

**Evolution of short inverted repeat in cupressophytes, transfer of *accD* to nucleus in *Sciadopitys verticillata* and phylogenetic position of Sciadopityaceae**

Jia Li<sup>1,2,†</sup>, Lei Gao<sup>1,†</sup>, Shanshan Chen<sup>1,2</sup>, Ke Tao<sup>1,2</sup>, Yingjuan Su<sup>3,4,\*</sup>, and Ting Wang<sup>1,5,\*</sup>

<sup>1</sup>CAS Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, Hubei, China, <sup>2</sup>University of Chinese Academy of Sciences, Beijing, China, <sup>3</sup>State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou, Guangdong, China, <sup>4</sup>Institute for Technology Research and Innovation of Sun Yat-sen University, Zhuhai, Guangdong, China, and <sup>5</sup>College of Life Sciences, South China Agricultural University, Guangzhou, Guangdong, China.

<sup>†</sup> These authors contributed equally to this paper.

\*Correspondence: Ting Wang, Tel: +86-27-87510677. Fax: +86-27-87510251. E-mail address:

[tingwang@wbgcas.cn](mailto:tingwang@wbgcas.cn);

Yingjuan Su, Tel: +86-20-84111939. Fax: +86-20-84036215. E-mail: [suyj@mail.sysu.edu.cn](mailto:suyj@mail.sysu.edu.cn)

Supplementary Table S1. Distribution of simple sequence repeats in *Sciadopitys verticillata* chloroplast genome.

SSR nr.	SSR	size	Start-end	Location
1	(G)12	12	7929--7940	<i>chlB/trnN-GUU</i> (IGS)
2	(A)21	21	9596--9616	<i>rrn4.5-rrn23</i> (IGS)
3	(TTTG)4	16	10797--10812	<i>rrn23</i>
4	(T)11	11	14615--14625	<i>trnI-GAU/rrn16</i> (IGS)
5	(AT)5	10	15300--15309	<i>trnI-GAU/rrn16</i> (IGS)
6	(AG)5	10	15702--15711	<i>trnI-GAU/rrn16</i> (IGS)
7	(A)13	13	19017--19029	<i>trnV-GAC/rps12 exon2</i> (IGS)
8	(A)19	19	20215--20233	<i>rps12 intron</i>
9	(AT)5	10	20288--20297	<i>rps12 intron</i>
10	(A)11	11	35682--35692	<i>rpl2 intron</i>
11	(A)13	13	38632--38644	<i>rpl16 intron</i>
12	(A)18	18	40655--40672	<i>rps8/infA</i> (IGS)
13	(T)10	10	41474--41483	<i>rpl36/rps11</i> (IGS)
14	(AGA)5	15	45759--45773	<i>petB intron</i>
15	(AT)21	42	46417--46458	<i>psbH/psbN</i> (IGS)
16	(AT)9	18	47048--47065	<i>psbT/atpF</i> (IGS)
17	(A)13	13	48467--48479	<i>atpF/atpH</i> (IGS)
18	(T)13	13	51081--51093	<i>rps2/rpoC2</i> (IGS)
19	(T)10	10	51289--51298	<i>rpoC2</i>
20	(T)10	10	51328--51337	<i>rpoC2</i>
21	(T)10	10	53294--53303	<i>rpoC2</i>
22	(A)10	10	57791--57800	<i>petL/psbE</i> (IGS)
23	(T)10	10	61455--61464	<i>petA/cemA</i> (IGS)
24	(A)11	11	61557--61567	<i>petA/cemA</i> (IGS)
25	(A)11	11	61584--61594	<i>petA/cemA</i> (IGS)
26	(T)10	10	62061--62070	<i>cemA</i>
27	(A)10	10	62908--62917	<i>cemA/ycf4</i> (IGS)
28	(AT)9	18	65597--65614	<i>trnE-UUC/trnT-GGU</i> (IGS)
29	(AT)18	36	66522--66557	<i>trnT-GGU/psbD</i> (IGS)
30	(T)16	16	69481--69496	<i>psbC/trnS-UGA</i> (IGS)
31	(A)10	10	69868--69877	<i>trnS-UGA/psbZ</i> (IGS)
32	(T)13	13	70583--70595	<i>trnG-GCC/trnfM-CAU</i> (IGS)
33	(A)10	10	70817--70826	<i>trnfM-CAU/rps14</i> (IGS)
34	(G)12	12	79121--79132	<i>trnS-GGA/rps4</i> (IGS)
35	(G)14	14	80451--80464	<i>trnT-UGU/trnL-UAA</i> (IGS)
36	(AT)19	38	81507--81544	<i>trnF-GAA/ndhJ</i> (IGS)
37	(T)11	11	81642--81652	<i>trnF-GAA/ndhJ</i> (IGS)
38	(ATT)4	12	85058--85069	<i>trnV-UAC intron</i>
39	(AT)9	18	85631--85648	<i>trnV-UAC/trnM-CAU</i> (IGS)
40	(T)10	10	86093--86102	<i>trnM-CAU/atpB</i> (IGS)
41	(TA)15	30	88468--88497	<i>atpE/rbcL</i> (IGS)

---

42	(T)10	10	90718--90727	<i>trnR-CCG/trnR-UCU</i> (IGS)
43	(A)11	11	90863--90873	<i>trnR-UCU/trnG-UCC</i> (IGS)
44	(A)12	12	101075--101086	<i>ycfI</i>
45	(A)10	10	102814--102823	<i>ycfI</i>
46	(AGA)7	21	106389--106409	<i>ycfI</i>
47	(TGA)8	24	107862--107885	<i>ycfI</i>
48	(A)13	13	122238--122250	<i>rpoB</i>
49	(T)16	16	130248--130263	<i>rps12 exon1</i>
50	(TTC)4	12	130571--130582	<i>rps12 exon1/clpP</i> (IGS)
51	(T)10	10	130748--130757	<i>clpP</i>
52	(A)11	11	135094--135104	<i>atpA</i>
53	(TA)8	16	137825--137840	<i>trnC-GCA/trnI-CAU</i> (IGS)

---

Note: IGS, intergenic spacers.

Supplementary Table S2. Distribution of tandem repeats in *Sciadopitys verticillata* chloroplast genome.

No.	Repeat length	Consensus copy number	size	Start-end	location
1	359	53 × 6.8		14932--15290	<i>trnI-GAU/rrn16</i> (IGS)
2	272	134 × 2		15306--15577	<i>trnI-GAU/rrn16</i> (IGS)
3	146	43 × 3.4		17637--17782	<i>rrn16/ trnV-GAC</i> (IGS)
4	77	39 × 2		28223--28299	<i>ycf2</i>
5	133	66 × 2		29468--29600	<i>ycf2</i>
6	66	33 × 2		29791--29856	<i>ycf2</i>
7	189	36 × 5.3		30734--30922	<i>ycf2</i>
8	314	75 × 4.2		31264--31577	<i>ycf2</i>
9	620	33 × 18.2		31975--32594	<i>ycf2</i>
10	389	179 × 2.2		33730--34118	<i>ycf2/ trnI-CAU</i> (IGS)
11	126	66 × 1.9		86082--86207	<i>trnM-CAU/atpB</i> (IGS)
12	207	73 × 2.8		94306--94512	<i>trnQ-UUG</i>
13	524	170 × 3.1		98374--98897	<i>rpl32/ycf1</i> (IGS)
14	116	57 × 2		99956--100071	<i>ycf1</i>
15	156	42 × 3.7		101424--101579	<i>ycf1</i>
16	273	39 × 6.9		101929--102201	<i>ycf1</i>
17	181	60 × 3		104785--104965	<i>ycf1</i>
18	137	63 × 2.2		125661--125797	<i>rpoC1</i> intron

Note: IGS, intergenic spacers.

Supplementary Table S3. Prediction of the presence of a transit peptide-encoding sequence in the nuclear *accD* transcript of *Sciadopitys verticilla* using five different softwares.

	ptTP score	mTP score	Secretory pathway	Elsewhere	cytoplasm
BaCelLO					<b>yes</b>
MultiLoc					<b>0.89</b>
Predotar	0.00	0.01		<b>0.99</b>	0.01
Protein Prowler	0.100	0.100			<b>0.650</b>
TargetP	0.058	0.179	0.087	<b>0.917</b>	

Supplementary Table S4. Primers used in this study

Primer name	direction	Primer Sequence (5'-3')
naccDF	F	ACTGTCTCCATTCAGTGGGTT
naccDR	R	CGCTTAATTGGAACAGCCCGATC
GSP3	R	TCCCTTTTAGAAGGCTCAGACC
GSP1	R	TTTATTTCGACCTGTGCCTGT
GSP4	F	ACCTTAAAGCAGCAAGTAGACG
GSP2	F	CTTACATCTCCCACAACCTGGT
GSP5R	R	GTCGAAGGTATTTCCGGCAGA
GSP6F	F	CCACGGATGTTGTTAAAAGGTGT
SP1L	R	TCGCATACTCAATTAGACGGGTGA
SP1R	F	TGACTGCTAGTTTTGCCATGTTGGG

Supplementary Table S5. Plant materials and GenBank accession numbers of the 29 species whose plastid genomes were completely sequenced.

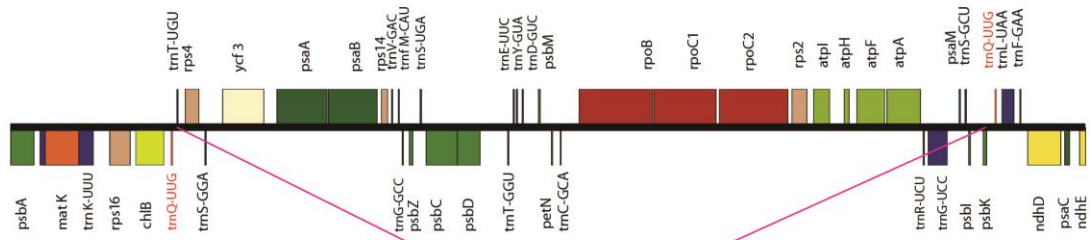
Taxa	Family	Genus	Species	GenBank Accession No.
Gymnosperm	Cephalotaxaceae	<i>Cephalotaxus</i>	<i>Cephalotaxus wilsoniana</i>	NC_016063
	Cephalotaxaceae	<i>Cephalotaxus</i>	<i>Cephalotaxus oliveri</i>	NC_021110
	Cupressaceae	<i>Cryptomeria</i>	<i>Cryptomeria japonica</i>	NC_010548
	Cupressaceae	<i>Taiwania</i>	<i>Taiwania cryptomerioides</i>	NC_016065
	Cupressaceae	<i>Calocedrus</i>	<i>Calocedrus formosana</i>	NC_023121
	Cupressaceae	<i>Cunninghamia</i>	<i>Cunninghamia lanceolata</i>	NC_021437
	Cupressaceae	<i>Juniperus</i>	<i>Juniperus virginiana</i>	NC_024024
	Cupressaceae	<i>Juniperus</i>	<i>Juniperus scopulorum</i>	NC_024023
	Araucariaceae	<i>Agathis</i>	<i>Agathis dammara</i>	NC_023119
	Podocarpaceae	<i>Agathis</i>	<i>Nageia nagi</i>	NC_023120
	Podocarpaceae	<i>Podocarpus</i>	<i>Podocarpus lambertii</i>	NC_023805
	Podocarpaceae	<i>Retrophyllum</i>	<i>Retrophyllum piresii</i>	NC_024827
	Taxaceae	<i>Taxus</i>	<i>Taxus mairei</i>	NC_020321
	Taxaceae	<i>Amentotaxus</i>	<i>Amentotaxus formosana</i>	NC_024945
	Sciadopityaceae	<i>Sciadopitys</i>	<i>Sciadopitys verticillata</i>	KT601210
	Cycadaceae	<i>Cycas</i>	<i>Cycas taitungensis</i>	NC_009618
	Ginkgoaceae	<i>Ginkgo</i>	<i>Ginkgo biloba</i>	NC_016986
	<i>Ephedraceae</i>	<i>Ephedra</i>	<i>Ephedra equisetina</i>	NC_011954
	<i>Gnetaceae</i>	<i>Gnetum</i>	<i>Gnetum montanum</i>	NC_021438
	Welwitschiaceae	<i>Welwitschia</i>	<i>Welwitschia mirabilis</i>	NC_010654
	Pinaceae	<i>Cathaya</i>	<i>Cathaya argyrophylla</i>	NC_014589
	Pinaceae	<i>Cedrus</i>	<i>Cedrus deodara</i>	NC_014575
	Pinaceae	<i>Keteleeria</i>	<i>Keteleeria davidiana</i>	NC_011930
	Pinaceae	<i>Larix</i>	<i>Larix decidua</i>	NC_016058
	Pinaceae	<i>Picea</i>	<i>Picea morrisonicola</i>	NC_016069
	Pinaceae	<i>Pinus</i>	<i>Pinus thunbergii</i>	NC_001631
Pinaceae	<i>Pseudotsuga</i>	<i>Pseudotsuga sinensis</i>	NC_016064	
Angiosperm	Solanaceae	<i>Nicotiana</i>	<i>Nicotiana tabacum</i>	NC_001879
	Amborellaceae	<i>Amborella</i>	<i>Amborella trichopoda</i>	NC_005086

Supplementary Table S6. GenBank accession numbers of the 18 *accD* sequences reported in previous research.

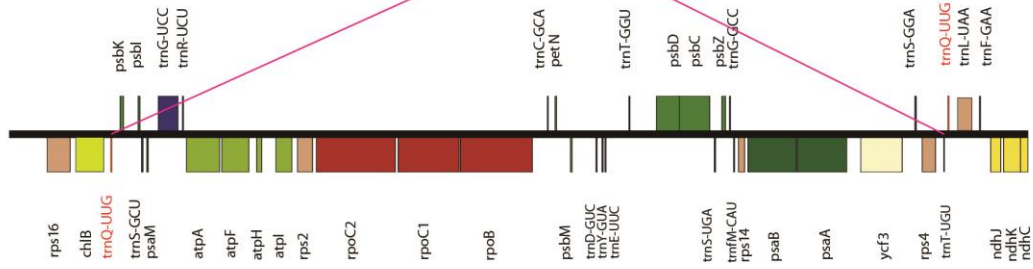
Family	Genus	Species	GenBank Accession No.	
Taxaceae	<i>Taxus</i>	<i>Taxus media</i>	KT030780	
		<i>Taxus wallichiana</i>	KT030784	
		<i>Taxus chinensis</i>	KT030782	
		<i>Taxus cuspidate</i>	KT030781	
		<i>Taxus yunnanensis</i>	KT030783	
	<i>Pseudotaxus</i>	<i>Pseudotaxus chienii</i>	KT030788	
	<i>Torreya</i>	<i>Torreya fargesii</i>	KT030785	
		<i>Torreya grandis</i>	KT030786	
	Cephalotaxaceae	<i>Amentotaxus</i>	<i>Amentotaxus argotaenia</i>	KT030787
		<i>Cephalotaxus</i>	<i>Cephalotaxus hainanensis</i>	KT030790
<i>Cephalotaxus sinensis</i>			KT030789	
Podocarpaceae	<i>Podocarpus</i>	<i>Podocarpus macrophyllus</i>	KT030792	
		<i>Podocarpus neriifolius</i>	KT030793	
	<i>Nageia</i>	<i>Nageia nagi</i>	KT030794	
		<i>Nageia fleuryi</i>	KT030795	
	Araucariaceae	<i>Araucaria</i>	<i>Araucaria cunninghamii</i>	KT030797
Taxodiaceae	<i>Metasequoia</i>	<i>Metasequoia glyptostroboides</i>	KT030796	



*Taiwania cryptomerioides*



*Cephalotaxus oliveri*



Supplementary Figure S1. The location of two copies of *trnQ-UUG* in cupressophytes chloroplast genome. One type of the genome organization around *trnQ-UUG* was represented by *Taiwania cryptomerioides*, while the other was represented by *Cephalotaxus oliveri*. The inversion between two plastomes was linked by carmine line.