

Table S5. The intron statistics in red algal and viridiplantae genomes. Only introns flanked by coding exons are included. Coding DNA sequences (CDSs) with in-frame stop codon, ambiguous nucleotide or incomplete codon (length is not divisible by 3) were removed. Frequencies of the two most prevalent splice sites found in most species (GT-AG and GC-AG) are shown for each species.

Species	Species	CDS with intron(s)*	Intron number	GT-AG	GC-AG
Rhodophyta	<i>G. sulphuraria</i>	4655	13,245	98.5%	1.4%
Rhodophyta	<i>C. merolae</i>	26	27	96.3%	3.7%
Rhodophyta	<i>P. purpureum</i>	201	245	94.7%	/
Rhodophyta	<i>G. chorda</i>	1815	2249	96.1%	/
Rhodophyta	<i>C. crisps</i>	1602	2926	95.8%	2.6%
Rhodophyta	<i>P. yezeonsis</i>	2520	3500	95.4%	0.05%
Viridiplantae	<i>C. reinhardtii</i>	11386	93749	98.5%	0.4%
Viridiplantae	<i>K. flaccidum</i>	16306	94296	98.4%	1.3%
Viridiplantae	<i>P. patens</i>	27765	143729	97.6%	2.2%
Viridiplantae	<i>S. moellendorffii</i>	29017	161879	99.3%	0.3%
Viridiplantae	<i>A. thaliana</i>	40106	236649	98.6%	1.1%

*includes multiple splice variants derived from single gene locus for Viridiplantae.