. **E2F/DP:** cell proliferation. **CPP:** cell division. **GRF:** growth regulation. **AP2:** development. **B3:** includes LAV, REM and RAV family. M-type&MIKC: MADS-box TFs. **GATA:** light responsive. **HD-ZIP:** development.

Table S7. Species distribution analysis of ionic homeostasis related gene families.

Gene Family	No. of genes		
	T. salsuginea	***T. parvula***	***A. thaliana***
NHX	8	11	8
HKT1	3	2	1
Shaker	9	9	9
KEA	6	6	6
KUP-HAK-KT	13	18	13
CNGC	27	21	20
TPK	4	7	6
PPa	7	6	6
AHA	10	10	11
ACA	16	12	11
ECA	3	4	4
CHX	28	28	29
CAX	5	5	6
AVP	4	3	2
VHA.a	3	3	3
VHA.c’	4	4	5
VHA.c’’	1	2	2
VHA.d	2	2	2
VHA.e	1	2	2
VHA-A	1	1	1
VHA-B	3	4	3
VHA-C	1	1	1
VHA-D	1	1	1
VHA-E	3	4	3
VHA-F	1	1	1
VHA-G	3	3	3
VHA-H	1	1	1
GLR	12	14	20
CCC	1	1	1
ATBGL	49	39	46
CBL	9	10	10
CIPK	30	28	25
CDPK	37	36	34

Table S8. Species distribution analysis of wax biosynthesis gene families.

Gene Family	No. of genes		
	T. salsuginea	***T. parvula***	***A. thaliana***
ACC	4	2	2
FATB	1	1	1
LACS	11	9	9
KCS	22	24	21
KCR	3	2	2
HCD	1	1	1
ECR	1	1	1
FAR	9	10	8
WS/DGAT	11	16	11
MAH1/CYP96A15	2	2	1
WBC11	1	1	1
CER5/WBC12	1	2	1
CER1/CER-like	3	3	4
CER2	1	0	1
CER3/WAX2/YRE/FLP1	1	1	1
CER7	1	1	1
WIN1/SHN1	1	0	1
Total	74	76	67

Table S9. Species distribution analysis of ABA biosynthesis and ABA signaling related gene families.

Gene Family	No. of genes		
	T. salsuginea	***T. parvula***	***A. thaliana***
ZEP	2	1	1
AAO	7	4	4
ABA3	1	1	1
NCED	7	7	7
CYP707A	5	4	4
SDIR1	1	1	1
PP2C	75	74	74
SNRK2	9	11	10
ABF	4	4	4
ABI5	1	1	1
AFP	4	4	4

Table S10. Species distribution analysis of other gene families related to salinity, drought and cold stress response or tolerance.

Gene Family	No. of genes		
	T. salsuginea	***T. parvula***	***A. thaliana***
PLD	15	11	12
P5CDH	1	1	1
P5CS	2	2	2
PDH	2	2	2
DREB	56	55	56
ERF	59	67	62
MAPK	18	19	20
MAPKK	10	11	10
MEKK	20	20	21
ZIK	11	11	11
Raf	45	50	48
AHK1	1	2	1
SKB1	3	2	1
SIZ1	2	2	1
LEA	42	41	40
OTS	2	3	2
ATSAT32	6	1	1