

## Supplemental Information

### **Full transcription of the chloroplast genome in photosynthetic eukaryotes**

Chao Shi<sup>1,3\*</sup>, Shuo Wang<sup>2\*</sup>, En-Hua Xia<sup>1,3\*</sup>, Jian-Jun Jiang<sup>2</sup>, Fan-Chun Zeng<sup>2</sup>, and Li-Zhi Gao<sup>1,2\*\*</sup>

<sup>1</sup>Plant Germplasm and Genomics Center, Germplasm Bank of Wild Species in Southwest China, Kunming Institute of Botany, the Chinese Academy of Sciences, Kunming 650204, China

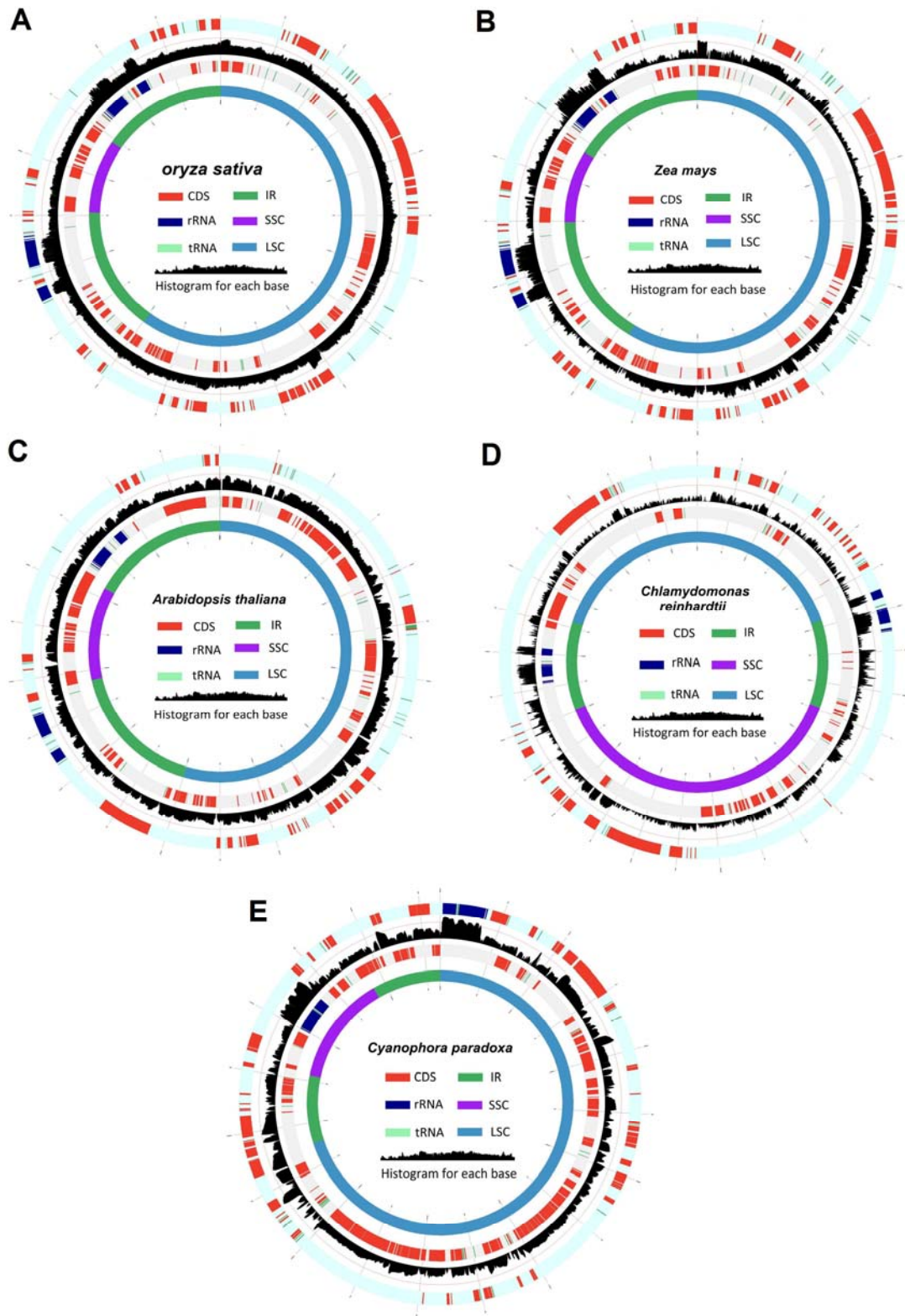
<sup>2</sup>Faculty of Life Science and Technology, Kunming University of Science and Technology, Kunming 650093, China

<sup>3</sup>University of the Chinese Academy of Sciences, Beijing 100039, China

\*These authors contribute equally to this work.

\*\*To whom correspondence should be addressed. Tel./Fax: +86-871-5223277, Email:

Lgao@mail.kib.ac.cn



**Supplementary Figure 1. Full transcription of the photosynthetic eukaryote chloroplast genome.** This is an enlarged version of the map in **Figure 1A and B**. The outer and third tracks represent the plastome genes, the inner track shows the four

genomic regions of the plastome, and the black histogram of the second track represents RNAseq reads mapping (scale  $\log_{10}$ -transformed numbers of sequence reads per nucleotide).

**Supplementary Table 1. The representative transcription units of chloroplast genes in higher plants.**

---

<b>Representative transcript units in higher plants*</b>	
1	<i>atpB-atpE</i>
2	3' <i>rps12-rps7</i>
3	<i>psbE-psbF-psbL-psbJ</i>
4	<i>psbD-psbC-orf62</i> (in dicots)
5	<i>psbK-psbI-psbD-psbC-orf62-trnG</i> (in monocots)
6	<i>ndhC-ndhK-ndhJ</i>
7	16S <i>rDNA-trnI-trnA</i> -23S <i>rDNA</i> -4.5S <i>rDNA</i> -5S <i>rDNA</i>
8	<i>rpoB-rpoC1-rpoC2</i>
9	<i>rpl23-rpl2-rps19-rpl22-rps3-rpl16-rpl14-rps8-infA-rpl36-rps11-rpoA</i>
10	<i>trnE-trnY-trnD</i>
11	<i>clpP-rps12-rpl20</i>
12	<i>orf31-petG-psaJ-rpl33-rps18</i>
13	<i>psaA-psaB-rps14</i>
14	<i>psaC-ndhD</i>
15	<i>psbB-psbH-petB-petD</i>
16	<i>psbK-psbI-trnG</i>
17	<i>ndhA-ndhI-ndhG-ndhE-psaC</i>
18	<i>rpl32-sprA</i>
19	<i>rps2-atpI-atpH-atpF-atpA</i>

---

\*Only major transcription units were listed. Please also see ref (12, 13).

**Supplementary Table 2. Summary of sequence reads used in this study.**

<b>Species</b>	<b>Tissues</b>	<b>Sequence data</b>	<b>Reference</b>
<i>O. sativa</i>	seedling shoots	ERA000212	(43)
<i>A. thaliana</i>	seedlings and flowers	SRA047499	(44)
	immature floral tissue (directional sequencing)	SRA000286	(45)
<i>Z. mays</i>	leaves	SRA050928	Not available
<i>C. reinhardtii</i>	whole cell	SRA026231	Not available
<i>C. paradoxa</i>	whole cell	SRA047914	(46)
<i>Prochlorococcus marinus</i> CCMP1986	whole cell	SRA050651	Not available
<i>Synechococcus</i> PCC7002	whole cell	SRA044901	Not available
<i>Synechocystis</i> PCC6803	whole cell	SRA025999	Not available

**Supplementary Table 3. Statistics of RNA-seq reads mapping with the chloroplast genomes.**

Species	Reads used for mapping	Average reads depth	Genome coverage (Bowtie) *	Genome coverage (PASS) *
<i>A. thaliana</i>	224,829,363	480	<b>98.76%</b>	<b>99.18%</b>
<i>A. thaliana</i> (F) <sup>b</sup>		503	<b>93.33%</b>	<b>94.12%</b>
<i>A. thaliana</i> (R) <sup>b</sup>	133,741,703	534	<b>94.25%</b>	<b>94.88%</b>
<i>O. sativa</i>	149,778,844	47,875	<b>99.91%</b>	<b>99.92%</b>
<i>Z. mays</i>	587,912,045	6,826	<b>98.53%</b>	<b>99.56%</b>
<i>C. reinhardtii</i>	575,401,566	988	<b>91.41% (99.34%)<sup>a</sup></b>	<b>90.73% (99.16%)<sup>a</sup></b>
<i>C. paradoxa</i>	119,434,879	589	<b>98.23%</b>	<b>98.92%</b>
<i>Prochlorococcus marinus</i> CCMP 1986	59,735,881	43	<b>94.33%</b>	<b>94.34%</b>
<i>Synechococcus</i> PCC 7002	195,192,665	78	<b>99.27%</b>	<b>99.27%</b>
<i>Synechocystis</i> PCC 6803	459,631,042	758	<b>99.99%</b>	<b>99.98%</b>

Note: \*All filtered reads were mapped using Bowtie or PASS.

<sup>a</sup> Genome coverage was calculated after trimming repeat sequences from the *C. reinhardtii* genome.

<sup>b</sup> Strand-specific sequencing data were mapped to forward (F) and reverse strand (R) of the chloroplast genome.

**Supplementary Table 4. Primers used to examine intergenic region transcription in the rice plastome.**

<b>Primer</b>	<b>Primer sequence (5'-3')</b>	<b>Position</b>	<b>Region</b>
1F	GCCGAGTACTCTACCATTGA	3902	<i>tRNA-Lys</i>
1R	CGTATTGTATTCTGCACGAGA	6567	<i>tRNA-Gln</i>
2F	GTTTGGGAAAGTCCGAAGAA	6921	<i>psbK</i>
2R	ACCATTTCCGAGCACCTTAT	8855	<i>psbD</i>
3F	TTTCTGTTCTGGCTCGGT	11364	<i>psbC</i>
3R	CGTCTCCGCACTTAGCAGT	12832	<i>tRNA-Met</i>
4F	CGGACGCCTTAAGTATATCA	13677	<i>tRNA-Gly</i>
4R	TCGCCTGTCCTACTCTAAGA	16159	<i>tRNA-Tyr</i>
5F	GGCAGGAAGCAAGATCAAATA	16362	<i>tRNA-Asp</i>
5R	AACGTGGATGAATACTTGCTC	18158	<i>tRNA-Cys</i>
6F	GCAAGTATTCATCCACGTTTC	18160	<i>tRNA-Cys</i>
6R	CGTCAAGTATACCCGTTTCAA	20101	<i>rpoB</i>
7F	GTCATTGGGTTAAGAGAAGTTCCGATTG	24,319	<i>rpoc1</i>
7R	GATGATACTGTAATGGAATGGCGAACC	26,456	<i>rpoc2</i>
8F	GATTCGCGATTTCCCTAAGA	27700	<i>rpoc2</i>
8R	AATGACAACGAGCCCTTATT	29804	<i>rps2</i>
9F	TATAGCTCTTCGGGAATGT	30062	<i>rps2</i>
9R	TTGTCCTGCAGCAGTACCTTGACCAACT	32114	<i>atpH</i>
10F	TCCCAGCCCTGATACCAAT	43344	<i>ycf3</i>
10R	GGCCATCTCTCCTCCATAATC	44430	<i>tRNA-Ser</i>
11F	GATTACTTCCGATTTAGGTG	45323	<i>rps4</i>
11R	ATCTACATTCCTTGCCGATA	47274	<i>tRNA-Leu</i>
12F	CCAATAATTTGTCAGGCGACT	50878	<i>tRNA-Val</i>
12R	AGTAGGGAACGGACTCGTGA	52689	<i>atpB</i>
13F	GGAGCTCCCGTGTCAATCACT	53022	<i>atpB</i>
13R	GAAGCCACCGGTATAACACCT	55211	<i>rbcL</i>
14F	CCGGTAGATAAACTAGATAGC	55,506	<i>rbcL</i>
14R	CGCCAACCTAGATACTCCT	57,598	<i>psaI</i>
15F	CCTTTACCGCGCATTATTG	56638	<i>psaI</i>
15R	GATTGGATATCCCGCATAAG	58063	<i>ycf4</i>
16F	GCGGGATATCCAATCAATTAG	58068	<i>ycf4</i>
16R	GGGAACAGGGCCTATCACAA	60044	<i>petA</i>
17F	GTGCTTCAGGATCCATTACGC	60436	<i>petA</i>
17R	GAACTATGACACAATCAAACCCGAATG	61910	<i>psbL</i>
18F	TGATCCTTTCTTTCTCCTCGCTTCAT	61226	<i>psbJ</i>
18R	AGTCCCTAATTCCGAATCAAG	63345	<i>psbE</i>
19F	CCTGTGACATTTTGTACCC	64250	<i>tRNA-Trp</i>
19R	ACGAGCCTGTTTTATAGCA	65914	<i>rps18</i>

---

20F	TTCTTCTTCCTCTTCGGTTTC	65110	<i>rpl33</i>
20R	TGCCGTCACGTGTATGGATAA	67070	<i>rpl20</i>
21F	CGCCATTCTGAGATTAAC	68053	<i>ClpP</i>
21R	CACACACTCATATTCCAGAGATACCGAAAC	71214	<i>psbH</i>
22F	GTTTCGGTCGGTCTTATTGTA	77899	<i>rpl14</i>
22R	TGAGAAATTGACTCGGAAAG	78704	<i>rpl16</i>
23F	GGCCTGTTATCTCTATCAAGA	71283	<i>psbH</i>
24R	CCCCTCAATATTGCAAGTTT	73298	<i>petD</i>
25F	GACCTTGCGGATCTGTAACA	75671	<i>rps11</i>
26R	TGAGAAATTGACTCGGAAAG	78704	<i>rpl16</i>
27F	AGGTTTGAATCTCGCAATGG	78668	<i>rpl16</i>
28R	AGAAGAAGAGTGGGCTAAGGA	80506	<i>rpl22</i>
29F	ATAGTACTCCTACTGACTTCGGCTTTAGTG	80223	<i>rps3</i>
29R	GTAACGGCAGCAAGTGATTGA	82128	<i>rpl2</i>
30F	TCTTCTCTCCATCGGAACAA	87396	<i>ndhB</i>
30R	TTATTCGAGGAGCCCTAGAT	89080	<i>rps12</i>
31F	GGGCGTGGAAACAGATAGAA	90548	<i>orf71</i>
31R	GTTTACGGCTAGGACTACTGG	92044	<i>rrn16</i>
32F	ATGCAAAGCGAAGAACCTTAC	92212	<i>rrn16</i>
32R	GGTCTGTGAAGATGCGTTGTT	93491	<i>ycf68</i>
33F	GGGTGAGTCAGGGCCTAAGAT	96581	<i>rrn23</i>
34R	ATAGAGTCCGACCGCAACGAC	99684	<i>tRNA-Asn</i>
35F	ATGAGTCTACCGCTTACAAGA	101230	<i>rps15</i>
35R	GTCTATCCTCATGAACCGGGAAATACTATG	102016	<i>ndhF</i>
36F	GACACCCTTGTCGAAAACTGCGTTAGT	101578	<i>rps15</i>
36R	GGGGTTGGTTACTAGTTCC	103213	<i>ndhF</i>
37F	AGCCCATATGCGACGAAGAT	103516	<i>ndhF</i>
37R	TGGGTAATGTAAATGAACCAT	106177	<i>ccsA</i>
38F	GCTCGTACACATTGAGTGCAT	108208	<i>psac</i>
38R	CGGGTGAATTTGTCTATT	110875	<i>ndhI</i>
39F	TGCCACAGCTACAAAGTAGGA	109576	<i>ndhE</i>
39R	ATTGGTCTTCTTATGGCAGGATATAGCTCA	112255	<i>ndhA</i>
40F	GAGCTATATCCTGCCATAAGA	112256	<i>ndhA</i>
40R	GCGGTAATTTTCATCGTTC	114018	<i>rps15</i>

---



**Supplementary Table 5. Tissue-specific transcriptome reads mapping in the rice plastome.**

**A. Transcriptome reads from seven different tissues.**

<b>Tissues</b>	<b>Normalized reads*</b>	<b>Mapped reads</b>	<b>Coverage (%)</b>
<b>Root</b>	38,903,554	7,780 (0.02%)	32.1
<b>Callus</b>	38,903,554	23,342 (0.06%)	35.1
<b>Seed</b>	38,903,554	151,723 (0.39%)	73.1
<b>PAAF</b>	38,903,554	482,404 (1.24%)	77.1
<b>PABF</b>	38,903,554	567,991 (1.46%)	76.8
<b>Leaf</b>	38,903,554	887,001 (2.28%)	73.6
<b>Shoot</b>	38,903,554	1,015,383 (2.61%)	78.4

**B. Transcriptome reads from four different tissues and with higher in-depth sequencing.**

<b>Tissues</b>	<b>Normalized reads*</b>	<b>Mapped reads</b>	<b>Coverage (%)</b>
<b>Root</b>	52,627,968	63,154 (0.12%)	72.6
<b>PABF</b>	52,627,968	931,515 (1.77%)	89.4
<b>Leaf</b>	52,627,968	1,405,166 (2.67%)	99.3
<b>Shoot</b>	52,627,968	1,505,159 (2.86%)	99.7

\*All reads were normalized to contain same number of reads by randomly extraction from different tissues.

**Supplementary Table 7. *De novo* chloroplast genome assembly from transcriptome data.**

<b>Taxa</b>	<b>Species</b>	<b>Data</b>	<b>Plastome assemble length</b>
Bryophytes	<i>Pohlia nutans</i>	SRP012293	122,067
	<i>Ceratodon purpureus</i>	SRX032867	120,749
angiosperms	<i>Dioscorea villosa</i>	SRX060329	152,589
	<i>Chlorophytum rhizopendulum</i>	SRX116253	148,744
	<i>Echinacea purpurea</i>	SRX085116	151,084
	<i>Lactuca serriola</i>	SRX098217	162,202
	<i>Valeriana officinalis</i>	SRX047631	152,768
	<i>Scabiosa columbaria</i>	SRX090087	162,104
	<i>Beta vulgaris</i>	ERX026027	150,673
	<i>Cajanus cajan</i>	SRA030523	120,869
	<i>Corylus avellana</i>	SRX154512	156,977
	<i>Hypericum perforatum</i>	SRX062064	156,712
	<i>Pomegranate peel</i>	SRX034876	159,871
	<i>Musa cavendish</i>	SRX146849	169,992

**Supplementary Table 8. Tissue-specific small RNA-seq reads mapping in rice plastome.**

<b>Tissues</b>	<b>Normalized reads*</b>	<b>Mapped reads</b>	<b>Coverage (%)</b>
<b>Root</b>	15,004,156	1,757 (0.01%)	6.3
<b>PABF</b>	15,004,156	11,436 (0.07%)	25.0
<b>Leaf</b>	15,004,156	420,375 (2.80%)	91.5
<b>Shoot</b>	15,004,156	574,317 (3.82%)	94.2

\*All reads were normalized to contain same number of reads by randomly extraction from different tissues.

## REFERENCES

43. T. Lu *et al.*, Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq. *Genome Research* **20**, 1238-1249 (2010).
44. Y. Marquez, J. W. Brown, C. Simpson, A. Barta, M. Kalyna, Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. *Genome Research* **22**, 1184-1195 (2012).
45. R. Lister *et al.*, Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**, 523-536 (2008).
46. D. C. Price *et al.*, *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. *Science* **335**, 843-847 (2012).