博士論文

Evolution of the Shine-Dalgarno Interaction

(Shine-Dalgarno相互作用の進化)

林 庚澤

Table of Contents

1. Abstract

1.1. Background

For translation of mRNA to occur, ribosome has to recognize the correct start codon. The site recognition mechanism differs considerably between prokaryotes and eukaryotes. The most well-known translation initiation site recognition mechanism in prokaryotic systems is the Shine-Dalgarno (SD) interaction. This interaction is mediated by the base pairing between rRNA and mRNA at their particular regions. A pyrimidine-rich, anti-SD sequence in the 3′ tail of a small subunit rRNA forms base pairing with a complementary, purine-rich, SD signal sequence in the 5′ untranslated region (UTR) of an mRNA. A core motif (i.e., the anti-SD motif), 3′CCUCC, is conserved among anti-SD sequences. This motif's extreme evolutionary constraint suggests a crucial role for SD interaction.

The anti-SD motif is so far known to be universally present in prokaryotes, suggesting the universality of SD interaction. SD sequences are not necessarily found in protein-coding genes in prokaryotes, showing considerably diverse usage among species. Because organelles of prokaryotic origin, mitochondria and plastids, originated from bacteria, SD interactions should have been present in their early endosymbiotic stages. The interaction seems still widely used in plastids, while used only in rare bacteria-like mitochondria.

Exponentially increased genome data now provide an unprecedented chance to obtain more detailed understanding of SD interactions in various taxonomic groups. Here I

conducted a large-scale analysis of all available complete genome sequences of bacteria and plastids for SD interactions with emphasis on their alterations and losses.

1.2. Parallel Losses of Shine-Dalgarno Interactions in Bacteria.

Contradicting to the conventional belief that prokaryotes universality use SD interactions for translation initiation of some of their genes, I found 15 bacteria without the classical anti-SD motif (referred to as lost anti-SD bacteria) by investigating 1,081 bacterial genome sequences.

This loss was accompanied by that of SD sequences, suggesting that SD interaction no longer operates in lost anti-SD bacteria. Lost anti-SD bacteria emerged independently in α-Proteobacteria, β-Proteobacteria, γ-Proteobacteria, Flavobacteria, and Mycoplasma. Many of lost anti-SD genomes belonged to obligate host-associated bacteria with highly reduced genomes (i.e., primary endosymbionts and mycoplasmas). The evolutionary forces toward massive gene/function loss during a period of host association may have brought about loss of important but non-essential regulatory functions such as SD interaction. A-rich motifs at the corresponding areas of the SD sequences were found in all Flavobacteria regardless of the conservation of SD interaction. This motif probably mediates an unknown translation initiation mechanism by which SD interactions have been replaced. Among Mycoplasma species, only those belonged to a subgroup that infects red blood cells showed this loss.

1.3. Parallel Losses and Alterations of Shine-Dalgarno Interactions in Plastids.

My research hereinabove reported the rare loss of SD interactions in several bacterial lineages. Many of these have been forming obligate association with eukaryotic cells. Such association is also seen in organelle genomes of prokaryotic origin, mitochondria and plastids. I attempted to understand what happened to SD interactions during the evolution of a cyanobacterial endosymbiont into modern plastids. I analyzed available complete plastid genome sequences (n = 429) to reveal that the majority of plastids retained SD interactions but with varying levels of usage in their protein-coding genes. Losses of SD interactions took place independently in plastids of Chlorophyta, Euglenophyta, and Chromerida/Apicomplexa lineages. I discovered that the canonical SD interaction (3′CCUCC/5′GGAGG (rRNA/mRNA)) was replaced by an altered SD interaction (3′CCCU/5′GGGA or 3′CUUCC/5′GAAGG) through coordinated changes in the sequences of the core rRNA motif and its paired mRNA signal in plastids of Chlorophyta and Euglenophyta. This rRNA-mRNA coevolution proceeded intermediate steps that permitted both the canonical and altered SD interactions, so that detrimental effects by the motif transition on the cells can be minimized. This coevolutionay phenomenon demonstrates unexpected plasticity in the translation initiation machinery.

1.4. Significance

This study demonstrates evolutionary plasticity of SD interactions by discovering their parallel losses (in bacteria and plastids) and alterations (in plastids) especially under

host-associated conditions. Furthermore, alterations in SD interactions were achieved by stepwise and coordinated changes in rRNA motif and its complementary mRNA signal. This represents, to the best of my knowledge, the first report of rRNA-mRNA coevolution. This coevolution caused unexpected plasticity in the translation initiation machinery, likely driving genome evolution by affecting all genes with mRNA signals.

2. Introduction

*This Introduction is written based on a published paper (Lim K, Furuta Y, Kobayashi I. 2012. Large variations in bacterial ribosomal RNA genes. Mol. Biol. Evol. 29:2937–2948.) and a submitted manuscript (Lim K, Kobayashi I, Nakai K. Alterations in rRNA-mRNA interaction during plastid evolution).

Translation is the process of synthesizing a protein using an mRNA as the template. Ribosome, formed by ribosomal RNAs (rRNA) and ribosomal proteins, serves for translation. Ribosome consists of two subunits; the small subunit recognizes mRNA, while the large subunit synthesizes peptide bonds. Translation is essential for all living organisms, so is genes for ribosomal subunits. Because many of these genes show relatively low evolutionary rates, sequence similarities among their orthologs are often clearly observed even for distantly related organisms. Therefore, rRNA genes and many ribosomal protein genes have been widely used as phylogenetic markers.

For translation of an mRNA to occur, the small subunit ribosome has to recognize the correct site of the start codon. The site recognition mechanism differs considerably between prokaryotes and eukaryotes (Malys and McCarthy 2011). In prokaryotes, the site recognition for initiates translation is often mediated by the Shine-Dalgarno (SD) interaction (Shine and Dalgarno 1974), rRNA-mRNA base pairing at particular regions.

A distinct base pairing rule of SD interactions is known; a pyrimidine-rich sequence (an anti-SD sequence) in the 3′ tail of a small subunit rRNA binds to a complementary,

purine-rich sequence (an SD sequence) in the 5′ untranslated region (UTR) of an mRNA (Fig. 1). A core motif (i.e., the anti-SD motif), 3′CCUCC, is conserved among anti-SD sequences (Ma et al. 2002; Nakagawa et al. 2010). This motif's extreme evolutionary constraint suggests a crucial role for SD interaction. Functional SD sequences should keep proper spacing, i.e., approximately 10 nt upstream from the start codon (Hirose and Sugiura 2004a; Chang et al. 2006), which has often been used for determining SD sequences (Nakagawa et al. 2010).

SD sequences are rarely seen in protein-coding genes in some prokaryotic genomes despite the conservation of classical anti-SD motif on their small subunit rRNA genes(Ma et al. 2002; Chang et al. 2006; Nakagawa et al. 2010). Loss of SD interaction acknowledged so far is the case of *Candidatus* Carsonella ruddii, which is a primary symbiont of insects, as sequencing of 16S–23S spacer regions revealed the loss of the classical anti-SD motif (Thao et al. 2000). This indicates that SD interaction is dispensable in some forms of living organisms and that there probably are other mechanisms for translation initiation site recognition. A well-known mechanism is direct translation of leaderless mRNAs (mRNAs with no or an extremely short 5′ UTR). In general, leaderless genes are widespread among prokaryotes, albeit not dominant (Zheng et al. 2011). For example, leaderless genes account for only 2.2% of the *Helicobacter pylori* protein-coding genes (Sharma et al. 2010). In an archaeon, *Halobacterium salinarum*, leaderless genes were translated more efficiently than the SD sequence carrying genes (Sartorius-Neef and Pfeifer 2004). This phenomenon seems not wide-

spread among prokaryotes as mRNAs with SD sequences show higher translational efficiency in an bacterium, *Escherichia coli* (Kosuri et al. 2013).

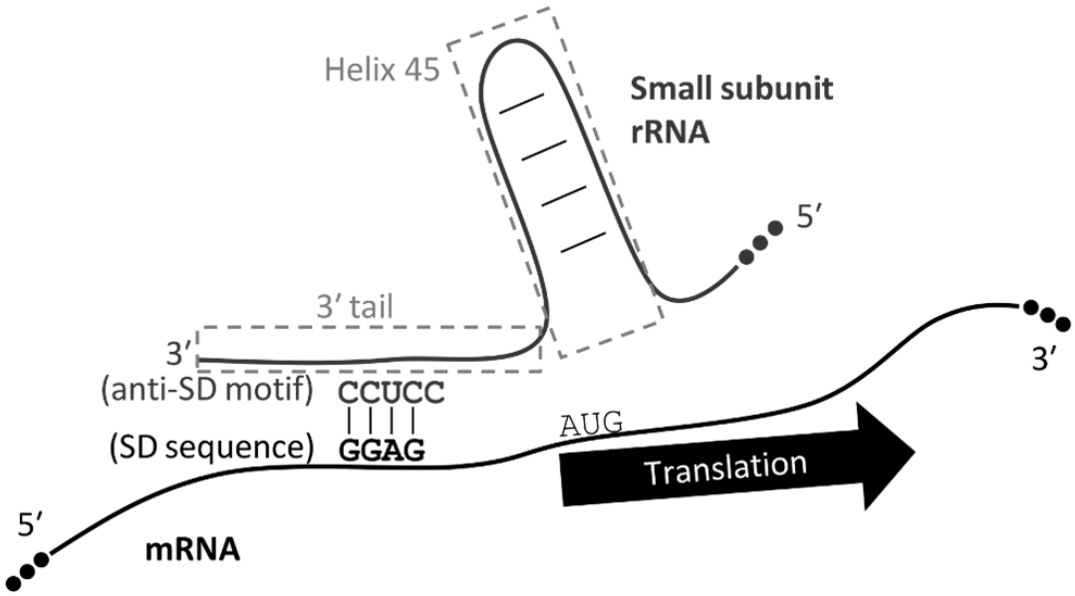


FIG. 1. The Shine-Dalgarno (SD) interaction for translation initiation in prokaryotes. The 3′ tail of rRNA in the small subunit of the ribosome recognizes a complementary sequence in the 5′ UTR of mRNA (i.e., the SD sequence) by rRNA-mRNA base paring. The 3′ tail contains a conserved motif (3′CCUCC), referred to as the anti-SD motif. This figure is adapted from Lim et al.(submitted)

Some evidence for SD interaction has been reported in several plastids (Bonham-Smith and Bourque 1989; Betts and Spremulli 1994; Hirose and Sugiura 2004a) and bacteria-like mitochondria (Lang et al. 1997; Hazle and Bonen 2007; Burger et al. 2013), supporting the endosymbiotic hypothesis that mitochondria and plastids (including chloroplasts) originated from bacterial endosymbionts (Sagan 1967; Gray and Doolittle 1982).

Plastids in photosynthetic eukaryotes evolved from a cyanobacterium by endosymbiosis (Sagan 1967; Gray and Doolittle 1982). Several endosymbiotic processes are responsible for the emergence of plastids in eukaryotes (Reyes-Prieto et al. 2007; Keeling 2013) (Fig. 2). First, direct endosymbiosis of a cyanobacterium with a eukaryote, known as primary endosymbiosis, resulted in plastids of the supergroup Archaeplastida (or Plantae). Archaeplastida includes Streptophyta (all land plants and a subgroup of green algae), Chlorophyta (a subgroup of green algae), Rhodophyta (red algae) and Glaucophyta (Yoon et al. 2004; Rodríguez-Ezpeleta et al. 2005). Second, endosymbiosis of Archaeplastida members with other eukaryotes, known as secondary endosymbiosis, propagated the formers' plastids: from Chlorophyta to Euglenophyta and Chlorarachniophyta; from Rhodophyta to Chromalveolata, Haptophyta and Cryptophyta (Janouskovec et al. 2010). Tertiary endosymbiosis and serial secondary endosymbiosis events involved plastids of Dinoflagellata (Tengs et al. 2000; Keeling 2013). One tertiary endosymbiosis event that originated the current plastids of the subgroup Dinophyceae was depicted in Fig. 2.

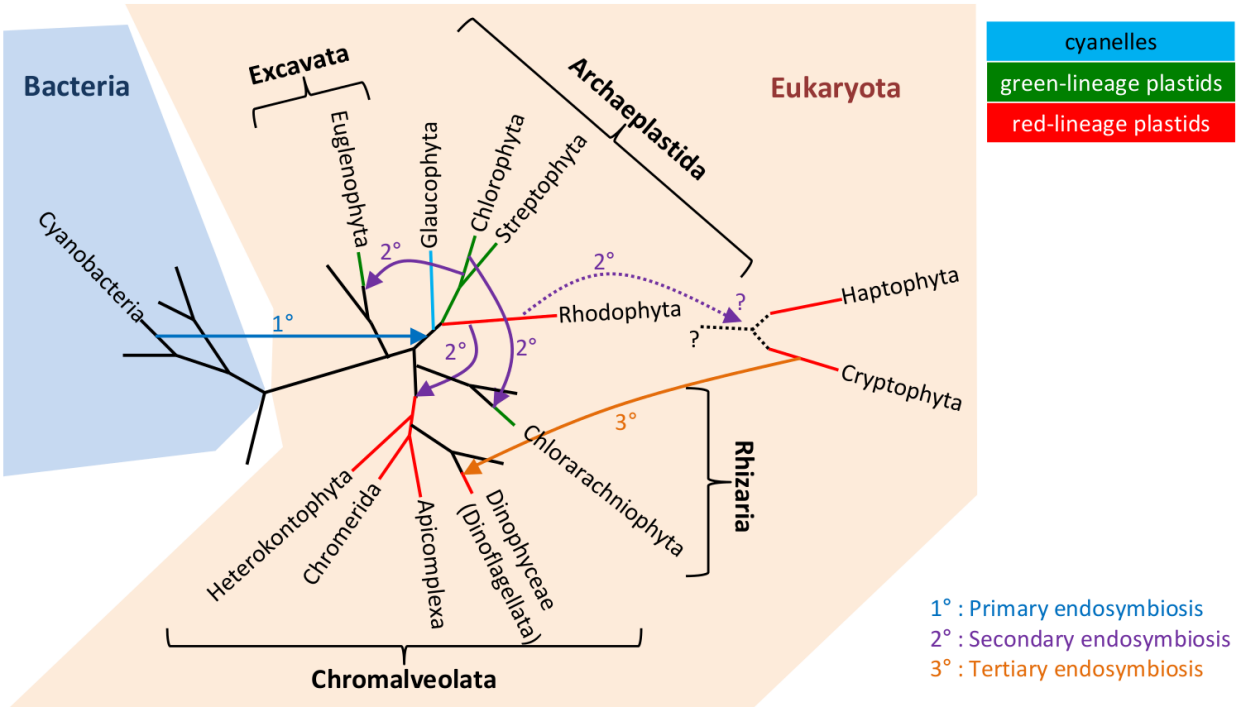


FIG. 2. Primary, secondary, and tertiary endosymbiosis events depicted on a diagram of bacterial and eukaryotic tree of life. A cyanobacterium has evolved to become the plastids of the eukaryotic supergroup Archaeplastida through primary endosymbiosis (labeled as 1°). Green and red lineage plastids within Archaeplastida have been transferred to other eukaryotic supergroups through secondary endosymbiosis (labeled as 2°). A Cryptophyta plastid was transferred to a subgroup of Dinoflagellata through tertiary endosymbiosis (labeled as 3°). For Haptophyta and Cryptophyta, phylogeny of them and their plastids has not been resolved, so they are shown as a dotted branch or an arrow. This figure is adapted from Lim et al.(submitted)

During their long history of endosymbiotic evolution, plastid genomes lost most of their genetic information, some of which were transferred to the nuclear genome (Martin et al. 2002; Timmis et al. 2004). Transferred genes even included genes coding for proteins that performed essential plastid functions such as translation (Gantt et al. 1991; Millen et al. 2001; Ueda et al. 2007; Jansen et al. 2011). This indicates that such essential gene product had to be transported back to the plastids, which is enabled by various transportation mechanisms (Agrawal and Striepen 2010). Unlike ribosomal protein genes, all rRNA genes for plastid ribosomes are present in plastid genomes. Some plastids even lost all genes for photosynthesis; nonphotosynthetic plastids have been reported in Streptophyta (dePamphilis and Palmer 1990; Delannoy et al. 2011; Logacheva et al. 2011), Chlorophyta (Boucias et al. 2001), Euglenophyta (Siemeister and Hachtel 1989), and Apicomplexa (Wilson et al. 1996).

In this study, I first aim to gain collective understanding of SD interactions in bacteria with emphasis on its losses. I show that some bacterial lineages no longer possess SD interactions. This loss was seen many bacteria under obligate association with eukaryotic host cells and some free-living bacteria only belonging to Flavobacteria. I next analyze plastid genomes to understand how SD interactions have evolved during an ancient history of endosymbiosis and genome reduction. I discovered that the anti-SD motif (rRNA) and the cognate SD signal (mRNA) have coevolved beyond the canonical SD interaction.

3. Part 1: Parallel Losses of Shine-Dalgarno Interactions in Bacteria

*This part is written based on a published paper (Lim K, Furuta Y, Kobayashi I. 2012. Large variations in bacterial ribosomal RNA genes. Mol. Biol. Evol. 29:2937–2948.).

3.1. Unusual 3′ Tail Sequences of Small Subunit rRNA Genes

Previous systematic analyses of bacterial genome sequences reported that evidence for SD interactions was observed in all surveyed genomes (Ma et al. 2002; Nakagawa et al. 2010), except for *Ca.* Carsonella ruddii, which is a primary endosymbiont of insects (Thao et al. 2000). Accelerated accumulation of complete genomic sequences motivated me to conduct a similar analysis of larger genomic data. I confirmed the loss of the canonical anti-SD motif not only in *Ca.* Carsonella ruddii but also in many other bacteria (Table 1). Fifteen among 1,182 complete genomes of bacteria did not carry the canonical anti-SD motif in any of their small subunit rRNA genes (I refer to the fifteen bacteria as lost anti-SD bacteria).

I categorized lost anti-SD bacteria into four groups (Table 1) in terms of phylogeny and life style. Group 1 consists of three bacteria with multiple small subuit rRNA genes belonging to the class Flavobacteria. Group 2 consists of six bacteria belonging to Flavobacteria and primary endosymbionts (obligate and mutualistic bacteria with an ancient history of host association) of insects (McCutcheon and Moran 2012). Group 3 bacteria were also primary endosymbionts of insects, but belonged to the phylum

Proteobacteria (McCutcheon and Moran 2012); Group 4 consists of three mycoplasmas living in red blood cells (hemotrophic mycoplasmas) (Guimaraes et al. 2011).

Table 1. Bacteria with unusual small subunit rRNA 3' tail sequences.

Group	Class	Strain	No. of SSU rRNA genes		Genomic size (nt)	Note
			Total	w/o anti-SD motif		
1	*Flavobacteria*	*Flavobacteriaceae bacterium* 3519-10	2	2	2768102	Psychrophile
	Flavobacteria	*Riemerella anatipestifer* DSM 15868	3	3	2155121	Pathogen of poultry
	Flavobacteria	*Weeksella virosa* DSM 16922	6	6	2272954	Isolated from human urine
2	*Flavobacteria*	*Blattabacterium* sp. str. Bge	1	1	636850	Primary endosymbionts of insects
	Flavobacteria	*Blattabacterium* sp. str. BPLAN	1	1	636994	
	Flavobacteria	*Candidatus* Sulcia muelleri CARI	1	1	276511	
	Flavobacteria	*Candidatus* Sulcia muelleri DMIN	1	1	243933	

	Flavobacteria	*Candidatus* Sulcia muelleri GWSS	1	1	245530	
	Flavobacteria	*Candidatus* Sulcia muelleri SMDSEM	1	1	276984	
3	*Alphaproteobacteria*	*Candidatus* Hodgkinia cicadicola Dsem	1	1	143795	
	Betaproteobacteria	*Candidatus* Zinderia insecticola CARI	1	1	208564	
	Gammaproteobacteria	*Candidatus* Carsonella ruddii PV	1	1	159662	
4	*Mollicutes*	*Mycoplasma haemofelis* str. Langford 1	1	1	1147259	Hemotrophic mycoplasmas
	Mollicutes	*Mycoplasma suis* str. Illinois	1	1	742431	
	Mollicutes	*Mycoplasma suis* KI3806	1	1	709270	

* This table is adapted from Lim et al. (2012)

In phylogenetic trees based on full-length small subunit rRNA genes of these lost anti-SD bacteria and other reference bacteria, Groups 1 (Fig. 3A), 2 (Fig. 4A), and 4 (Fig. 5A) clustered into distinctive clades, individually. Each of these groups possibly shares a history of the anti-SD motif loss. Members of Group 1 share a variation: the anti-SD motif, 5′CCTCC, was changed to 5′TCTCA (Fig. 3B). In the other groups, the variation was diverse within a group. In Group 2, the classical anti-SD motif, 5′CCTCC, was changed to 5′TCTCT or 5′TTTCT (Fig. 3B). There is divergence even within the *Candidatus* Sulcia muelleri: 5′TCTCT from CARI and SMDSEM and 5′TTTCT from DMIN and GWSS. Group 3 featured extensively degenerated 3′ tail sequences; 5′CCTCC was changed to 5′TTTGA, 5′CATTT, or 5′TTTTT (Fig. 4B). In Group 4, *Mycoplasma haemofelis* has a degenerated sequence, 5′TCTTC, and the two *Mycoplasma suis* strains have 5′CTTTT, instead of the classical anti-SD motif (Fig. 5B).

The conserved C-rich characteristic of the classical anti-SD motif possibly are pivotal for firm rRNA–mRNA binding by forming C-G hydrogen bonds stronger than the A-U bonds (Freier et al. 1986). The degeneration of the anti-SD motif, mentioned above, predominantly resulted in substitutions from C to T or A, thereby likely hampering the binding capability of anti-SD sequences to SD sequence.

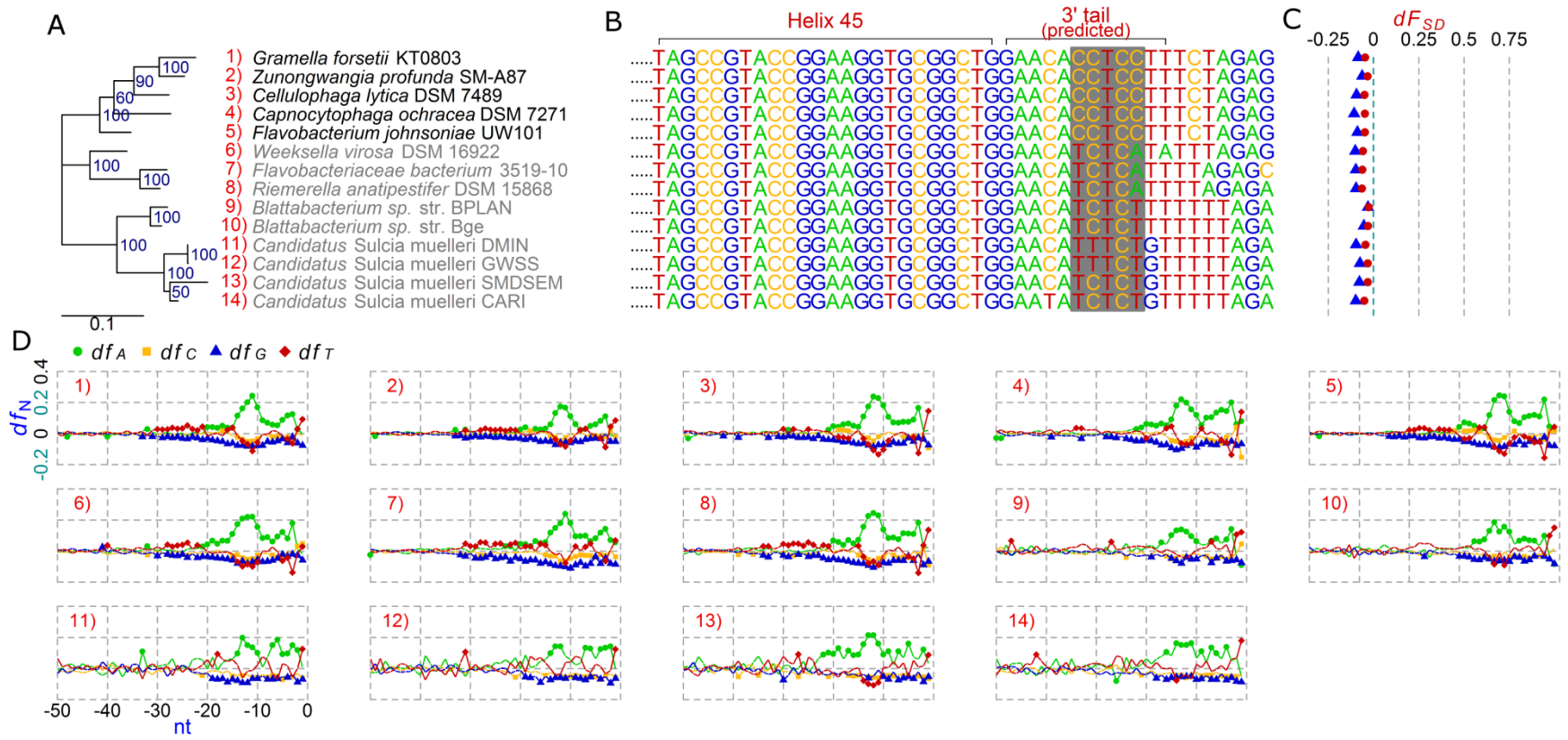


FIG. 3. Comparative analysis of SD interactions in Flavobacteria.(A) Maximum likelihood phylogenetic tree based on small subunit rRNA gene sequences. Lost anti-SD bacteria are shown in gray. (B) Predicted small subunit rRNA 3′ tail sequences. The regions corresponding to the anti-SD motif are shaded in gray. (C) SD indexes (dF_{SD}). Triangle: cutoff value < 3.4535 kcal/mol; Dot: cutoff value < 4.4 kcal/mol. (D) Fractions of the four nucleotides, df_N (N = A, C, G, or T), at specific positions between -50 and -1 nt from start codons in each genome. The background fraction was subtracted. This figure is adapted from Lim et al. (2012).

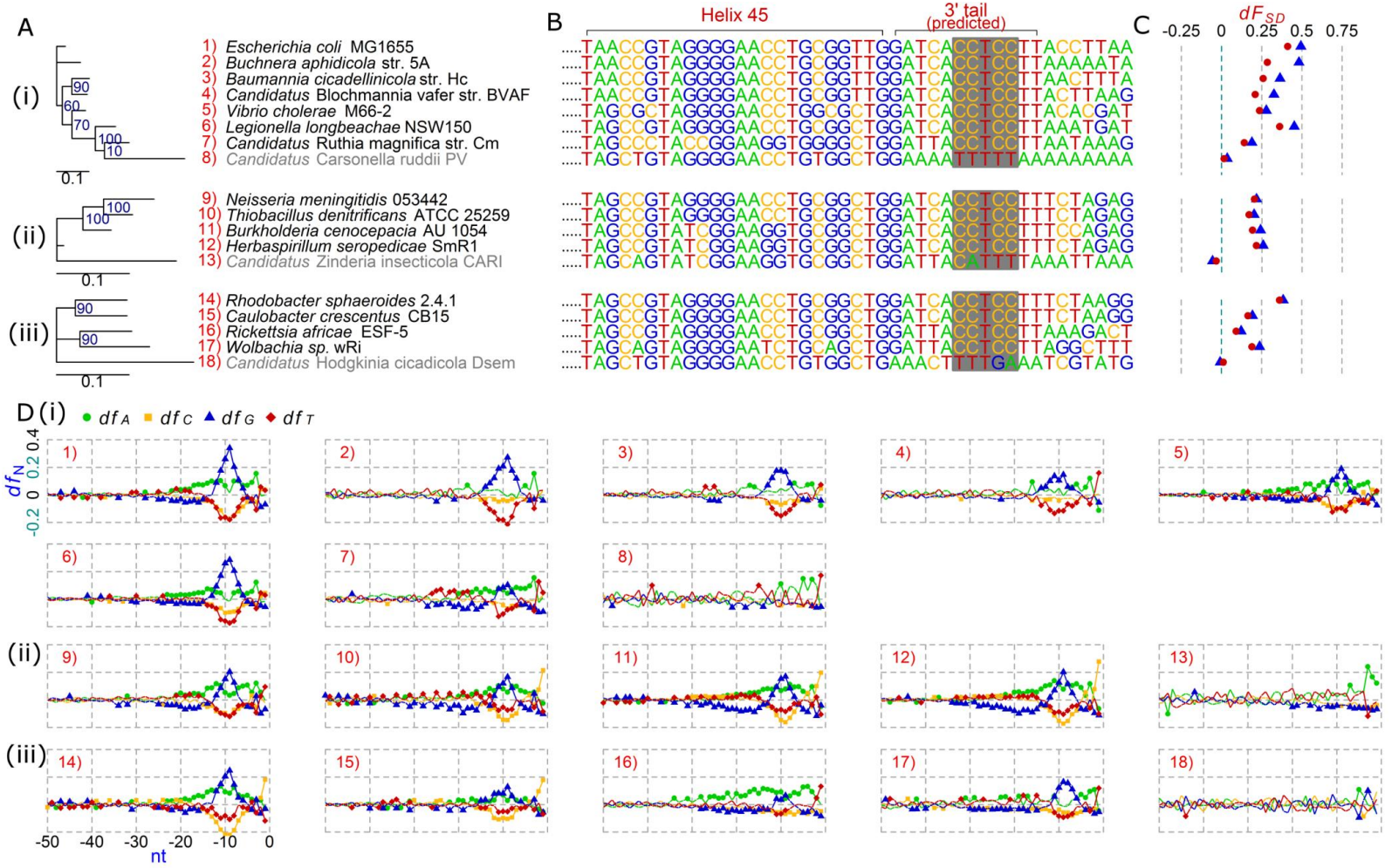

A
(i)
(ii)
(iii)
90
60
70
100
10
0.1
1) Escherichia coli MG1655
2) Buchnera aphidicola str. 5A
3) Baumannia cicadellinicola str. Hc
4) Candidatus Blochmannia vafer str. BVAF
5) Vibrio cholerae M66-2
6) Legionella longbeachae NSW150
7) Candidatus Ruthia magnifica str. Cm
8) Candidatus Carsonella ruddii PV
9) Neisseria meningitidis 053442
10) Thiobacillus denitrificans ATCC 25259
11) Burkholderia cenocepacia AU 1054
12) Herbaspirillum seropedicae SmR1
13) Candidatus Zinderia insecticola CARI
14) Rhodobacter sphaeroides 2.4.1
15) Caulobacter crescentus CB15
16) Rickettsia africae ESF-5
17) Wolbachia sp. wRi
18) Candidatus Hodgkinia cicadicola Dsem
B
Helix 45
3' tail
(predicted)
C
dF_SD
-0.25 0 0.25 0.5 0.75
D (i)
df_A df_C df_G df_T
df_N
0.4 0.2 0 -0.2
-50 -40 -30 -20 -10 0
nt

FIG. 4. Comparative analysis of SD interactions in Proteobacteria.(A) Maximum likelihood phylogenetic tree. Lost anti-SD bacteria are shown in gray. (B) Predicted small subunit rRNA 3′ tail sequences. The regions corresponding to the anti-SD motif are shaded in gray. (C) SD indexes (dF_{SD}). Triangle: cutoff value < 3.4535 kcal/mol; Dot: cutoff value < 4.4 kcal/mol. (D) Fractions of the four nucleotides, df_N (N = A, C, G, or T), at specific positions between -50 and -1 nt from start codons in each genome. The background fraction was subtracted. (i) Gammaproteobacteria. (ii) Betaproteobacteria. (iii) Alphaproteobacteria. This figure is adapted from Lim et al. (2012).

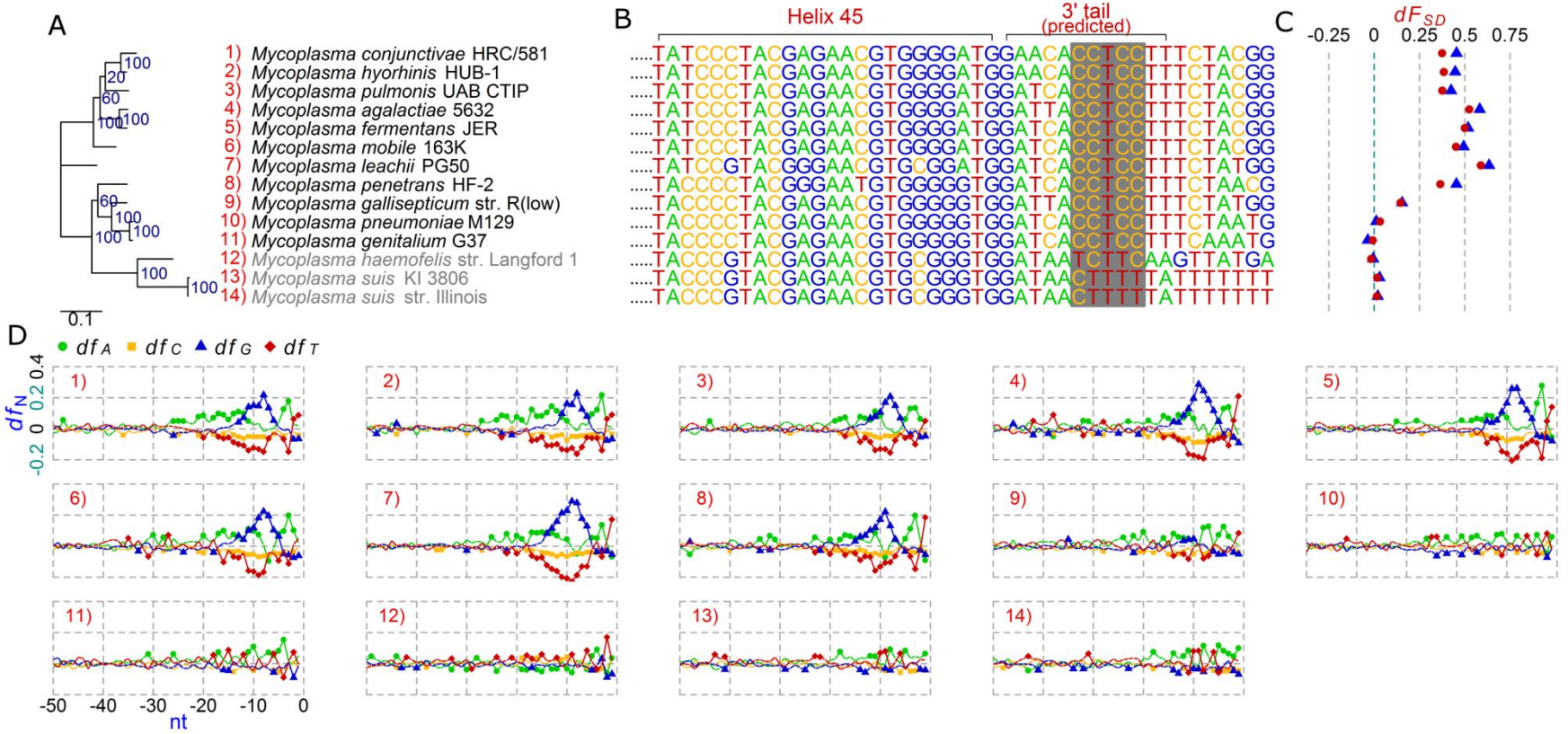


FIG. 5. Comparative analysis of SD interactions in Mycoplasma.(A) Maximum likelihood phylogenetic tree. Lost anti-SD bacteria are shown in gray. (B) Predicted small subunit rRNA 3′ tail sequences. The regions corresponding to the anti-SD motif are shaded in gray. (C) SD indexes (dF_{SD}). Triangle: cutoff value < 3.4535 kcal/mol; Dot: cutoff value < 4.4 kcal/mol. (D) Fractions of the four nucleotides, df_N (N = A, C, G, or T), at specific positions between -50 and -1 nt from start codons in each genome. The background fraction was subtracted. This figure is adapted from Lim et al. (2012).

3.2. Lack of SD Sequences in Genomes without the Classical Anti-SD Motif

Next, I analyzed SD sequences in lost anti-SD bacteria. A widely accepted strategy to determine SD-like sequences is to test whether the SD region (usually defined as the region -20 to -5 nt from the start codon) of a gene is able to form a RNA-RNA duplex with the host's small subunit rRNA 3′ tail (Schurr et al. 1993; Ma et al. 2002; Starmer et al. 2006; Nakagawa et al. 2010). I used FREE_SCAN, which is a tool for finding the minimum free energy (ΔG) RNA-RNA structure from given two RNA strands (Starmer et al. 2006). I set two cutoff ΔG values, -3.4535 and -4.4 kcal/mol, following earlier studies (Ma et al. 2002; Starmer et al. 2006) for determining potential SD sequences. For each cut off value, I measured the gene fraction carrying the potential SD sequences among all protein-coding genes for a given genome by the equation: F_{SD} = "Number of protein-coding genes with the SD-like sequences" / "Number of total protein-coding genes". I regarded F_{SD} as a proxy for intragenomic SD interaction usage. Based on a previous study (Nakagawa et al. 2010), I used another index, which was calculated by F_{SD} after substituting the background SD fraction in random artificial sequences generated based on its background nucleotide fraction (see Materials and Methods for details). The adjusted value (dF_{SD}) thus indicates a gene fraction with the SD interaction relative to a fraction with random genomic region/anti-SD interaction. In other words, a dF_{SD} value < 0 indicates that fewer 5′ UTRs than random genomic regions have capacities for binding to small subunit rRNA 3′ tails.

I applied another method that directly showed nucleotide bias in the 5′ UTR by calculating changes in the nucleotide fraction (df_N; N = A, C, G, or T) at specific positions in the 5′ UTR. A standard SD signal is the G enrichment in the SD region because the classical anti-SD motif is C-rich, as clearly seen in the *E. coli* SD regions (Fig. 4D(i)).

Signal intensities of the SD sequences have not been studied in Groups 1 and 2 lost anti-SD bacteria belonging to the class Flavobacteria (Table 1). I found that members of this class showed dF_{SD} values < 0 and mean ΔG values > -1 kcal/mol with the standard deviations ranging from 1.4 to 2.12, regardless of the conservation of an classical anti-SD motif (Fig. 3C). This suggests that Flavobacteria rarely use SD interactions for translation initiation, and that some members lost their classical anti-SD motifs no longer allowing SD interactions.

This may be due to A-rich signals at the corresponding areas of the SD region in all surveyed Flavobacteria (Fig. 3D), as opposed to the G-rich signal for SD interactions. In *E. coli*, ribosomal protein S1 contributes to translation initiation complex formation through its high affinity to AU-rich regions often observed on 5′ UTRs (Draper et al. 1977; Boni et al. 1991; Sengupta et al. 2001; Salah et al. 2009). The protein seems to assist binding between mRNA and the ribosome together with SD interaction, but it was dispensable when the SD interaction was strong (Farwell et al. 1992). Although the ability of ribosomal protein S1 to initiate translation by itself without SD interaction remains to be determined, an A-rich stretch may aid ribosomal protein S1 to bind and assist in translation initiation without an SD interaction.

It is also conceivable that an A-rich strand forms a RNA-RNA duplex with a U-rich strand in rRNA. A plausible U-rich region is a stretch of U-rich sequences right after the anti-SD sequence (Fig. 3B). Because our small subunit rRNA 3′ tail prediction was based on sequence comparison with the reference *E. coli* small subunit rRNA, Flavobacteria small subunit rRNA 3′ tails may be longer than the predicted length. Accurate annotation of small subunit rRNA ends for all bacteria remains to be achieved, which is important to more correctly understand anti-SD motif degeneration processes.

Figure 4 shows the comparison analysis of Group-3 lost anti-SD bacteria with several reference genomes in the phylum Proteobacteria. Diverse dF_{SD} values (range: 0.09–0.41 with a cutoff value of -4.4 kcal/mol) are seen in the reference genomes, whereas those near 0 were observed in Group-3 lost anti-SD bacteria (Fig. 4C). This result was in agreement with the nucleotide bias analysis (Fig. 4D): G enrichment were seen in the references but not in Group 3 lost anti-SD bacteria. Mean ΔG values of Group 3 lost anti-SD bacteria (range: -1.43 to -0.73) were higher than those of the references (range: -5.81 to -1.78) (Table S1). The ΔG standard deviation values more clearly distinguish Group 3 (range: 1.07–1.60) from the references (range: 2.64–3.18) (Table S1). Many reference bacteria shown in Fig 4 were host-obligate bacteria with similar features, in terms of host association, to Group 3 members. For example, *Buchnera aphidicola*, *Baumannia cicadellinicola*, and *Candidatus* Blochmannia vafer are primary endosymbionts of insects, as are all Group 3 strains (McCutcheon and Moran 2012). Our result supports that SD interactions were lost in Group 3 members and that being a primary endosymbiont of insects is not necessarily indicative of SD interaction loss.

The phylum Mollicutes, which includes the genus Mycoplasma, is diverse in SD interactions usage (Nakagawa et al. 2010). Three hemotrophic mycoplasmas (Group 4) that lost classical anti-SD motifs (Fig. 5B) did not indicate SD interactions; there were no G-rich signals or any nucleotide enrichments within the SD regions (Fig. 5C and D). Among mycoplasmas with the classical anti-SD motif, *Mycoplasma genitalium* and *Mycoplasma pneumonia*, which are phylogenetically closely related (Fig. 5A), showed no G enrichments within the SD regions and had very low dF_{SD} (Fig. 5C and D). The ΔG standard deviation values (range: 2.60–3.47) of these two strains, however, were distinct from Group 4 (range: 1.69–1.73) (Table S1). In other Mycoplasma species, SD interactions appeared to largely involve translation initiation, as considerable dF_{SD} values were observed.

These results substantiate our hypothesis that the loss of classical anti-SD motifs equates to loss of SD interactions. Conservation of the classical anti-SD motif, however, does not always correspond to a high frequency in SD-like sequences, dramatic examples of which were seen in Flavobacteria (with the anti-SD motif) as well as in *M. genitalium* and *M. pneumonia*. It is presumable that SD-led and non SD-led mechanisms (for example, direct translation of leaderless mRNA) for translation initiation coexist in most bacteria. SD-led mechanisms appear in a very small number of Flavobacteria (with the anti-SD motif), *M. genitalium*, and *M. pneumonia* genes or not at all in the lost anti-SD bacteria

3.3. Reductive Evolution May be Associated with the SD Interaction Loss

The loss of the classical anti-SD motif in primary endosymbionts (Groups 2 and 3) and hemotrophic mycoplasmas (Group 4) possibly is related to a host-associated lifestyle. Extreme host association restricts the effective population size and causes frequent population bottlenecks at the time of transmission. Because host-associated bacteria can stably obtain copious metabolites from hosts, most of their genes for metabolism and other cellular processes have been eliminated, resulting in massive genomic downsizing (Toft and Andersson 2010; McCutcheon and Moran 2012).

Primary endosymbionts and mycoplasmas are thought to be at the extreme of this host association, as they lack genes most free-living bacteria have including those for DNA repair, recombination, and transfer, and they carry the smallest genomes among sequenced bacteria to date as a consequence (Toft and Andersson 2010; McCutcheon and Moran 2012). SD interaction, which may be a useful regulatory mechanism but not essential for life, might have lost in response to such genomic minimization. I assume that ancestors of obligate lost anti-SD bacteria (Groups 2, 3, and 4) in a free-living state had multiple mechanisms for translation initiation. During a period of host association, the evolutionary forces toward large-scale gene/function loss may have leaded to elimination of SD interactions, remaining other option for translation initiation.

I looked for the gene loss pattern in several lost anti-SD bacteria belonging to primary endosymbionts. I first categorized orthologous clusters of gamma- and beta-

proteobacteria according to functional categories in the KEGG orthology database (Kanehisa et al. 2014). SD interaction usage (F_{SD}) was calculated for each category to investigate differences in SD interaction usage among functional categories. For the both bacterial groups, genes for key biological functions such as transcription, transport, catabolism, energy metabolism, signal transduction, cell growth, and cell death showed high SD interaction usage (Figs. 6 and 7). The other categories also contained some SD sequence-carrying genes with lower SD interaction usage. Despite the diversity in usage, SD sequences were wide-spread among all functional categories. This probably suggests that translation initiation by SD interactions is not a mechanism specific to some particular functions.

Gene loss processes during genome reduction of primary endosymbionts seem to have undergone massively in all functional categories except for categories of essential processes such as transcription and translation (Figs. 6 and 7). Gene loss in primary endosymbionts did not occur specifically in genes with a SD sequence, suggesting SD interaction loss was not due to direct loss of genes with a SD sequence. It is presumable that reduction of absolute number of SD sequence-carrying genes during extreme genome reduction likely increased the chance for radical evolution of SD sequences towards their elimination.

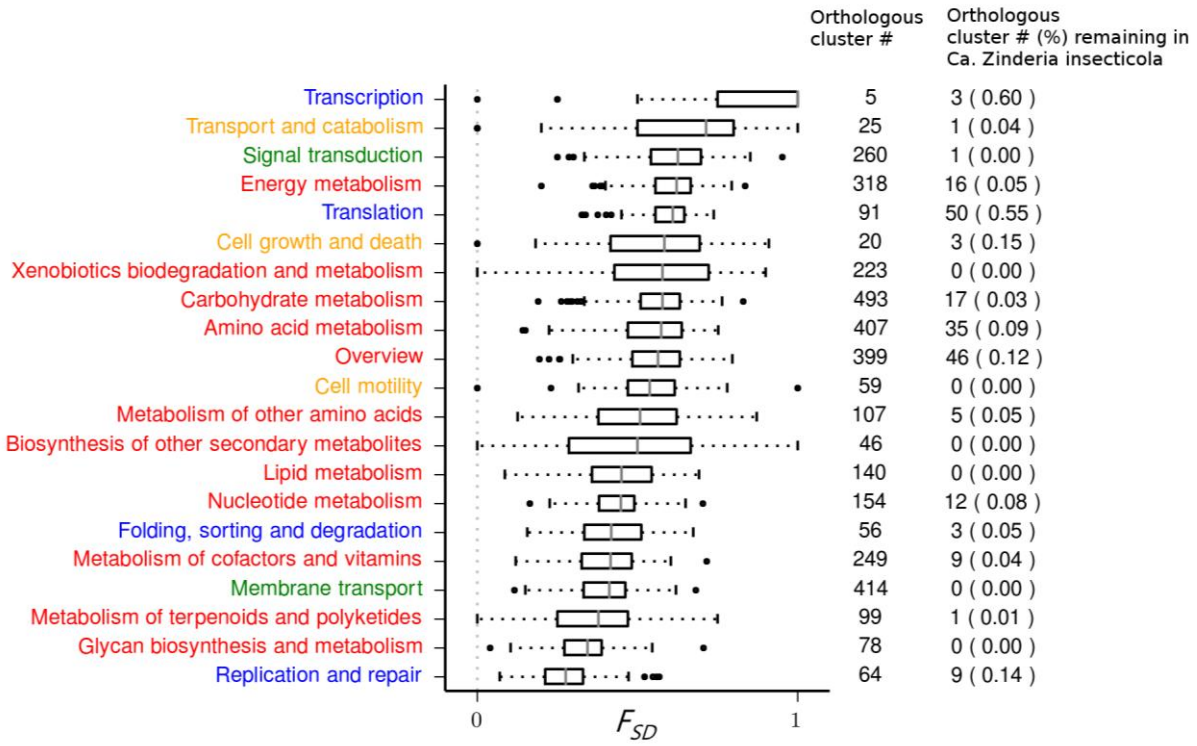


FIG. 6. F_{SD} values (SD interaction usage) of various functional gene categories for Gammaproteobacteria. F_{SD} values resulted from 184 species with genome size > 1 Mb were shown in a boxplot for each gene category. Functional gene categories with orthologous clusters > 4 are shown. Number and percentage of orthologous clusters remaining in *Ca.* Zinderia insecticola (a lost-anti-SD bacterium) are shown. Blue fonts: Genetic information. Green fonts: Environmental information. Red fonts: Metabolism. Orange: Cellular processes.

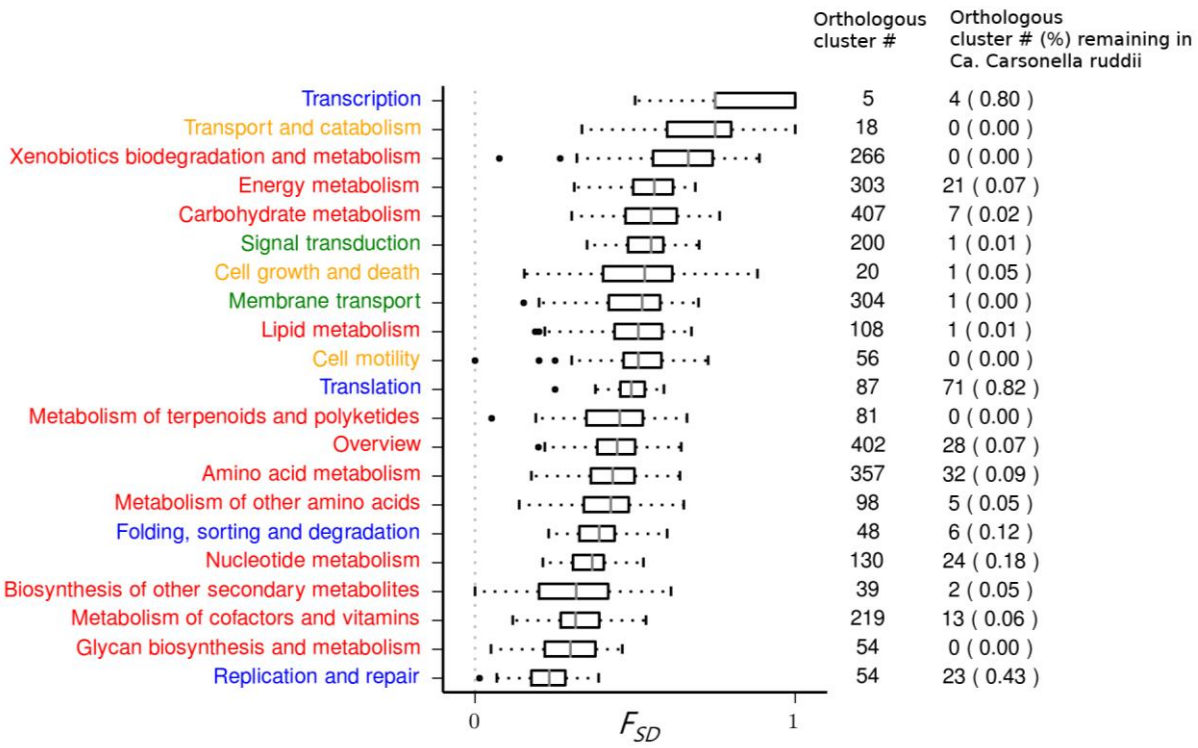


FIG. 7. F_{SD} values (SD interaction usage) of various functional gene categories for Beta-proteobacteria. F_{SD} values resulted from 81 species with genome size > 1 Mb were shown in a boxplot for each gene category. Functional gene categories with orthologous clusters > 4 are shown. Number and percentage of orthologous clusters remaining in *Ca.* Carsonella ruddii (a lost-anti-SD bacterium) are shown. Blue fonts: Genetic information. Green fonts: Environmental information. Red fonts: Metabolism. Orange: Cellular processes.

SD interaction loss, however, is not strongly correlated with reduced genomic size. Groups 2 and 3 lost anti-SD bacteria that are members of the smallest genomes among those used in this study (Table 1) supported the hypothesis. However, this was not supported by *B. aphidicola*, *B. cicadellinicola* and *Ca.* Blochmannia vafer (Fig. 4), because they are primary endosymbionts of insects with genomic sizes as small as those in Groups 2 and 3, despite the considerable SD interaction usage. Moreover, the genomic size of *M. haemofelis* str. Langford 1 (Group 4) is the third largest among 26 mycoplasmas I analyzed.

It is obvious that genome reduction is a feature shared by many lost anti-SD bacteria. The loss, however, cannot be explained simply by genome reduction. Each bacterium is on its own evolutionary history and path, which makes it difficult to point out causes for SD interaction loss.

3.4. Emergence of a Novel Translation Initiation Mechanism May Have Led to SD Interaction Loss

Flavobacteria showed a distinct pattern in the 5′ UTR, which was an A-rich pattern before a start codon (Fig. 3D). No primary endosymbionts in this class that has been completely sequenced to date contained classical anti-SD motifs. Interestingly, three free-living in this class without any signs of extreme genomic reduction (Group 1) also belonged to lost anti-SD bacteria. An unknown alternative mechanism that uses the A-rich signal is superior to the SD-led mechanism in this class, hence possibly promoting

loss of SD interactions. The alternative mechanism, however, is still hypothetical, thus its experimental confirmation remains to be achieved.

3.5. Materials and Methods

3.5.1. Determining Small Subunit rRNA Genes and Anti-SD Sequences

Refseq sequences of complete bacterial genomes (n = 1081) and their annotation information were downloaded from the National Center for Biotechnology Information website (ftp://ftp.ncbi.nih.gov/genomes/Bacteria/) on 26 April 2011. I first retrieved small subunit rRNA gene sequences from these genomes based on their registered annotations. The 3′ tail of the reference small subunit rRNA gene, which was a sequence of the small subunit rRNA gene of E. coli str. K-12 substr. DH10B, was used to locate corresponding regions in other genomes. The ClustalW algorithm (Larkin et al. 2007) was used for the sequence alignment with default parameters. When the retrieved sequence did not include the 3′ tail sequence, I searched for it downstream. I searched the 3′ tail for the classical anti-SD motif, 5′CCUCC.

3.5.2. Phylogenetic Tree

Small subunit rRNA gene sequences were aligned by ClustalW with default parameters. For the alignment, trees were constructed using MEGA5 (Tamura et al. 2011) (maximum likelihood method, Tamura–Nei model (Tamura and Nei 1993), and bootstrap replication of 1,000 times).

3.5.3. Analysis of SD Interaction

For every protein-coding gene, I retrieved the sequence in the SD region (the region -20 to -5 nt from the start codon). To determine a given SD region had a SD sequence, I followed previous approaches that quantified the minimum change in free energy (ΔG) in duplex formation between the SD region and the 3′ tail of small subunit rRNA (Schurr et al. 1993; Ma et al. 2002; Starmer et al. 2006; Nakagawa et al. 2010). SD sequences were defined when ΔG was lower than a certain cutoff; I used two cutoff values, -3.4535 and -4.4 kcal/mol following earlier studies (Ma et al. 2002; Starmer et al. 2006). FREE_SCAN algorithm (Starmer et al. 2006) was used for measuring ΔG between two strands. Next I calculated the gene fraction carrying the SD sequences among all for a given genome by the equation: F_{SD} = "Protein-coding genes number with the SD sequences" / "Total number of protein-coding genes". For each genome, F_{SD} was adjusted to dF_{SD} as an earlier study (Nakagawa et al. 2010) described by the equation: $dF_{SD} = dF_{SD} - rF_{SD}$, where rF_{SD} was the fraction of SD sequences among artificially generated sequences (n = 20000, 16 nt). The probability for generating a particular nucleotide equated the background fraction of each nucleotide, where the background nucleotide fraction was defined as the fraction of each nucleotide in sequences ranging from -21 to -100 nt relative to start codons in the given genome.

3.5.4. Analysis of Nucleotide Bias in 5′ UTRs

I calculated nucleotide bias in 5′ UTRs (from -50 to -1 relative to the start codon) of all protein-coding genes in a given genome. The nucleotide bias df_N (N = A, C, G, or T)

was calculate by the equation: df_N = “fraction of a nucleotide N at the specific position” / “the background fraction of nucleotide N”.

4. Part 2. Alterations in Shine-Dalgarno Interaction during Plastid Evolution

*This part is written based on a submitted manuscript (Lim K, Kobayashi I, Nakai K. Alterations in rRNA-mRNA interaction during plastid evolution)

4.1. Little is Known about SD Interactions in Plastids.

The Shine-Dalgarno (SD) interaction is rRNA-mRNA interaction used in the prokaryotic system for translation initiation (Shine and Dalgarno 1974). This mechanism is never observed in the nuclear genetic systems of eukaryotes (Malys and McCarthy 2011). There is a distinct base pairing rule common in SD interactions; a pyrimidine-rich, anti-SD sequence in the 3′ tail of a small subunit rRNA binds to a complementary, purine-rich, SD signal sequence in the 5′ untranslated region (UTR) of an mRNA (Fig. 1). A core motif (i.e., the anti-SD motif), 3′CCUCC, is conserved among anti-SD sequences (Ma et al. 2002; Nakagawa et al. 2010), suggesting an extreme evolutionary constraint and a crucial role for SD interaction.

Although the SD interaction is observed in the overwhelming majority of prokaryotes, its usage varies considerably (Ma et al. 2002; Chang et al. 2006; Nakagawa et al. 2010). In the previous chapter, I reported the rare loss of the anti-SD motif and its complement SD sequence in several bacterial groups. Most of these bacteria are under obligate association with their eukaryotic hosts and have undergone massive genome reduction (primary endosymbionts of insects or hemotrophic mycoplasmas). These obligate host-associated bacteria share many biological and evolutionary features with eukaryotic

organelles of prokaryotic origin, such as mitochondria and plastids (Toft and Andersson 2010; McCutcheon and Moran 2012). This made me hypothesize that SD interactions in plastids may have undergone drastic molecular evolution as obligate bacteria. To gain insight into this evolutionary process, I analyzed SD interactions in plastids. I discovered that the classical anti-SD motif in rRNA and the complementary SD signal in mRNA have coevolved beyond the classical SD interaction.

4.2. Variations in Canonical Anti-SD Motifs.

I investigated variations in the canonical anti-SD motif sequence (3'CCUCC) from plastid genome sequences (n = 429, including chloroplasts and nonphotosynthetic plastids). Analyzing the 3′ tail that follows helix 45 in small subunit rRNA (Figs. 1 and 8) allowed me to find variations in the canonical anti-SD motif in all copies of small subunit rRNA genes in 17 plastid genomes (Table 2 and Fig. 8). I referred to these as *mutated anti-SD plastids*. The other plastids (n = 412) with canonical anti-SD motifs were referred to as *conserved anti-SD plastids*.

4.3. Conserved anti-SD plastids are Highly Diverse in SD Interaction Usage.

I next quantified the usage of SD interactions in conserved anti-SD plastids. Our strategy for determining the SD sequence was based on previous studies (Schurr et al. 1993; Ma et al. 2002; Starmer et al. 2006; Nakagawa et al. 2010) that predicted the minimum free energy (MFE) structure between an anti-SD sequence (on the 3′ tail of a small subunit

rRNA) and a 5′ UTR of an mRNA to obtain the most presumable rRNA-mRNA base pairing structure.

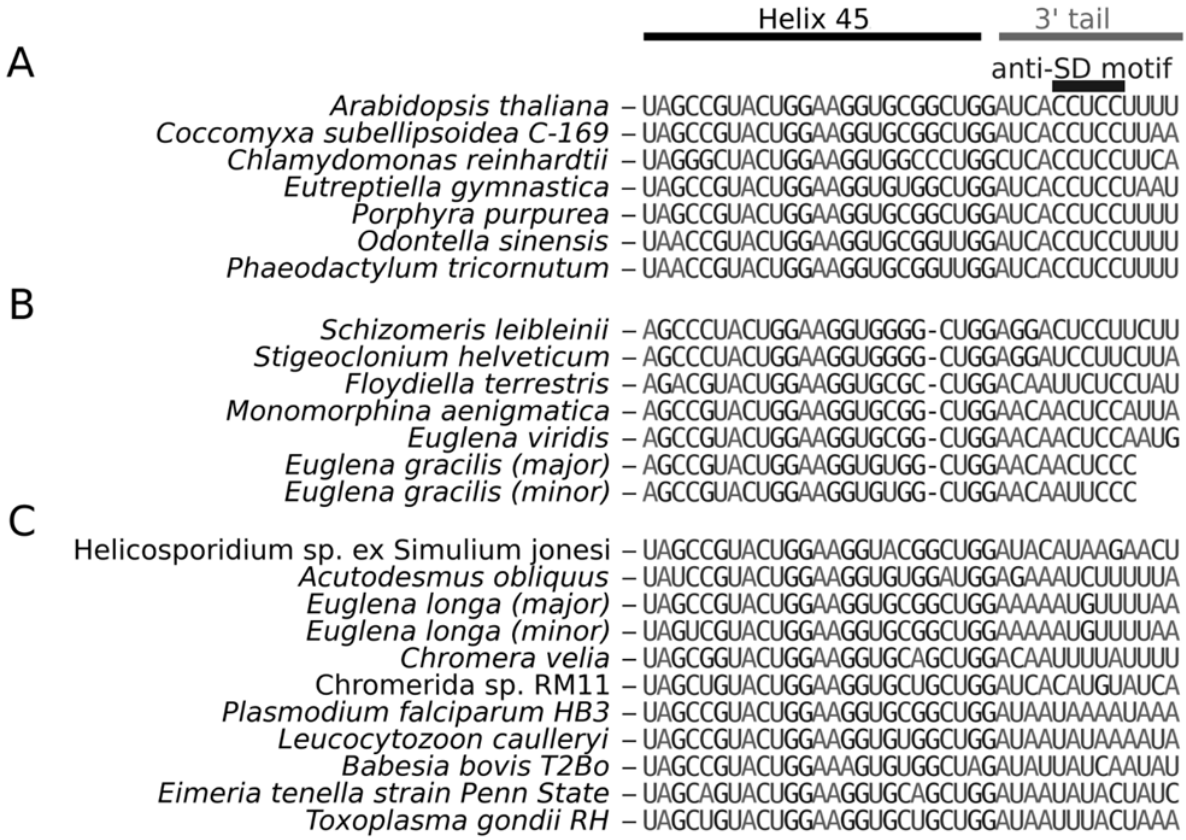


FIG. 8. Multiple alignment of 3′ end regions of small subunit rRNAs, for (A) selected conserved anti-SD plastids; (B) reduced anti-SD plastids; and (C) lost anti-SD plastids. In this study, reduced and lost anti-SD plastids are collectively referred to as mutated anti-SD plastids. Major: major small subunit rRNA sequence type. Minor: minor small subunit rRNA sequence type. The 3′ tail length was defined as 13 nt. The 3′ tail sequence of *E. gracilis* was shown as previously reported (Steege et al. 1982). This figure is adapted from Lim et al.(submitted)

Table 2. Plastids with an rRNA lacking the canonical anti-SD motif

Taxonomic group (supergroup) species	Canonical anti-SD motif[a]	Small subunit rRNA of plastids	
		3′ tail[b]	Gene Copy
Chlorophyta (Archaeplastida)			
Helicosporidium sp. ex Simulium jonesi	Lost	3′UCAAGAAUACAUA	1
Acutodesmus obliquus	Lost	3′AUUUUUCUAAAGA	2
Schizomeris leibleinii	Reduced	3′UUCUUCCUCAGGA	1
Stigeoclonium helveticum	Reduced	3′AUUCUUCCUAGGA	1
Floydiella terrestris	Reduced	3′UAUCCUCUUAACA	1
Euglenophyta (Excavata)			
Monomorphina aenigmatica	Reduced	3′AUUACCUCAACAA	1
Euglena viridis	Reduced	3′GUAACCUCAACAA	1
Euglena gracilis	Reduced	3′CCCUCAACAA	3
	Reduced	3′CCCUUAACAA	1
Euglena longa	Lost	3′AAUUUUGUAAAAA	4
Chromerida (Chromalveolata)			
Chromera velia	Lost	3′UUUUAUUUUAACA	1
Chromerida sp. RM11	Lost	3′ACUAUGUACACUA	2
Apicomplexa (Chromalveolata)			
Plasmodium falciparum	Lost	3′AAAUAAAAUAAUA	1
Leucocytozoon	Lost	3′AUAAAAUAUAAUA	2

caulleryi			
Babesia bovis	Lost	3′UAUAACUAUUAUA	2
Theileria parva	Lost	3′AAUGUGUAUUUUA	1
Eimeria tenella	Lost	3′CUAUCAUAUAAUA	2
Toxoplasma gondii	Lost	3′AAAUCAUUUAAUA	2

[a]Reduced: changes ≤ 2 nt (insertion, deletion, or substitution) compared to the canonical anti-SD motif (3′CCUCC). Lost: more changes.

[b]The 3′ tail length was collectively set as 13 nt except for *E. gracilis*, where the 3′ tail sequence is shown as previously validated (Steege et al. 1982). The putative remnants of the canonical anti-SD motif are underlined.

See Table S2 for more detailed information on the above plastids.

This Table is adapted from Lim et al.(submitted)

More precisely, these studies established that a given sub-region of 5′ UTR had a potential SD sequence when the MFE < particular cutoff value, with a lower value denoting greater RNA-RNA stability. The determination of the cutoff was described in Materials and Methods. I next calculated how much fraction of SD sequence-carrying protein-coding genes is present in each genome, which was denoted as F_{SD}. I used F_{SD} as an index for SD interaction usage in this study.

F_{SD} allowed comparison of SD interaction usage between genomes and taxonomic groups as shown in Fig. 9A(i). SD interaction usage varied among individual plastids and taxonomic groups. As asserted for prokaryotes (Schurr et al. 1993; Chang et al. 2006; Nakagawa et al. 2010), conservation of the classical anti-SD motif in rRNA was not necessarily indicative of high SD interaction usage; some conserved anti-SD plastids showed very low F_{SD} values (Fig. 9A(i)).

A high median F_{SD} with little diversity was seen in Streptophyta plastids (Fig. 9A(i)), whereas a high diversity in the usage was shown in Chlorophyta plastids (Fig. 9A(i)). It was impossible to deduce more taxon-specific patterns of SD interaction usage due to the extreme taxonomic bias in current plastid genome sampling.

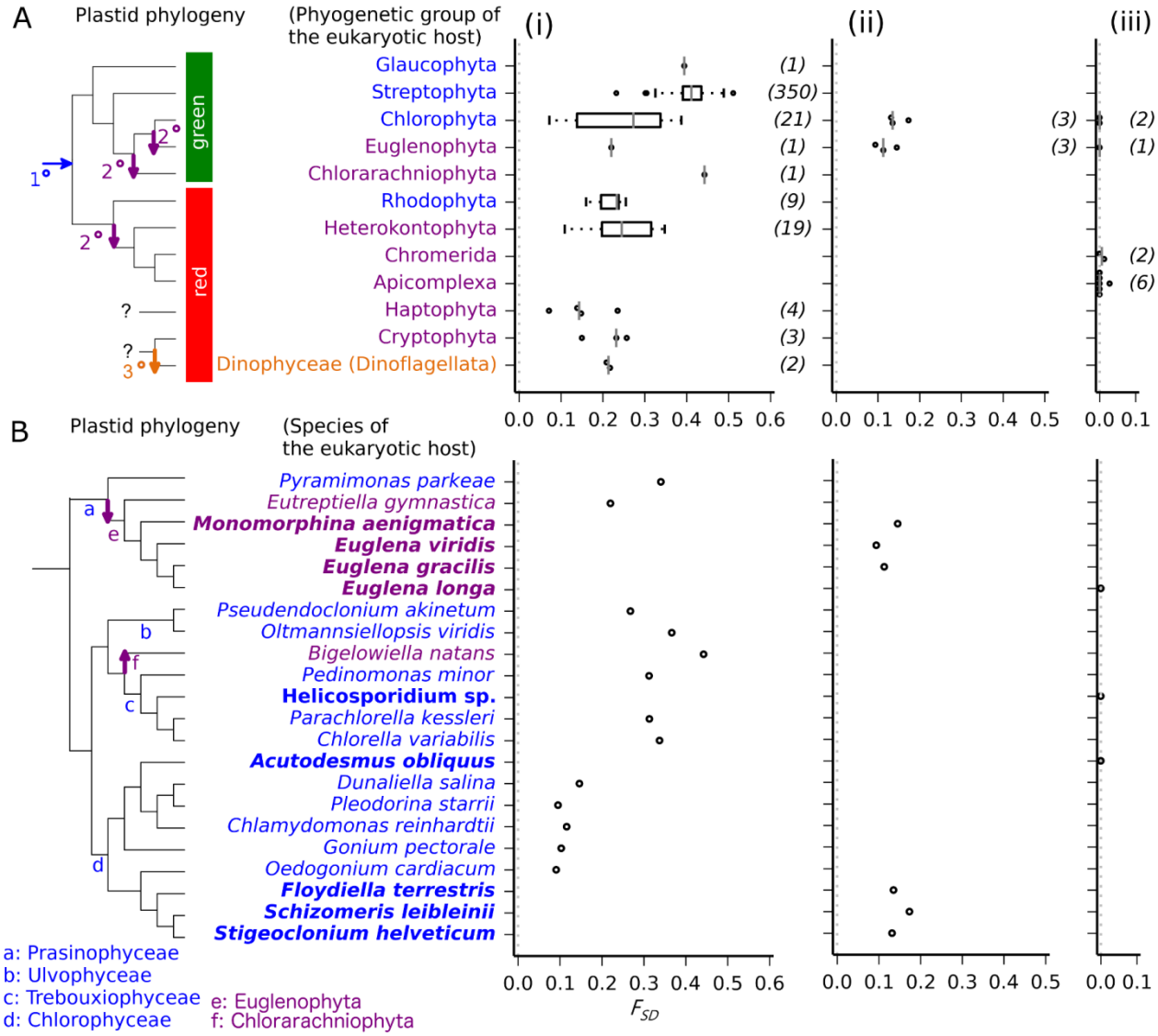


FIG. 9. Analysis of SD interaction usage in plastids. (A) F_{SD} values (the fraction of SD sequence-carrying genes in a genome) were shown on plastid phylogeny. A vertical bar: median F_{SD}. A group with > 6 members: a boxplot was drawn. (B) F_{SD} values and multi-gene phylogeny are shown for plastids in selected species of Chlorophyta, Euglenophyta, or Chlorarachniophyta. (i) Plastids carrying canonical anti-SD motifs refered to as conserved anti-SD plastids; (ii) plastids with reduced anti-SD motifs refered to as reduced anti-SD plastids; (iii) plastids lacking classical anti-SD motifs referes to as lost anti-SD plastids. The species containing (ii) and (iii) are emphasized in bold font. 1°: primary endosymbiosis. 2°: secondary endosymbiosis. 3°: tertiary endosymbiosis. The tree in panel B was drawn by plastid protein-coding gene sequences. A detailed tree with scaled branch lengths and bootstrap values is shown in Fig. S1. See Materials and Methods for details on tree construction. This figure is adapted from Lim et al.(submitted)

4.4. Mutations in the Canonical Anti-SD Motif in Multiple Plastid Lineages.

Mutated anti-SD plastids arose in four taxonomic groups, Chlorophyta, Euglenophyta, Apicomplexa, and Chromerida (Table 2 and Fig. 9A). Although there were multiple copies of small subunit rRNA genes in some of these plastids (Table 2), intra-genomic copies of small subunit rRNA genes had identical 3′ tail sequences, with the exception of the plastid of *Euglena gracilis*. In *E. gracilis*, there is a single nucleotide (nt) difference between the major (three / four copies) and minor (one / four copies) genes (Table 2). It is known that intra-genomic homogeneity of rRNA genes is maintained by gene conversion between them (Palmer 1985; Hashimoto et al. 2003; Khakhlova and Bock 2006).

The mutated 3′ tails of rRNAs may still maintain their functions for SD interactions. To examine this possibility, the capability of putative anti-SD sequences in mutated anti-SD plastids to form base pairing with 5′ UTRs of genes in the plastids was tested, assuming that the realative position of the putative anti-SD sequences is identical with that of anti-SD sequences in conserved anti-SD plastids (see Material and Methods for details). The rRNA-mRNA interactions that met the MFE criterion applied above to conserved anti-SD plastids were considered as SD interactions in a broad sense (see Material and Methods for details). This approach allowed measuring F_{SD} for mutated anti-SD plastids. For plastids with multiple types of anti-SD sequences within a genome, the major type (the type observed in the largest number of small subunit rRNA gene copies) was used for the F_{SD} calculation.

Mutated anti-SD plastids could be classified into two groups based on the level of anti-SD motif variation. The first group (n = 6) contained plastids with ≤ 2 nt changes (insertions, deletions, and substitutions) in the canonical anti-SD motif refered to as *reduced anti-SD plastids*. Reduced anti-SD plastids presumably maintain SD interactions, because they displayed F_{SD} values > 0.09 (Table 2 and Fig. 9A(ii)). The other group (n = 11) contained plastids with a larger variation in the canonical anti-SD motif. These *lost anti-SD plastids* (Table 2 and Fig. 9A(iii)) had negligible F_{SD} values (< 0.05) indicating their loss of SD interactions.

Mutated anti-SD plastids (reduced and lost anti-SD plastids combined) could also be grouped into two large phylogenetic clusters (Fig. 9A). The first cluster (labeled green in Fig. 9A) included primary plastids in Chlorophyta (green algae) (n = 5), and secondary plastids in Euglenophyta (n = 4) that had originated in Chlorophyta through secondary endosymbiosis (Fig. 2). The second cluster (labeled red in Fig. 9A) included secondary plastids in Apicomplexa (n = 6) and Chromerida (n = 2) (Fig. 9A), which originated in Rhodophyta (red algae) through secondary endosymbiosis (Fig. 2).

The green cluster (green algae) contained both reduced and lost anti-SD plastids that apparently emerged independently multiple times in phylogeny (Fig. 9B). Phylogeny suggests that one of the mutation events in the canonical anti-SD motif may have occurred in a Euglenida (Euglenophyta) plastid after secondary endosymbiosis of an ancestor of Chlorophyta (Fig. 9B(e)). Among eukaryotes with completely sequenced plastids, Euglenida has three genera, Eutreptiella, Monomorphina, and Euglena. A conserved anti-SD plastid belonged to Eutreptiella, whereas reduced (*Monomorphina*

aenigmatica, *Euglena viridis*, and *Euglena gracilis*) and lost (*Euglena longa*) anti-SD plastids belonged to the other genera. The order of evolution deduced from phylogeny is the conserved anti-SD plastid of Eutreptiella, the reduced anti-SD plastids of *M. aenigmatica* and *E. viridis* and the reduced anti-SD plastid of *E. gracilis*, and the lost anti-SD plastid of *E. longa*. Additionally, lost anti-SD plastids were found independently in Trebouxiophyceae (Fig. 9B(c), *Helicosporidium* sp. ex Simulium jonesi) and Chlorophyceae (Fig. 9B(c), *Acutodesmus obliquus*). Three reduced anti-SD plastids clustered phylogenetically in a clade within Chlorophyceae (Chlorophyta) (Fig. 9B(c)).

In the red cluster (Apicomplexa and Chromerida), only lost anti-SD plastids were observed (Fig. 9A). These two groups are neighboring in phylogeny, suggesting a shared history of the loss of SD interaction.

4.5. SD Interaction Loss and Genome Reduction.

Next, I examined a possible association between SD interaction loss and the reductive evolution of plastid genomes. I then found tendency that lost anti-SD plastids tend to have smaller genome than their close relatives (Fig. 10).

Evidence for SD-like interaction was reported in bacteria-like mitochondria with much larger and gene-rich genomes compared to other mitochondria (Lang et al. 1997; Burger et al. 2013). SD interactions in other known mitochondria seem to have been eliminated (Hazle and Bonen 2007) These reports support our assumption of association between genome reduction and the loss of SD interaction.

Such trend was clearly seen by comparisons between plastids in the red lineage. A Rhodophyta (red algae) plastid transferred to an ancestor of Chromalveolata through secondary endosymbiosis forming Chromalveolata subgroups in the order of Heterokontophyta, Chromerida and Apichomplexa (Keeling 2013). The gene number decreased in the indentical order (Fig. 10A), the last two groups of which showed the loss of SD interactions (Fig. 9A and 10A).

Among the Euglenophyta plastids originated from Chlorophyta by secondary endosymbiosis (Fig. 2), the *E. longa* plastid, a lost anti-SD plastid, has undergone the most drastic genome reduction (protein-coding gene number < 40) (Fig. 10A).

In two lost anti-SD plastids belong to Chlorophyta, only one plastid, the *Helicosporidium* sp. ex Simulium jonesi plastid, showed extreme genome reduction (Fig. 10A). Except for this, there are several evidences that contradict association between genome reduction and loss of SD interaction in primary plastids of the green lineage (Chlorophyta and Streptophyta) (Fig. 10A). For example, the other lost anti-SD plastid in Chlorophyta, the *Acutodesmus obliquus* plastid, has average genome size compared to its relatives (Fig. 10A). In addition, nonphotosynthetic plastids in parasitic plants belonging to Streptophyta such as *Rhizanthella gardneri*, *Epifagus virginiana*, and *Neottia nidus-avis* have very tiny genome size (protein-coding gene number < 40) (Fig. 10A) but show high SD interaction usage ($F_{SD} > 0.3$).

Drastic plastid genome reduction often resulted in the loss of photosynthetic function and associated genes, hence nonphotosynthetic plastids often coincide with the loss of SD interactions as asserted in plastids of Apicomplexa, *E. longa*, and *Helicosporidium* sp. ex Simulium jonesi (Fig. 10).

For genomes containing a small number of genes, the loss of just a few genes with an SD signal will eliminate SD interactions, resulting in the loss of the anti-SD motif by the rRNA. The reverse is also possible; the loss of SD interaction through a mutation in the anti-SD motif of rRNA hampered the expression of many genes simultaneously, promoting their decay and thereby driving extreme genome reduction. An examination of more plastid and nuclear genomes could provide evidence to these scenarios.

It is also presumable that the two phenomena, genome reduction and SD interaction loss, are not in a cause and effect relationship. Another factor such as relaxed selection due to accelerated reductive evolution may have facilitated the elimination of both SD interaction and gene content, contributing to a correlation of the two phenomena.

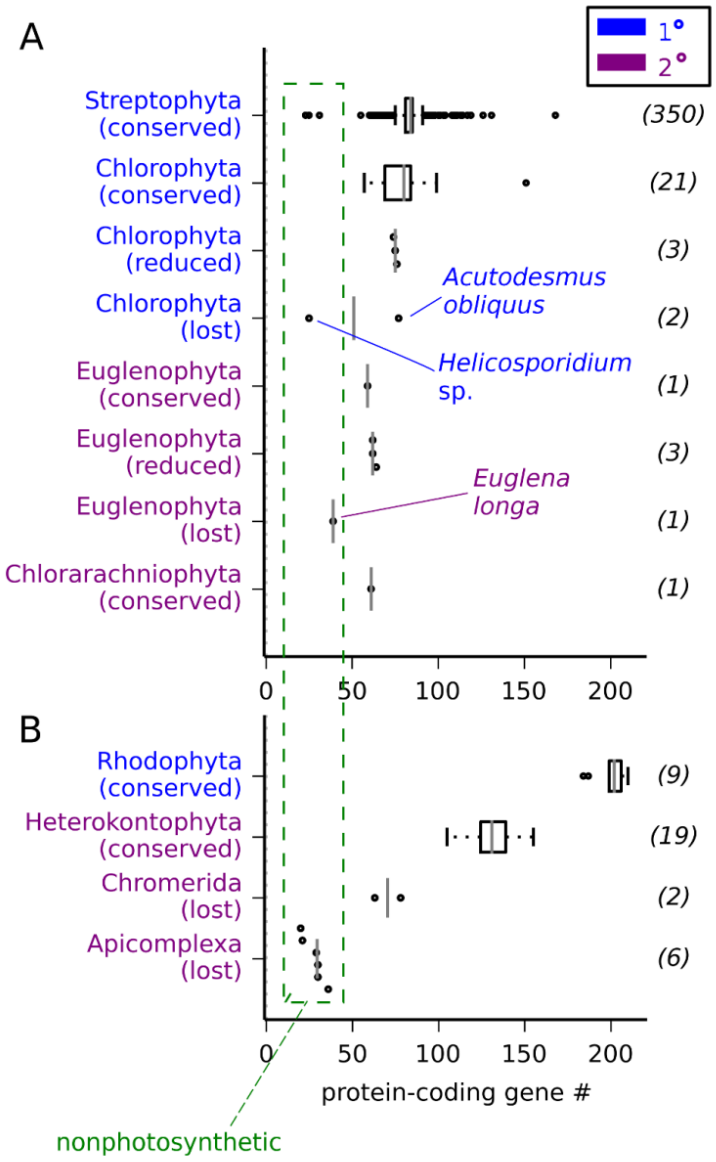
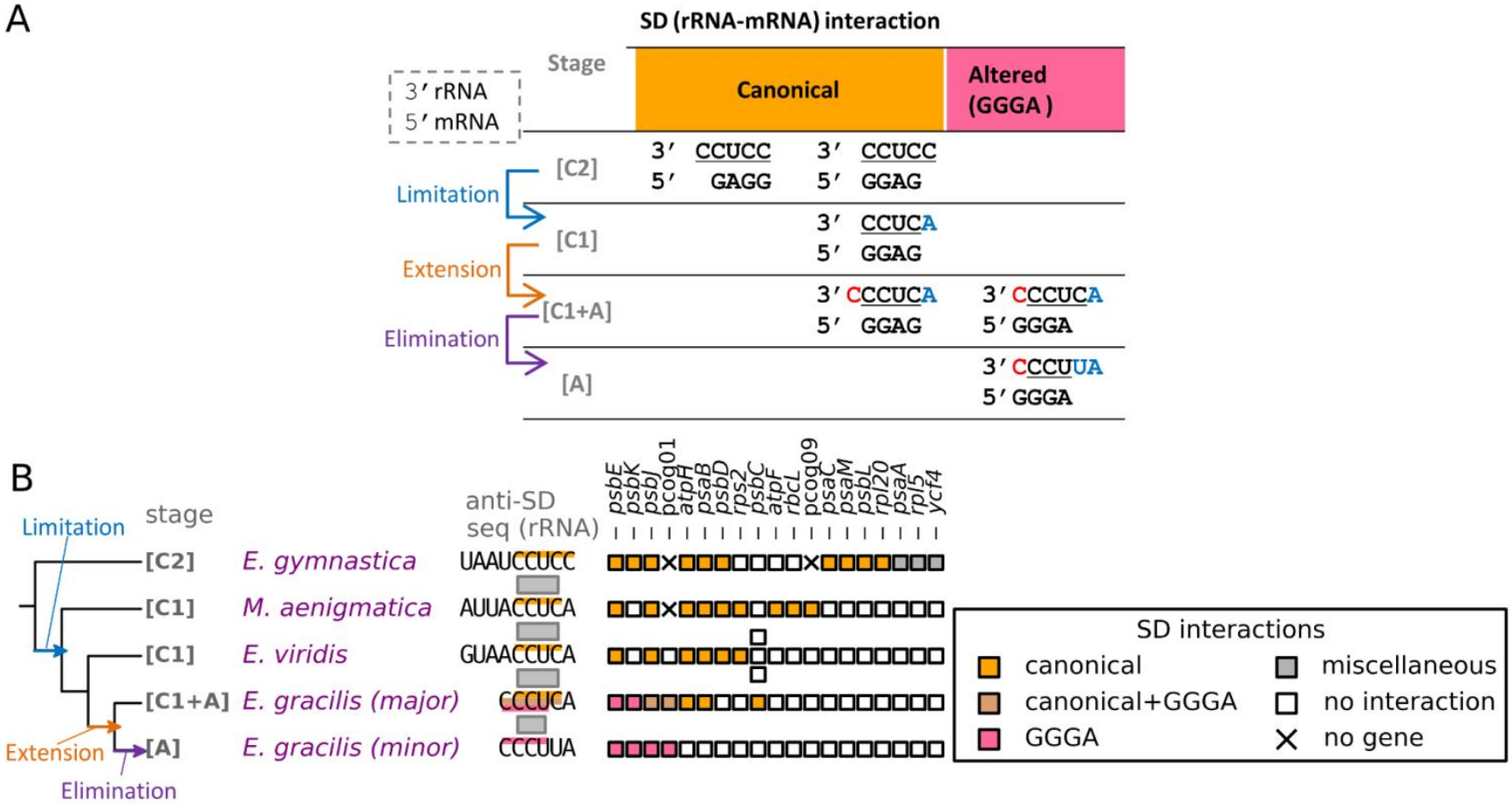


FIG. 10. Lost anti-SD plastids tend to carry a small genome. (A) Comparison among green-lineage plastids (Streptophyta, Chlorophyta, Euglenophyta, and Chlorarachniophyta plastids). (B) Comparison among red-lineage plastids (Rhodophyta, Heterokontophyta, Chromerida and Apicomplexa plastids). 1°: plastids from primary endosymbiosis. 2°: plastids from secondary endosymbiosis. Conserved: conserved anti-SD plastids. Reduced: reduced anti-SD plastids. Lost: lost anti-SD plastids. A vertical bar: a median of the protein-coding gene numbers. A boxplot is drawn for a group with > 6 members. This figure is adapted from Lim et al.(submitted)

Another likely relevant factor is host-driven innovation within the plastid translational system. This is indicated by the existence of many plastid-specific translation-related elements such as Nac2, RBP40, plastid-specific ribosomal proteins, and ribosomal proteins with plastid-specific domains (Yamaguchi et al. 2002; Hirose and Sugiura 2004b; Manuell et al. 2007; Schwarz et al. 2007). Collectively, these elements may have been functionally substituted for the SD interaction. Their intra-genomic usage and precise roles in translational regulation across plastid lineages need to be determined to gain a better understanding of their influence on the evolution of SD interaction.

4.6. Coevolution between Anti-SD Motifs and Complementary SD Signals.

The above F_{SD} calculation revealed that the altered anti-SD regions of reduced anti-SD plastids were able to form base paurings with their own 5′ UTRs. I examined the possibility that reduced anti-SD plastids have maintained SD interactions through coordinated changes in the anti-SD motif (in rRNA) and the complementary SD signal (in mRNA). There can be two possible 4-base rRNA/mRNA pairs for the SD interaction (canonical SD pairs) mediated by the canonical anti-SD motif (3′CCUCC): 3′CUCC/5′GAGG and 3′CCUC/5′GGAG (referred to as *canonical SD interactions*) (Fig. 11A). In reduced anti-SD plastids, I detected two types of SD interactions different from these, 3′CCCU/5′GGGA and 3′CUUCC/5′GAAGG, refered to these interactions as *altered SD interactions.*


A
SD (rRNA-mRNA) interaction
Stage
Canonical
Altered (GGGA)
3′ rRNA
5′ mRNA
[C2]
Limitation
3′ CCUCC
5′ GAGG
3′ CCUCC
5′ GGAG
[C1]
Extension
3′ CCUCA
5′ GGAG
[C1+A]
Elimination
3′ CCCUCA
5′ GGAG
3′ CCCUCA
5′ GGGA
[A]
3′ CCCUUA
5′ GGGA
B
stage
Limitation
anti-SD seq (rRNA)
psbE psbK psbJ pcog01 atpH psaB psbD rps2 psbC atpF rbcL pcog09 psaC psaM psbL rpl20 psaA rpl5 ycf4
[C2] E. gymnastica UAAUCCUCC
[C1] M. aenigmatica AUUACCUCA
[C1] E. viridis GUAACCUCA
[C1+A] E. gracilis (major) CCCUCA
[A] E. gracilis (minor) CCCUUA
Extension
Elimination
SD interactions
canonical
canonical+GGGA
GGGA
miscellaneous
no interaction
no gene

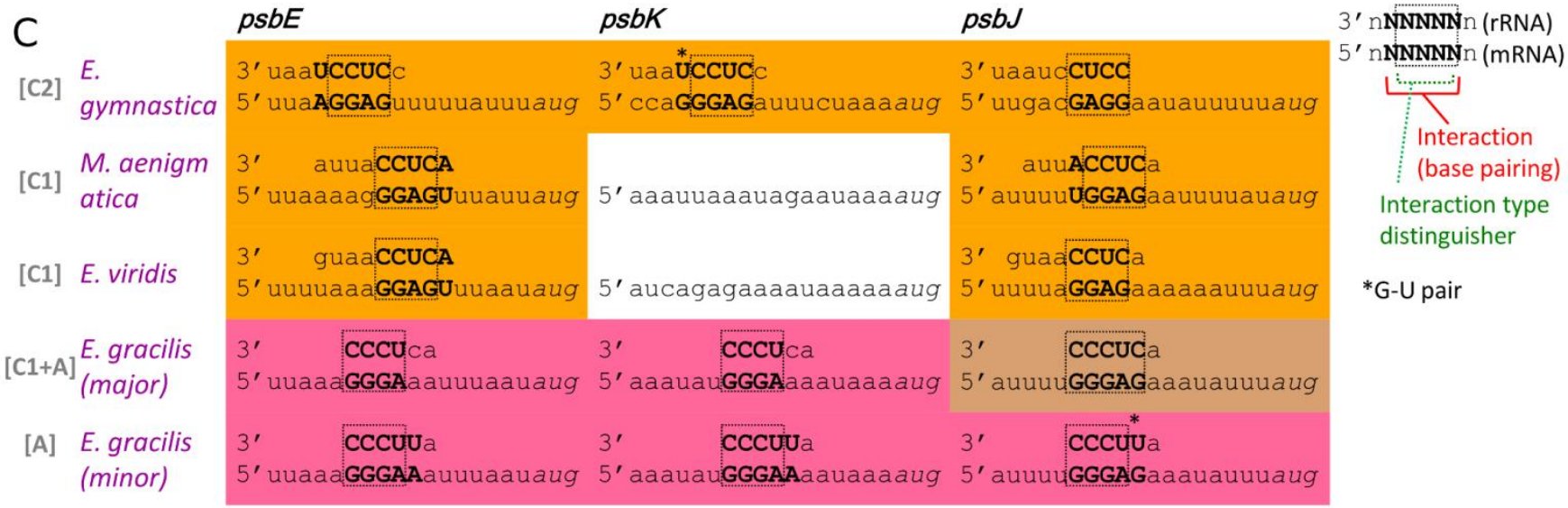


FIG. 11. rRNA-mRNA coevolution in Euglenophyta plastids. (A) Inferred stages of the coevolution. Two canonical SD interactions were limited to one interaction, then extended for the emergence of an altered SD interaction (GGGA interaction), and the remaining canonical SD interaction was then eliminated. (B) Anti-SD motifs (in anti-SD sequences) and a matrix for plastid genes at various evolutionary stages. Boxes in the matrix indicate genes and are colored according to their SD interactions. Multiple boxes denote different paralogous genes. Major: the type present in three of the four small subunit rRNA genes. Minor: the type present in one of the four rRNA genes. Miscellaneous: secondary structures with the bulge or internal loop structure, or non-Watson-Crick base paring. See Table S3 for the orthologous group, pcog01 and pcog09. (C) Alignments of predicted SD interactions for several orthologous genes. This figure is adapted from Lim et al.(submitted).

The altered SD interactions evolved likely via the following coordinated changes: (i) *extension*, where an altered SD interaction mediated by a sequence that combined part of the canonical anti-SD motif with its flanking sequence was emerged; (ii) *limitation*, a 1-nt mutation in the canonical anti-SD motif eliminated one of the two canonical SD pairs; and (iii) *elimination*, where another mutation destroyed the other canonical SD pair and only the altered interaction retained. There were two presumable orders of this evolution: limitation, extension, and elimination in the plastid lineage of Euglenophyta (Fig. 11), and extension, limitation, and elimination in the plastid lineage of Chlorophyta (Fig. 12), as detailed below.

(a) Euglenophyta

The stepwise changes of limitation, extension, and elimination were clearly seen in secondary plastids in Euglenophyta (Fig. 11), as below.

Limitation: A "C → A" substitution at the 5′ end of the canonical anti-SD motif (3′CCUCC to 3′CCUCA) disallowed one of the two canonical SD pairs (stage [C1] in Fig. 11A). This stage was observed in the *M. aenigmatica* and *E. viridis* plastids (Fig 11B).

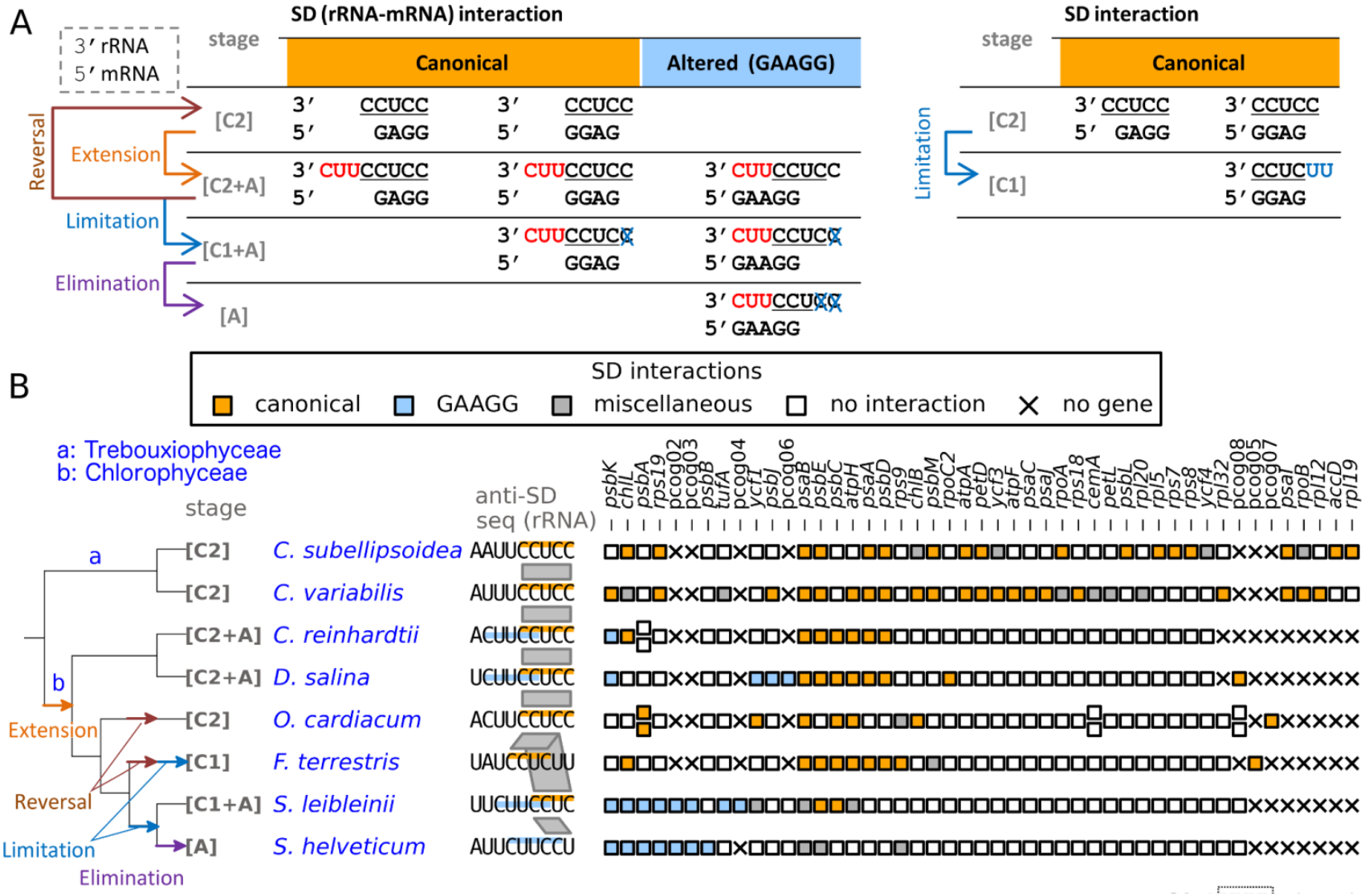
Extension: One C was generated at next to the 3′ end of the above rRNA motif (3′CCUCA to 3′CCCUCA), followed by mRNA (5′ UTR) adaptation to this alteration, enabling an altered SD interaction, referred to as a GGGA interaction (3′CCCU/5′GGGA) (stage [C1+A] in Fig. 11A). This allowed both a canonical SD

interaction (with one of the two canonical SD pairs, 3’CCUC/5’GGAG) and a GGGA interaction (3’CCCU/5’GGGA) (Fig. 11B). Such rRNA change was observed in a major type rRNA (three / four small subunit rRNA genes) of the *E. gracilis* plastid. There was emergence of GGGA mRNA signals in four genes, three of which are for photosystems II (*psbE*, *psbK*, and *psbJ*) (Fig. 11BC). Two genes, *psbJ* and pcog01, had intermediate mRNA signals that covered both interactions (canonical and GGGA interactions) (Fig. 11BC). This intermediate signals suggested that the transition from canonical SD to GGGA interaction may have proceeded through an intermediate step. Fig. 11C shows that the mRNA signals were present at similar locations relative to the start codon regardless of the type of SD interaction, suggesting that the GGGA signals follow the behavior of the canonical SD interactions.

Elimination: Another “C → U” substitution in 3′CCCUCA in the rRNA motif, which generated 3′CCCUUA, was found in a minor type rRNA (i.e., one of the four small subunit rRNA genes) of the *E. gracilis* plastid, completely disallowing the canonical SD interaction and allowing only the GGGA interaction (stage [A] in Fig. 11A).

(b) Chlorophyta

Plastids of Chlorophyceae belonging to Chlorophyta represnted the most clear sequential changes of extension, limitation, and elimination (Fig. 12). In this case, the altered SD interaction was 3’CUUCC/5’GAAGG (rRNA/mRNA) (referred to as GAAGG interaction).


A
3′ rRNA
5′ mRNA
stage
SD (rRNA-mRNA) interaction
Canonical
Altered (GAAGG)
Reversal
Extension
Limitation
Elimination
[C2]
[C2+A]
[C1+A]
[A]
3′ CCUCC
5′ GAGG
3′ CCUCC
5′ GGAG
3′ CUUCCUCC
5′ GAAGG
SD interaction
[C1]
3′ CCUCUU
B
SD interactions
canonical
GAAGG
miscellaneous
no interaction
no gene
a: Trebouxiophyceae
b: Chlorophyceae
anti-SD
seq (rRNA)
psbK psbI psbA rps19 pcog02 pcog03 psbB tufA pcog04 ycf1 psbJ pcog06 psaB psbE psbC atpH psaA psbD rps9 chlB psbM rpoC2 atpA petD ycf3 atpF psaC psaJ rpoA rps18 cemA petL psbL rpl20 rpl5 rps7 rps8 ycf4 rpl32 pcog08 pcog05 pcog07 psaI rpoB rpl12 accD rpl19
C. subellipsoidea AAUUCCUCC
C. variabilis AUUUCCUCC
C. reinhardtii ACUUCCUCC
D. salina UCUUCCUCC
O. cardiacum ACUUCCUCC
F. terrestris UAUCCUCUU
S. leibleinii UUCUUCCUC
S. helveticum AUUCUUCCU

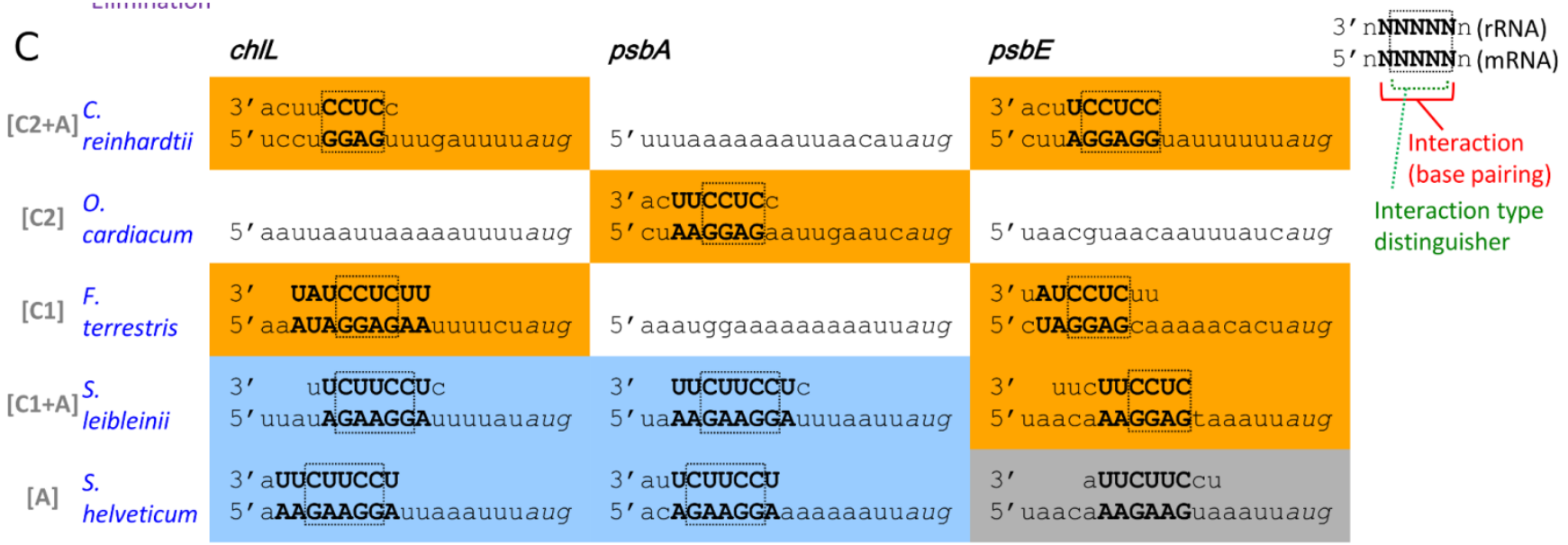


FIG. 12. rRNA-mRNA coevolution in Chlorophyta plastids. (A) Inferred stages of the coevolution. Left: The emergence of an altered SD interaction (GAAGG interaction) through an extension of the canonical anti-SD motif and mRNA adaptation, was followed by serial loss of the canonical SD interactions. Right: One of the canonical SD interactions was disallowed. (B) Anti-SD motifs (in anti-SD sequences) and a matrix for plastid genes at various evolutionary stages. Boxes in the matrix indicate genes and are colored according to their SD interactions. Multiple boxes denote different paralogous genes. Miscellaneous: secondary structures with the bulge or internal loop structure, or non-Watson-Crick base paring. See Table S3 for the orthologous group, pcog01 and pcog09. (C) Alignments of predicted SD interactions for several orthologous genes. This figure is adapted from Lim et al.(submitted).

Extension: In the extension step, the canonical anti-SD motif (3'CCUCC) in rRNA is extended to the 3' side by three nucleotides (3'CUUCCUCC). The extend motif can engage in both the GAAGG and canonical SD interactions (stage [C2+A] in Fig. 12A). My analysis of the phylogeny (Fig. 12B) revealed that this may have occurred in a common ancestor of plastids belonging to Chlorophyceae (Chlorophyta).

Limitation: Next, a canonical SD interaction was disallowed by deletion of one C at the 5′ end from the canonical anti-SD motif (3′CCUCC) within the extended motif, (3′CUUCCUCC), generating 3′CUUCCUC (stage [C1+A] in Fig. 12A). The phylogeny indicated taht this took place in the common ancestor of *Schizomeris leibleinii* and *Stigeoclonium helveticum* plastids (stage [C1+A] in Fig. 12B). Although this anti-SD motif variant allowed both types of interaction, genes with GAAGG interactions outnumbered genes with the canonical SD interaction in the plastid of *S. leibleinii* (Fig. 12B).

Elimination: Finally, evolution of the rRNA/mRNA pairs was followed by deletion of another C to generate 3′CUUCCU from 3′CUUCCUC in the plastid of *S. helveticum* (stage [A] in Fig. 12 A and B). This variant used only the altered (GAAGG) interaction because a severe decay in the canonical anti-SD motif prohibited a canonical SD interaction (Fig. 12B).

In support of this route, 5′ UTRs in some protein-coding genes have evolved to be paired with the altered anti-SD motifs and their flanking regions in the plastids of *S. leibleinii* and *S. helveticum* (Fig. 12C, *chlL* and *psbA*). A similarly coordinated change was also

observed between *S. leibleinii* and *S. helveticum* plastids. In the *psbE* gene, the mRNA signal change (5′GGAG to 5′GAAG) offset the alteration in the anti-SD motif (3′CUUCCUC to 3′CUUCCU) (Fig. 12C). During this coevolution, there were systematic changes in the repertoire of SD signal-carrying genes. For example, some genes (*rps19*, *psbB*, and *tufA*) that had previously lacked an SD interaction signal at some point acquired an altered signal (GAAGG), to stages [C1+A] (of the *S. leibleinii* plastid) and [A] (of the *S. helveticum* plastid) (Fig. 12B). In these plastids, some newly transferred genes (refered to as pcog02, pcog03, and pcog04) also acquired a GAAGG interaction signal in the 5′UTRs (Fig. 12B). These genes appeared to be DNA endonucleases such as homing endonucleases (Table S3). During the same period, other genes (*psaA*, *psbD*, *psbC*, and *atpH*) that had depended on the canonical anti-SD motif lost their capacity for SD interaction likey becasue serial decay events involving the canonical motif (Fig. 12B) weakened it complementarity with canonical SD signals. These phenomena suggest a strong association between rRNA-mRNA coevolution and gene translation patterns.

In addition to the above-described route, evolution of the canonical and altered SD interactions also proceeded via two other routes. First, the [C2+A] stage may have independently reverted to the original [C2] stage in the plastids of *Oedogonium cardiacum* and *Floydiella terrestris* (Fig. 12B). Second, a substitution at the 5′ end of the canonical anti-SD motif took place in the *F. terrestris* plastid, which has an SD-dependent gene inventory similar to that of closely-related plastids (stage [C1] in Fig. 12 A (right) and B).

(c) Statistical significance

I tested whether the observed occurrence of altered mRNA signals for altered SD interaction was statistically significant compared to the probability of observing the same consensus signal in random plastid genome regions. I found that GGGA (in the *E. gracilis* plastid) and GAAGG (in *S. leibleinii* and *S. helveticum* plastids) interactions showed statistically significant abundance ($p < 0.01$), supporting the mRNA adaptations to the altered SD interactions were not just casual association (Table 3).

Table 3. Statistically significant ($p < 0.01$) altered SD interactions

Canonical anti-SD motif[a]	Taxonomic group	Species	mRNA signal consensus[b]	Gene # (%) with the consensus[c]	Expected % (p) [d]
Conserved	Streptophyta	*Zygnema circumcarinatum*	GAAGG	6 (6.5)	0.55 (1.35e-05)
Conserved	Chlorophyta	*Pseudendoclonium akinetum*	GAAGG	8 (8.3)	0.57 (9.8e-08)
Conserved	Chlorophyta	*Dunaliella salina*	GAAGG	4 (4.9)	0.58 (1.40e-03)
Conserved	Chlorarachniophyta	*Bigelowiella natans*	GAAGG	17 (28)	0.52 (6.3e-25)
Reduced	Euglenophyta	*Euglena gracilis*	GGGA	4 (6.5)	0.90 (2.4e-03)
Reduced	Chlorophyta	*Schizomeris leibleinii*	GAAGG	8 (10.7)	0.45 (2.2e-09)
Reduced	Chlorophyta	*Stigeoclonium helveticum*	GAAGG	7 (9.2)	0.44 (5.1e-08)

[a]Reduced: changes ≤ 2 nt (insertion, deletion, or substitution) compared to the canonical anti-SD motif (3′CCUCC).

[b]mRNA signal consensus of an altered rRNA-mRNA interaction. See Figs. 11, 12, and 13 for details.

[c]Number and percentage of protein-coding genes with the consensus in their SD regions (−20 to −5 (16 nt) from the start codon).

[d]Probability (%) of observing the consensus among artificially generated 16 nt sequences ($n = 10^6$) using the nucleotide frequencies of the genome. p for a binomial test of the null hypothesis that the actual number of consensus observed is equal to the expected number.

This table is adapted from Lim et al.(submitted).

(d) Extensions without limitation or elimination.

Extension (the emergence of altered SD interactions) did not necessarily entail decay of the canonical anti-SD motif. GAAGG interaction showed statistically significant abundance ($p < 0.01$) in four conserved anti-SD plastids belonging to Streptophyta, Chlorophyta, and Chlorarachniophyta (Table 3).

Among these, the *B. natans* (Chlorarachniophyta) plastid showed the highest abundance of GAAGG signal-carrying genes. In addition, an extension in the *B. natans* plastid yielded a GAAGG interaction (3′CUUCC/5′GAAGG) as described for Chlorophyceae plastids (stage [C2+A] in Fig. 13A). Next steps of the coevolution have not been obsered. The GAAGG interaction accounted for the majority of SD interaction signals (Fig. 13B), and contributed to *B. natans* showing the highest dependency on SD interaction (F_{SD}) of all plastids outside Streptophyta (Fig. 9B). This could be due to the plastid undergoing the early stages of secondary endosymbiosis, as inferred from the retention of the nucleomorph (a remnant of the nucleus of an endosymbiotic alga), in addition to other lines of evidence (Gilson et al. 2006).

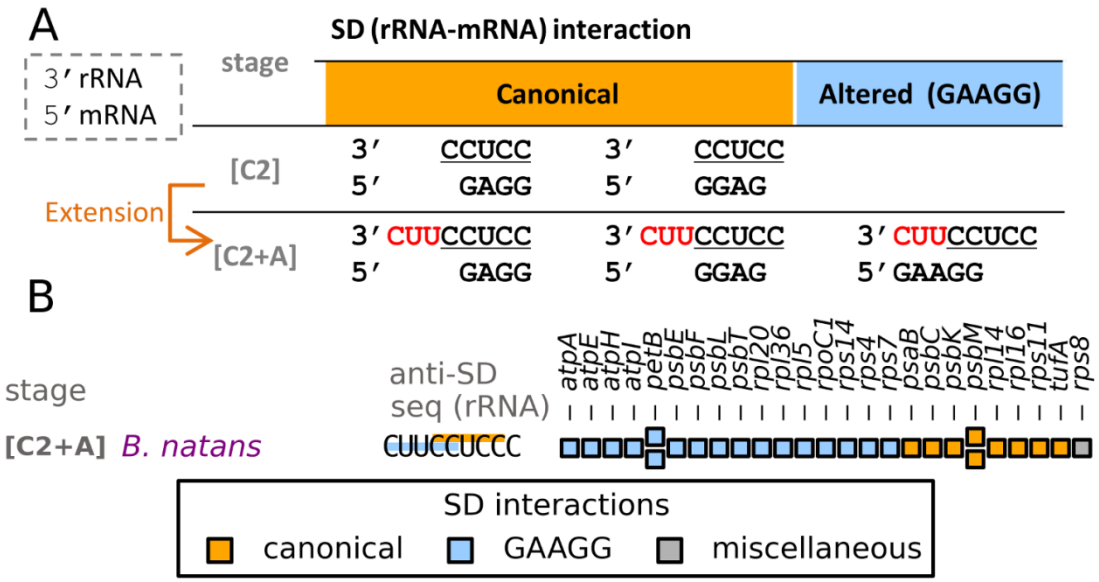


FIG. 13. rRNA-mRNA coevolution in the *B. natans* (Chlorarachniophyta) plastid. (A) Inferred stages of the coevolution, leading to the emergence of an altered SD interaction (GAAGG interaction). (B) Anti-SD motifs (in anti-SD sequences) and genes. Genes (boxes) are colored according to their SD interactions. Two boxes are shown when two paralogous genes are present. Miscellaneous: secondary structures that cannot be categorized as one of the interactions in panel A, mostly due to the bulge or internal loop structure, or non-Watson-Crick base paring. This figure is adapted from Lim et al.(submitted).

Abrupt replacement of the canonical anti-SD motif would presumably be detrimental to the plastid and its host, as it would hinder the expression of many genes. The proposed route with intermediate steps could have provided a safer route for the evolution of the prokaryotic translation initiation system. This model is comparable to significant changes in codon-anticodon interaction (e.g., arising from codon reassignment) that hae often been described using stepwise models that can tolerate potentially deleterious effects resulting from such changes (Osawa and Jukes 1989; Andersson and Kurland 1991; McCutcheon et al. 2009).

I assume that a reduced number of intra-genomic protein-coding genes that depend on canonical SD interactions may have acclerated the rRNA-mRNA coevolution, by increasing the chances for a newly emerged SD interaction to dominate the genome. The new SD interaction caused a relaxation of the evolutionary constraints on the canonical anti-SD motif, allowing the canonical anti-SD motif to undergo a series of single nucleotide mutations, until the newly emerged interaction was fixed within the genome.

The relationship between the replacement (by an altered SD interaction) and loss of SD interaction is elusive. In Euglenophyta, the replacement (in *E. gracilis*) appears to have preceded loss (in *E. longa*); it remains to be determined whether this represents the last stage of the coevolution, or an independent phenomenon. In other cases of alteration, there was no obvious association between the two events.

4.7. rRNA: a Driving Force of mRNA Evolution Leading to Adaptive Evolution?

It was proposed that ribosome has evolved through coevolution of rRNAs and ribosomal proteins (Harish and Caetano-Anollés 2012). rRNA evolution likely drove the compensated changes in rRNA-interacting ribosomal proteins (Barreto and Burton 2013) leading to ribosomal evolution. Beyond this view that regarded rRNA evolution as a driver of the evolution of ribosomal proteins, our study demonstrates that rRNA evolution can also drive mRNA evolution affecting the translational efficiency. Each step of our stepwise coevolutionary route can be approximately explained by the occurrence of a mutation in or near the anti-SD motif of an rRNA gene, followed by adaptive evolution of SD signals in mRNAs. The sequences of SD signals were consistent with this order of events (Figs. 11 and 12). Furthermore, I hypothesize that the rRNA-driven mRNA evolution was potentially relevant to changes in SD signal-dependent genes, possibly affecting their expression patterns. This event might have happened during the evolution towards the *S. leibleinii* and *S. helveticum* plastids, where the gain and loss of SD signals in some genes were observed (Fig. 12B). Together, our results suggest that the rRNA-driven evolutionary force potentially exerts a broad impact on the genome, thereby driving its evolution.

We hypothesize that the biological significance of the alteration in SD interaction is in systematic change of expression pattern of a set of genes. This can be directly tested by proteome comparison of this plastid and closely-related plastids although we do not have such closely-related plastid genome sequences yet. Furthermore, we imagine that such

changes may have helped in adaptation during progression of secondary symbiosis in Euglenophyta (*Euglena gracilis*) and in Chlorachniophyta (*Biogelowiella natans*), and during some other processes in the primary plastids (Chlorophyceae). In *E. gracilis* plastid, a few genes involving photosystem II acquired altered SD signals (Fig. 11B). In *B. natans*, many genes including several photosystem II genes and several ATPase genes for ATP synthesis acquired the signal (Fig. 13B). In the primary plastids in Chlorophyceae, the genes with the altered SD signal include two genes for photosystem II (Fig. 12B).

Photosystem II oxidizes water with a stronger oxidizing agent. It is sensitive to strong light and other environmental stresses(Murata et al. 2007; Saibo et al. 2009). Reactive oxygen species generated by different stresses inhibit photosystem II repair by suppressing the transcription and translation of a protein of the photosystem II complex. Several transcription factors are implicated in the stress response. Based on these observations, we imagine the translational control of photosystem II genes, suggested above, could have helped in adaptation to secondary endosymbiosis or other environmental challenges in terms of photosynthesis. Such possible biological significance of the SD signal alteration is worth being examined by further genome comparison, omics analysis and biological experiments.

4.8. Materials and Methods

4.8.1. Genome Sequences and Protein-Coding Genes.

The RefSeq collection (http://www.ncbi.nlm.nih.gov/refseq/) of complete plastid genome sequences (n = 430) was downloaded on November 20, 2013 (the list can be found in Table S2). Structural annotation information for the genomes was retrieved from the nucleotide database (http://www.ncbi.nlm.nih.gov/nuccore/) of NCBI. I could obtain registered coding sequence annotation information of 428 of the total 430 plastid genomes sequences. Among the two genomes lacking annotation information, I annotated the *P. falciparum* HB3 plastid genome (NC_017928.1) for the small subunit rRNA gene and protein-coding genes (Table S4) according to a previous annotation report for the plastid of the same species (Wilson et al. 1996). I searched highly similar sequences between the two genomes using BLASTn 2.2.27+ for the annotation (Zhang et al. 2000). Including the *P. falciparum* HB3 plastid genome, the final plastid genome set had 429 sequences (Table S2).

I only used protein-coding genes occurring in multiple genomes to avoid pseudogenes and to standardize the content of protein-coding gene sets. More specifically, I found similar protein sequences across genoems where the similar protein sequence was defined as a hit with the Expect value < 0.001 when searched using BLASTp 2.2.27+ (Altschul et al. 1997) against the nr BLAST database (ftp://ftp.ncbi.nih.gov/blast/db/) of NCBI.

4.8.2. Predicting the minimum free energy (MFE) structure between two RNA strands.

I used RNAcofold algorithm in the Vienna RNA package 2.0.7 (Lorenz et al. 2011) with the "-d0" and "--noLP" options to find the best hybridization structure between two given RNA strands. The free energy change, referred to as MFE, was measured for the resulting best structure only considering inter-strand base parings, using RNAeval algorithm in the Vienna RNA package 2.0.7 with the "-d0" option.

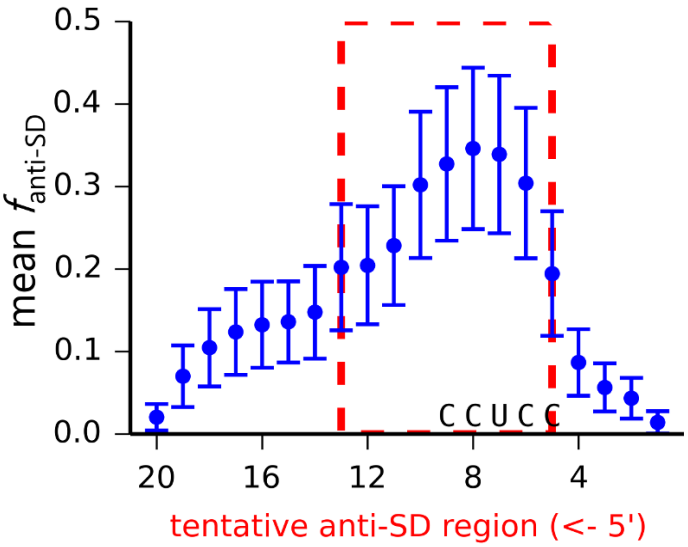


FIG. 14. Means and standard deviations of $f_{anti\text{-}SD}$ values in the tentative anti-SD regions (+20 nt from the 5′ start of the small subunit rRNA 3′ tail) for all plastid genomes assessed in this study. $f_{anti\text{-}SD}$ represents the site-specific base pairing frequency calculated for each genome. Inter-strand base pairings from the hybridization structures with MFE ≤ −3.7 kcal/mol were included. Mean $f_{anti\text{-}SD}$ values within the area defined by the broken line are significantly higher than other tentative anti-SD positions, and are thus defined as the anti-SD region for the corresponding plastid genome set. The position of the canonical anti-SD motif (3'CCUCC) within the tentative anti-SD region is shown. See Materials and Methods for detail. This figure is adapted from Lim et al.(submitted)

4.8.3. Locating Small Subunit rRNA 3′ Tails and Anti-SD Sequences.

I located 3′ tails that follow helix 45 of small subunit rRNAs (Figs. 1 and 8) based on experimentally verified locations and sequences of several 3′ tails (Steege et al. 1982). According to the conservation level of the canonical anti-SD motif in the 3′ tails, I classified plastids into conserved, altered, and lost anti-SD plastids as described in Table 2 and Fig. 8.

I next attempted to locate the anti-SD sequences of these plastids. I hypothesized that the region of an anti-SD sequence (i.e., the anti-SD region) was likely to form secondary structures with SD regions in its own genome. I defined the +20 nt region from the 5′ start of the 3′ tail, which was larger than known 3′ tail lengths, as a tentative anti-SD region. Based on an earlier study (Nakagawa et al. 2010), I next set the sub-region within the 5′ UTR of an mRNA in which SD sequences were searched (i.e., the SD region) as -20 nt to -5 nt from the start codon. MFE structures between a tentative anti-SD region and SD regions of the same genome were predicted and only those with an MFE ≤ -3.7 kcal/mol were retained for further analysis, since this value represented the free energy of the four base pairs in canonical SD interactions, 3′CCUC/5′GGAG and 3′CUCC/5′GAGG. Next, I measured the site-specific base-pairing frequency ($f_{\text{anti-SD}}$) in the tentative anti-SD region for each genome. For example, an $f_{\text{anti-SD}}$ value of 0.5 indicated the specific anti-SD position formed by base pairings with the SD regions in 50% of the protein-coding transcriptome (Fig. 14). For the plastid set, the 5 nt to 13 nt sub-region of the tentative anti-SD region showed significantly higher mean $f_{\text{anti-SD}}$ values than other positions with the region of the canonical anti-SD motif included (Fig.

14). I labeled this sub-region an anti-SD region. The 3′ end of this sub-region was considered as the 3′ end of the 3′ tail for all plastid genomes I assessed. The only exception was the plastid of *E. gracilis*, whose 3′ tail sequence was already known (Steege et al. 1982).

4.8.4. Predicting SD Interactions

I predicted MFE structures between an anti-SD sequence and all SD regions of the same genome. When there were multiple types of anti-SD sequences within a genome, the type observed in the largest number of small subunit rRNA gene copies was used. A sequence within the SD region was designated as an SD sequence if the MFE value of the hybridization structure was ≤ -3.7 kcal/mol. Finally, I calculated the fraction of SD sequence-containing protein-coding genes for each genome (F_{SD}) and used it as an index of intra-genomic SD interaction usage. For reduced and lost anti-SD plastids that had variations in their anti-SD motifs, this strategy detected potential rRNA-mRNA interactions that did not obey the canonical SD interaction. These interactions were regarded as SD interactions in a broad sense.

4.8.5. Constructing Phylogenetic Trees.

I references previous literature (Keeling 2013) for plastid phylogeny shown in Figs. 2 and 9A. In addition, I constructed a multi-gene phylogenetic tree for the plastids of selected species in Chlorophyta, Euglenophyta, and Chlorarachniophyta. The procedure of tree construction was:

(i) I found common ribosomal protein genes (rpl2, rpl4, rpl14, rpl16, rpl20, rpl36, rps3, rps4, rps7, rps8, rps11, rps12, rps14, and rps19) in the plastids.

(ii) I aligned protein sequences of each orthologous gene set using MAFFT 6.927b (Katoh and Toh 2008) with the L-INS-I option.

(iii) I removed poorly aligned regions using Gblocks 0.91b (Talavera and Castresana 2007). The trimmed alignments were concatenated for each species.

(iv) After the best-fit model selection (the model was cpREV+I+G4), maximum likelihood tree deduction with 1000 bootstrap replications for the concatenated alignments were conducted using IQ-TREE tool version 0.9.5 (Minh et al. 2013). The resulting tree shown in Fig. S1 was in good agreemen with previously deduced phylogenetic trees (Turmel, Gagnon, et al. 2009; Turmel, Otis, et al. 2009; Brouard et al. 2011; Hrdá et al. 2012).

5. Closing Remarks

Precise RNA-RNA interaction is central to many biological processes; for example, rRNA and tRNA folding into their highly conserved structures is indispensible to translation (Brink et al. 1993; Sherlin et al. 2000). Due to their functional importance, these RNA-RNA interactions have structural constraint, thus their molecular evolution often necessitates coordinated variation of paired loci (Gultyaev et al. 2000). As the interacting partners grow, cost for the coordinated variation becomes greater, so does a stronger evolutionary constraint. This indicates that biologically crucial RNA-RNA interactions that involve a multitude of loci probably are under an extensive evolutionary constraint, examples of which are codon-anticodon (tRNA-mRNA) interaction and the SD interaction (rRNA-mRNA). Supporting this hypothesis, their pairing motifs, the genetic code and the SD interaction, respectively, are completely conserved with rare exceptions in the genetic code (Knight et al. 2001).

In the present study, I revealed numerous exceptions of SD interactions. Such exceptions can be categorized into two types: one is the SD interaction loss (entirely from a genome) and the other is the alteration in SD interaction motifs. It has been already known that SD interactions were eliminated in most mitochondria (Burger et al. 2013) and a primary endosymbiont (Thao et al. 2000). This study demonstrate that such loss was occurred many more lineages of bacteria and plastids. The loss in Flavobacteria was attributable to a distinct evolutionary force, the emergence and prevalence of an alternative A-rich signal whose location overlaps SD signals'. The loss in other lineages commonly occurred in tiny genomes with a few exceptions. This indicates some relationship of the

loss with extensive genome reduction. It is, however, noteworthy that a tiny genome is not necessarily an indicative of SD interaction loss as I showed in many lineages of bacteria and plastids. This suggests that there are other unknown factors associated with this loss.

The alteration in SD interaction motifs I found in several plastid lineages are comparable to the exceptional changes in the codon-anticodon interaction (the codon reassignment), which have been seen mitochondria and Mycoplasma species (Knight et al. 2001). The 'genome streamlining' hypothesis posits that reductive evolution of their genomes has likely driven the codon reassignment by selection to smaller number of tRNA species (Andersson and Kurland 1991; Andersson and Kurland 1998; Knight et al. 2001). One possible explanation for this is that decreased abundance of protein-coding genes thus that of codons and anticodons relaxed the evolutionary constraint on codon-anticodon (tRNA-mRNA) interaction. In a similar manner, genome reduction, together with decreased dependency on the SD interaction, can weaken the evolutionary constraint on the SD interaction by decreasing the number of intra-genomic SD sequences. This explains our unexpected discovery of coevolution of the anti-SD/SD (rRNA/mRNA) motif pair in plastid genomes with a small number of SD-sequence-carrying protein-coding genes. It seems that the evolution has proceeded by a series of single nt level changes, eventually bringing about plasticity in the rRNA-mRNA interaction, which has never seen in other prokaryotic systems. It is unclear whether the new interactions will persist as they are because their plastids are still keeping small numbers of protein-coding genes that depend on the rRNA-mRNA interaction for their translation initiation.

Here I reported unexpected fates of SD interactions through a large-scale analysis of bacterial and plastid genome sequences. Such fates are 1) loss and 2) alteration of SD interactions. The loss and alteration occurred in many lineages in parallel especially in genomes that have experienced a massive genome reduction process. In Flavobacteria, a novel interaction that uses A-rich signals appeared to have alternated SD interactions, resulting in the loss. This result demonstrates an aspect of evolutionary plasticity in translation regulation mechanisms.

6. Acknowledgements

I would like to express my sincere thanks to Professor Ichizo Kobayashi, from whom I first learned the field of bioinformatics. He also taught me scientific thinking and writing, without which this dissertation would not have been possible.

I would like to express my sincere appreciation to Professor Kenta Nakai, who kindly supervised and supported me to complete my research project. Thanks to him, I could have learned various bioinformatics skills and trends, which broadened my scientific vision.

Finally, I would like to express the most special thanks to my lovely wife, Saaya Lim Tsutsué, who was the strongest motivation for pursuing my career as a researcher. Her existence itself was the greatest power that accomplished this dissertation.

This dissertation was co-authored by Assistant Professor Yoshikazu Furuta, Professor Ichizo Kobayashi and Professor Kenta Nakai. Part 1 (chapter 3) of this dissertation is based on previously published work, Lim et al. (2012), which is Open Access (cc by-nc), and on a submitted manuscript, Lim et al. (submitted). Financial support was provided by Japanese government (MEXT) scholarship.

7. References

Agrawal S, Striepen B. 2010. More membranes, more proteins: complex protein import mechanisms into secondary plastids. Protist 161:672–687.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

Andersson GE, Kurland CG. 1991. An extreme codon preference strategy: codon reassignment. Mol. Biol. Evol. 8:530–544.

Andersson SG, Kurland CG. 1998. Reductive evolution of resident genomes. Trends Microbiol. 6:263–268.

Barreto FS, Burton RS. 2013. Evidence for compensatory evolution of ribosomal proteins in response to rapid divergence of mitochondrial rRNA. Mol. Biol. Evol. 30:310–314.

Betts L, Spremulli LL. 1994. Analysis of the role of the Shine-Dalgarno sequence and mRNA secondary structure on the efficiency of translational initiation in the Euglena gracilis chloroplast atpH mRNA. J. Biol. Chem. 269:26456–26463.

Bonham-Smith P, Bourque D. 1989. Translation of chloroplast-encoded mRNA: potential initiation and termination signals. Nucleic Acids Res. 17:2057–2080.

Boni I V, Isaeva DM, Musychenko ML, Tzareva N V. 1991. Ribosome-messenger recognition: mRNA target sites for ribosomal protein S1. Nucleic Acids Res. 19:155–162.

Boucias DG, Becnel JJ, White SE, Bott M. 2001. In vivo and in vitro development of the protist Helicosporidium sp. J. Eukaryot. Microbiol. 48:460–470.

Brink MF, Verbeet MP, de Boer HA. 1993. Formation of the central pseudoknot in 16S rRNA is essential for initiation of translation. EMBO J. 12:3987–3996.

Brouard J-S, Otis C, Lemieux C, Turmel M. 2011. The chloroplast genome of the green alga Schizomeris leibleinii (Chlorophyceae) provides evidence for bidirectional DNA replication from a single origin in the chaetophorales. Genome Biol. Evol. 3:505–515.

Burger G, Gray MW, Forget L, Lang BF. 2013. Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. Genome Biol. Evol. 5:418–438.

Chang B, Halgamuge S, Tang S-L. 2006. Analysis of SD sequences in completed microbial genomes: non-SD-led genes are as common as SD-led genes. Gene 373:90–99.

Delannoy E, Fujii S, Colas des Francs-Small C, Brundrett M, Small I. 2011. Rampant gene loss in the underground orchid Rhizanthella gardneri highlights evolutionary constraints on plastid genomes. Mol. Biol. Evol. 28:2077–2086.

dePamphilis CW, Palmer JD. 1990. Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. Nature 348:337–339.

Draper DE, Pratt CW, von Hippel PH. 1977. Escherichia coli ribosomal protein S1 has two polynucleotide binding sites. Proc. Natl. Acad. Sci. U. S. A. 74:4786–4790.

Farwell MA, Roberts MW, Rabinowitz JC. 1992. The effect of ribosomal protein S1 from Escherichia coli and Micrococcus luteus on protein synthesis in vitro by E. coli and Bacillus subtilis. Mol. Microbiol. 6:3375–3383.

Freier SM, Kierzek R, Jaeger JA, Sugimoto N, Caruthers MH, Neilson T, Turner DH. 1986. Improved free-energy parameters for predictions of RNA duplex stability. Proc. Natl. Acad. Sci. U. S. A. 83:9373–9377.

Gantt JS, Baldauf SL, Calie PJ, Weeden NF, Palmer JD. 1991. Transfer of rpl22 to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. EMBO J. 10:3073–3078.

Gilson PR, Su V, Slamovits CH, Reith ME, Keeling PJ, McFadden GI. 2006. Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. Proc. Natl. Acad. Sci. U. S. A. 103:9566–9571.

Gray MW, Doolittle WF. 1982. Has the endosymbiont hypothesis been proven? Microbiol. Rev. 46:1–42.

Guimaraes AMS, Santos AP, SanMiguel P, Walter T, Timenetsky J, Messick JB. 2011. Complete genome sequence of Mycoplasma suis and insights into its biology and adaption to an erythrocyte niche. PLoS One 6:e19574.

Gultyaev AP, Franch T, Gerdes K. 2000. Coupled nucleotide covariations reveal dynamic RNA interaction patterns. RNA 6:1483–1491.

Harish A, Caetano-Anollés G. 2012. Ribosomal history reveals origins of modern protein synthesis. PLoS One 7.

Hashimoto JG, Stevenson BS, Schmidt TM. 2003. Rates and consequences of recombination between rRNA operons. J. Bacteriol. 185:966–972.

Hazle T, Bonen L. 2007. Comparative analysis of sequences preceding protein-coding mitochondrial genes in flowering plants. Mol. Biol. Evol. 24:1101–1112.

Hirose T, Sugiura M. 2004a. Functional Shine-Dalgarno-like sequences for translational initiation of chloroplast mRNAs. Plant Cell Physiol. 45:114–117.

Hirose T, Sugiura M. 2004b. Multiple elements required for translation of plastid atpB mRNA lacking the Shine-Dalgarno sequence. Nucleic Acids Res. 32:3503–3510.

Hrdá Š, Fousek J, Szabová J, Hampl V, Vlček Č. 2012. The plastid genome of Eutreptiella provides a window into the process of secondary endosymbiosis of plastid in euglenids. PLoS One 7:e33746.

Janouskovec J, Horák A, Oborník M, Lukes J, Keeling PJ. 2010. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. Proc. Natl. Acad. Sci. U. S. A. 107:10949–10954.

Jansen RK, Saski C, Lee S-B, Hansen AK, Daniell H. 2011. Complete plastid genome sequences of three Rosids (Castanea, Prunus, Theobroma): evidence for at least two independent transfers of rpl22 to the nucleus. Mol. Biol. Evol. 28:835–847.

Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2014. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 42:D199–205.

Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. 9:286–298.

Keeling PJ. 2013. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. Annu. Rev. Plant Biol. 64:583–607.

Khakhlova O, Bock R. 2006. Elimination of deleterious mutations in plastid genomes by gene conversion. Plant J. 46:85–94.

Knight RD, Freeland SJ, Landweber LF. 2001. Rewiring the keyboard: evolvability of the genetic code. Nat. Rev. Genet. 2:49–58.

Kosuri S, Goodman DB, Cambray G, Mutalik VK, Gao Y, Arkin AP, Endy D, Church GM. 2013. Composability of regulatory sequences controlling transcription and translation in Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. 110.

Lang B, Burger G, O'Kelly C. 1997. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature.

Larkin MA, Blackshields G, Brown NP, et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948.

Lim K, Furuta Y, Kobayashi I. 2012. Large variations in bacterial ribosomal RNA genes. Mol. Biol. Evol. 29:2937–2948.

Lim K, Kobayashi I, Nakai K. Alterations in rRNA-mRNA interaction during plastid evolution (submitted).

Logacheva MD, Schelkunov MI, Penin A a. 2011. Sequencing and analysis of plastid genome in mycoheterotrophic orchid Neottia nidus-avis. Genome Biol. Evol. 3:1296–1303.

Lorenz R, Bernhart SH, Höner Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. Algorithms Mol. Biol. 6:26.

Ma J, Campbell A, Karlin S. 2002. Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. J. Bacteriol. 184:5733–5745.

Malys N, McCarthy JEG. 2011. Translation initiation: variations in the mechanism can be anticipated. Cell. Mol. Life Sci. 68:991–1003.

Manuell AL, Quispe J, Mayfield SP. 2007. Structure of the chloroplast ribosome: novel domains for translation regulation. PLoS Biol. 5:e209.

Martin W, Rujan T, Richly E, et al. 2002. Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc. Natl. Acad. Sci. U. S. A. 99:12246–12251.

McCutcheon JP, McDonald BR, Moran N a. 2009. Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. PLoS Genet. 5:e1000565.

McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. Nat. Rev. Microbiol. 10:13–26.

Millen RS, Olmstead RG, Adams KL, et al. 2001. Many parallel losses of infA from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell 13:645–658.

Minh BQ, Nguyen MAT, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30:1188–1195.

Murata N, Takahashi S, Nishiyama Y, Allakhverdiev SI. 2007. Photoinhibition of photosystem II under environmental stress. Biochim. Biophys. Acta 1767:414–421.

Nakagawa S, Niimura Y, Miura K, Gojobori T. 2010. Dynamic evolution of translation initiation mechanisms in prokaryotes. Proc. Natl. Acad. Sci. U. S. A. 107:6382–6387.

Osawa S, Jukes TH. 1989. Codon reassignment (codon capture) in evolution. J. Mol. Evol. 28:271–278.

Palmer JD. 1985. Comparative organization of chloroplast genomes. Annu. Rev. Genet. 19:325–354.

Reyes-Prieto A, Weber APM, Bhattacharya D. 2007. The origin and establishment of the plastid in algae and plants. Annu. Rev. Genet. 41:147–168.

Rodríguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Löffelhardt W, Bohnert HJ, Philippe H, Lang BF. 2005. Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. Curr. Biol. 15:1325–1330.

Sagan L. 1967. On the origin of mitosing cells. J. Theor. Biol. 14:255–274.

Saibo NJM, Lourenço T, Oliveira MM. 2009. Transcription factors and regulation of photosynthetic and related metabolism under environmental stresses. Ann. Bot. 103:609–623.

Salah P, Bisaglia M, Aliprandi P, Uzan M, Sizun C, Bontems F. 2009. Probing the relationship between Gram-negative and Gram-positive S1 proteins by sequence analysis. Nucleic Acids Res. 37:5578–5588.

Sartorius-Neef S, Pfeifer F. 2004. In vivo studies on putative Shine-Dalgarno sequences of the halophilic archaeon Halobacterium salinarum. Mol. Microbiol. 51:579–588.

Schurr T, Nadir E, Margalit H. 1993. Identification and characterization of E.coli ribosomal binding sites by free energy computation. Nucleic Acids Res. 21:4019–4023.

Schwarz C, Elles I, Kortmann J, Piotrowski M, Nickelsen J. 2007. Synthesis of the D2 protein of photosystem II in Chlamydomonas is controlled by a high molecular mass complex containing the RNA stabilization factor Nac2 and the translational activator RBP40. Plant Cell 19:3627–3639.

Sengupta J, Agrawal RK, Frank J. 2001. Visualization of protein S1 within the 30S ribosomal subunit and its interaction with messenger RNA. Proc. Natl. Acad. Sci. U. S. A. 98:11991–11996.

Sharma CM, Hoffmann S, Darfeuille F, et al. 2010. The primary transcriptome of the major human pathogen Helicobacter pylori. Nature 464:250–255.

Sherlin LD, Bullock TL, Newberry KJ, Lipman RS, Hou YM, Beijer B, Sproat BS, Perona JJ. 2000. Influence of transfer RNA tertiary structure on aminoacylation efficiency by glutaminyl and cysteinyl-tRNA synthetases. J. Mol. Biol. 299:431–446.

Shine J, Dalgarno L. 1974. The 3’-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. U. S. A. 71:1342–1346.

Siemeister G, Hachtel W. 1989. A circular 73 kb DNA from the colourless flagellate Astasia longa that resembles the chloroplast DNA of Euglena: restriction and gene map. Curr. Genet. 15:435–441.

Starmer J, Stomp a, Vouk M, Bitzer D. 2006. Predicting Shine-Dalgarno sequence locations exposes genome annotation errors. PLoS Comput. Biol. 2:e57.

Steege DA, Graves MC, Spremulli LL. 1982. Euglena gracilis chloroplast small subunit rRNA. Sequence and base pairing potential of the 3’ terminus, cleavage by colicin E3. J. Biol. Chem. 257:10430–10439.

Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56:564–577.

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10:512–526.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28:2731–2739.

Tengs T, Dahlberg OJ, Shalchian-Tabrizi K, Klaveness D, Rudi K, Delwiche CF, Jakobsen KS. 2000. Phylogenetic analyses indicate that the 19’Hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. Mol. Biol. Evol. 17:718–729.

Thao ML, Moran NA, Abbot P, Brennan EB, Burckhardt DH, Baumann P. 2000. Cospeciation of psyllids and their primary prokaryotic endosymbionts. Appl. Environ. Microbiol. 66:2898–2905.

Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat. Rev. Genet. 5:123–135.

Toft C, Andersson SGE. 2010. Evolutionary microbial genomics: insights into bacterial host adaptation. Nat. Rev. Genet. 11:465–475.

Turmel M, Gagnon M-C, O'Kelly CJ, Otis C, Lemieux C. 2009. The chloroplast genomes of the green algae Pyramimonas, Monomastix, and Pycnococcus shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplasts of euglenids. Mol. Biol. Evol. 26:631–648.

Turmel M, Otis C, Lemieux C. 2009. The chloroplast genomes of the green algae Pedinomonas minor, Parachlorella kessleri, and Oocystis solitaria reveal a shared ancestry between the Pedinomonadales and Chlorellales. Mol. Biol. Evol. 26:2317–2331.

Ueda M, Fujimoto M, Arimura S, Murata J, Tsutsumi N, Kadowaki K. 2007. Loss of the rpl32 gene from the chloroplast genome and subsequent acquisition of a preexisting transit peptide within the nuclear gene in Populus. Gene 402:51–56.

Wilson RJ, Denny PW, Preiser PR, et al. 1996. Complete gene map of the plastid-like DNA of the malaria parasite Plasmodium falciparum. J. Mol. Biol. 261:155–172.

Yamaguchi K, Prieto S, Beligni MV, Haynes PA, McDonald WH, Yates JR, Mayfield SP. 2002. Proteomic characterization of the small subunit of Chlamydomonas reinhardtii chloroplast ribosome: identification of a novel S1 domain-containing protein and unusually large orthologs of bacterial S2, S3, and S5. Plant Cell 14:2957–2974.

Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. Mol. Biol. Evol. 21:809–818.

Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. J. Comput. Biol. 7:203–214.

Zheng X, Hu G-Q, She Z-S, Zhu H. 2011. Leaderless genes in bacteria: clue to the evolution of translation initiation mechanisms in prokaryotes. BMC Genomics 12:361.

8. Supplementary Information

Table S1. Analysis of SD Signals of the Lost Anti-SD and Reference Genomes

a: Mean of ΔG values smaller than the cut-off

Pylum /Class	Name	Accession	Predicted anti-SD sequence	The anti-SD motif	All protein-coding genes								
					Number	Cut off: ΔG < -4.4 kcal/mol			Cut off: ΔG < -3.4535 kcal/mol			Mean ΔG	Stdev ΔG
						F_{SD}	dF_{SD}	Mean ΔG[a]	F_{SD}	dF_{SD}	Mean ΔG[a]		
Bacteroidetes /Flavobacteria	Gramella forsetii KT0803	NC_008571	GAACACCUCCUUU	O	3584	0.05	-0.05	-6.31	0.11	-0.09	-5.06	-0.48	2.07
	Zunongwangia profunda SM-A87	NC_014041	GAACACCUCCUUU	O	4653	0.06	-0.03	-6.24	0.11	-0.06	-5.11	-0.57	2.12
	Cellulophaga lytica DSM 7489	NC_015167	GAACACCUCCUUU	O	3284	0.02	-0.05	-6.29	0.05	-0.09	-4.95	0.21	1.81
	Capnocytophaga ochracea DSM 7271	NC_013162	GAACACCUCCUUU	O	2171	0.06	-0.05	-6.67	0.08	-0.11	-5.70	-0.21	2.17
	Flavobacterium johnsoniae UW101	NC_009441	GAACACCUCCUUU	O	5017	0.03	-0.05	-5.95	0.06	-0.09	-4.86	0.07	1.78
	Weeksella virosa DSM 16922	NC_015144	GAACAUCUCAUAU	X	2049	0.02	-0.06	-5.75	0.04	-0.10	-4.78	-0.41	1.45
	Flavobacteriaceae bacterium 3519-10	NC_013062	GAACAUCUCAUUU	X	2534	0.03	-0.06	-5.93	0.06	-0.10	-4.82	-0.57	1.64
	Riemerella anatipestifer DSM 15868	NC_014738	GAACAUCUCAUUU	X	1972	0.02	-0.06	-6.13	0.06	-0.10	-4.53	-0.45	1.55
	Blattabacterium sp. str. BPLAN	NC_013418	GAACAUCUCUUUU	X	578	0.04	-0.03	-6.15	0.09	-0.03	-4.85	-0.91	1.83
	Blattabacterium sp. str. Bge	NC_013454	GAACAUCUCUUUU	X	586	0.03	-0.03	-5.68	0.06	-0.05	-4.76	-0.59	1.69
	Candidatus Sulcia muelleri DMIN	NC_014004	GAACAUUUCUGUU	X	226	0.04	-0.04	-5.60	0.05	-0.10	-5.12	-0.37	1.58
	Candidatus Sulcia muelleri GWSS	NC_010118	GAACAUUUCUGUU	X	227	0.04	-0.03	-5.95	0.07	-0.08	-5.26	-0.54	1.75
	Candidatus Sulcia muelleri SMDSEM	NC_013123	GAACAUCUCUGUU	x	242	0.04	-0.03	-6.16	0.06	-0.08	-5.41	-0.56	1.73
	Candidatus Sulcia muelleri CARI	NC_014499	GAAUAUCUCUGUU	x	246	0.02	-0.05	-5.47	0.04	-0.10	-4.83	-0.61	1.40
Proteobacteria	Escherichia coli str. K-12 substr.	NC_000913	GAUCACCUCCUUA	o	4145	0.61	0.41	-6.84	0.80	0.49	-6.11	-5.14	2.74

/γ-proteobacteria	MG1655												
	Buchnera aphidicola str. 5A	NC_011833	GAUCACCUCCUUA	o	555	0.32	0.29	-5.99	0.55	0.48	-5.04	-3.05	2.64
	Baumannia cicadellinicola str. Hc	NC_007984	GAUCACCUCCUUA	o	595	0.34	0.26	-6.22	0.50	0.36	-5.38	-2.95	2.90
	Candidatus Blochmannia vafer str. BVAF	NC_014909	GAUCACCUCCUUA	o	587	0.27	0.21	-6.40	0.44	0.33	-5.35	-2.58	2.96
	Vibrio cholerae M66-2	NC_012578	GAUCACCUCCUUA	o	3693	0.43	0.24	-6.52	0.57	0.28	-5.82	-3.75	2.93
	Legionella longbeachae NSW150	NC_013861	GAUCACCUCCUUA	o	3403	0.48	0.36	-6.64	0.65	0.45	-5.87	-4.14	2.99
	Candidatus Ruthia magnifica str. Cm	NC_008610	GAUUACCUCCUUA	o	976	0.25	0.14	-6.32	0.37	0.19	-5.50	-2.32	2.87
	Candidatus Carsonella ruddii PV	NC_008512	GAAAAUUUUUAAA	x	182	0.03	0.02	-5.54	0.08	0.04	-4.54	-1.44	1.38
Proteobacteria /β-proteobacteria	Neisseria meningitidis 053442	NC_010120	GAUCACCUCCUUU	o	2020	0.40	0.20	-6.45	0.52	0.22	-5.81	-3.43	2.99
	Thiobacillus denitrificans ATCC 25259	NC_007404	GAUCACCUCCUUU	o	2827	0.46	0.17	-7.02	0.61	0.20	-6.22	-4.20	3.17
	Burkholderia cenocepacia AU 1054	NC_008061	GAUCACCUCCUUU	o	6477	0.47	0.19	-6.95	0.64	0.24	-6.12	-4.30	3.08
	Herbaspirillum seropedicae SmR1	NC_014323	GAUCACCUCCUUU	o	4735	0.47	0.22	-6.75	0.63	0.26	-5.99	-4.15	2.97
	Candidatus Zinderia insecticola CARI	NC_014497	GAUUACAUUUUAA	x	202	0.01	-0.03	-5.42	0.02	-0.06	-4.35	-0.80	1.07
Proteobacteria /α-proteobacteria	Rhodobacter sphaeroides 2.4.1	NC_007493	GAUCACCUCCUUU	o	3857	0.66	0.36	-7.53	0.81	0.38	-6.86	-5.81	3.09
	Caulobacter crescentus CB15	NC_002696	GAUCACCUCCUUU	o	3737	0.46	0.17	-7.04	0.61	0.19	-6.27	-4.28	3.13
	Rickettsia africae ESF-5	NC_012633	GAUUACCUCCUUA	o	1030	0.19	0.09	-6.26	0.28	0.12	-5.40	-1.78	2.68
	Wolbachia sp. wRi	NC_012416	GAUUACCUCCUUA	o	1150	0.32	0.19	-6.77	0.45	0.24	-5.91	-2.97	3.18
	Candidatus Hodgkinia cicadicola Dsem	NC_012960	AAACUUUUGAAAU	x	169	0.04	0.01	-5.07	0.05	-0.01	-4.77	-0.74	1.60
Tenericutes /Mollicutes	Mycoplasma conjunctivae HRC/581	NC_012806	GAACACCUCCUUU	o	692	0.43	0.37	-6.84	0.57	0.45	-6.07	-3.53	3.47
	Mycoplasma hyorhinis HUB-1	NC_014448	GAACACCUCCUUU	o	654	0.43	0.38	-6.46	0.54	0.45	-5.90	-3.23	3.36
	Mycoplasma pulmonis UAB CTIP	NC_002771	GAUCACCUCCUUU	o	782	0.44	0.38	-6.70	0.54	0.42	-6.13	-3.29	3.52
	Mycoplasma agalactiae	NC_013948	GAUUACCUCCUUU	o	813	0.59	0.53	-6.93	0.70	0.58	-6.43	-4.70	3.25
	Mycoplasma fermentans JER	NC_014552	GAUCACCUCCUUU	o	797	0.56	0.50	-7.51	0.63	0.52	-7.09	-4.44	4.06
	Mycoplasma mobile 163K	NC_006908	GAUCACCUCCUUU	o	633	0.51	0.45	-6.71	0.60	0.49	-6.26	-3.66	3.59
	Mycoplasma leachii PG50	NC_014751	GAUCACCUCCUUU	o	882	0.64	0.59	-7.09	0.74	0.63	-6.64	-5.06	3.34
	Mycoplasma penetrans HF-2	NC_004432	GAUCACCUCCUUU	o	1037	0.41	0.37	-6.55	0.55	0.45	-5.84	-3.21	3.39
	Mycoplasma gallisepticum str. R(low)	NC_004829	GAUUACCUCCUUU	O	763	0.23	0.15	-6.28	0.30	0.15	-5.71	-2.08	2.78
	Mycoplasma pneumoniae M129	NC_000912	GAUCACCUCCUUU	O	689	0.17	0.03	-8.06	0.24	0.01	-6.76	-1.80	3.47

	Mycoplasma genitalium G37	NC_000908	GAUCACCUCCUUU	O	475	0.10	-0.01	-7.35	0.14	-0.04	-6.17	-1.12	2.60
	Mycoplasma haemofelis str. Langford 1	NC_014970	GAUAAUCUUCAAG	X	1545	0.07	-0.02	-5.68	0.15	-0.01	-4.71	-1.70	1.73
	Mycoplasma suis KI3806	NC_015153	GAUAACUUUUUAU	X	794	0.07	0.02	-5.62	0.13	0.03	-4.81	-1.57	1.73
	Mycoplasma suis str. Illinois	NC_015155	GAUAACUUUUUAU	X	844	0.07	0.02	-5.45	0.13	0.02	-4.78	-1.59	1.69

This table is adapted from Lim et al. (2012).

Table S2. Genome information and SD interaction usage of plastids

b: Protein-coding genes accepted by our filtering strategy described in Materials and Methods.

c: Protein-coding genes with an SD sequence.

Name (host)	Accession	Phylum	Class	Anti-SD region[a]	Classical anti-SD motif	Genome size	Protein-coding gene #		F_{SD} (= b/c)	Mean MFE of SD interactions
							Total[b]	w/ SD seq[c]		
Pellia endiviifolia	NC_019628.1	Streptophyta	Jungermanniopsida	CCUCCUUUU	Classical	120546	88	45	0.51	-5.84
Cheilanthes lindheimeri	NC_014592.1	Streptophyta	Polypodiopsida	CCUCCUUUU	Classical	155770	86	42	0.49	-6.20
Ptilidium pulcherrimum	NC_015402.1	Streptophyta	Jungermanniopsida	CCUCCUUUU	Classical	119007	80	39	0.49	-6.25
Aneura mirabilis	NC_010359.1	Streptophyta	Jungermanniopsida	CCUCCUUUU	Classical	108007	62	30	0.48	-6.47
Erodium carvifolium	NC_015083.1	Streptophyta	no_rank	CCUCCUUUU	Classical	116935	75	36	0.48	-6.47
Podocarpus totara	NC_020361.1	Streptophyta	Coniferopsida	CCUCCUGCU	Classical	133259	75	36	0.48	-6.28
Mankyua chejuensis	NC_017006.1	Streptophyta	Ophioglossopsida	CCUCCUUUU	Classical	146221	82	39	0.48	-6.15
Pisum sativum	NC_014057.1	Streptophyta	no_rank	CCUCCUUUU	Classical	122169	74	35	0.47	-6.42
Lathyrus sativus	NC_014063.1	Streptophyta	no_rank	CCUCCUGUU	Classical	121020	74	35	0.47	-6.69
Selaginella moellendorffii	NC_013086.1	Streptophyta	Isoetopsida	CCUCCUUCC	Classical	143780	70	33	0.47	-6.93
Pteridium aquilinum subsp. aquilinum	NC_014348.1	Streptophyta	Polypodiopsida	CCUCCUUUU	Classical	152362	87	41	0.47	-6.40
Marchantia polymorpha	NC_001319.1	Streptophyta	Marchantiopsida	CCUCCUUUU	Classical	121024	85	40	0.47	-6.20
Cuscuta gronovii	NC_009765.1	Streptophyta	no_rank	CCUCCUUUU	Classical	86744	62	29	0.47	-6.73
Erodium guttatum	NC_018762.1	Streptophyta	no_rank	CCUCCUUUU	Classical	128510	77	36	0.47	-6.59
Ophioglossum californicum	NC_020147.1	Streptophyta	Ophioglossopsida	CCUCCUUUU	Classical	138270	84	39	0.46	-5.89

Syntrichia ruralis	NC_012052.1	Streptophyta	Bryopsida	CCUCCUUUU	Classical	122630	80	37	0.46	-6.06
Erodium texanum	NC_014569.1	Streptophyta	no_rank	CCUCCUUUU	Classical	130812	78	36	0.46	-6.59
Cicer arietinum	NC_011163.1	Streptophyta	no_rank	CCUCCUUUU	Classical	125319	74	34	0.46	-6.39
Adiantum capillus-veneris	NC_004766.1	Streptophyta	Polypodiopsida	CCUCCUUUU	Classical	150568	83	38	0.46	-6.62
Isoetes flaccida	NC_014675.1	Streptophyta	Isoetopsida	CCUCCUUUC	Classical	145303	83	38	0.46	-6.53
Trachelium caeruleum	NC_010442.1	Streptophyta	no_rank	CCUCCUUUU	Classical	162321	81	37	0.46	-6.89
Alsophila spinulosa	NC_012818.1	Streptophyta	Polypodiopsida	CCUCCUUUU	Classical	156661	88	40	0.45	-6.50
Marsilea crenata	NC_022137.1	Streptophyta	Polypodiopsida	CCUCCUUUU	Classical	151628	88	40	0.45	-6.74
Picea morrisonicola	NC_016069.1	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	124168	71	32	0.45	-6.68
Trifolium subterraneum	NC_011828.1	Streptophyta	no_rank	CCUCCUUUU	Classical	144763	69	31	0.45	-6.25
Cymbidium aloifolium	NC_021429.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	156904	78	35	0.45	-6.88
Cymbidium sinense	NC_021430.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	155548	78	35	0.45	-6.88
Cymbidium tortisepalum	NC_021431.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	155627	78	35	0.45	-6.88
Cymbidium tracyanum	NC_021432.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	156286	78	35	0.45	-6.86
Cymbidium mannii	NC_021433.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	155308	78	35	0.45	-6.88
Citrus sinensis	NC_008334.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160129	87	39	0.45	-6.55
Lygodium japonicum	NC_022136.1	Streptophyta	Polypodiopsida	CCUCCUUUU	Classical	157260	87	39	0.45	-6.23
Pinus nelsonii	NC_011159.4	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	116834	67	30	0.45	-7.26
Monsonia speciosa	NC_014582.1	Streptophyta	no_rank	CCUCCAUUU	Classical	128787	76	34	0.45	-5.34
Medicago truncatula	NC_022099.1	Streptophyta	no_rank	CCUCCUUUU	Classical	124033	76	34	0.45	-6.36
Medicago truncatula	NC_022100.1	Streptophyta	no_rank	CCUCCUUUU	Classical	123833	76	34	0.45	-6.36
Medicago truncatula	NC_022101.1	Streptophyta	no_rank	CCUCCUUUU	Classical	123706	76	34	0.45	-6.36
Ipomoea purpurea	NC_009808.1	Streptophyta	no_rank	CCUCCUUUU	Classical	162046	85	38	0.45	-6.54
Fragaria mandshurica	NC_018767.1	Streptophyta	no_rank	CCUCCUUUU	Classical	129805	74	33	0.45	-6.42
Rhynchoryza subulata	NC_016718.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	136303	83	37	0.45	-7.01
Bigelowiella natans	NC_008408.1	Chlorarachniophyta	Chlorarachniophyceae	CCCUCCUUC	Classical	69166	61	27	0.44	-6.34
Elaeis guineensis	NC_017602.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	156973	86	38	0.44	-6.71
Pinus monophylla	NC_011158.4	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	116479	68	30	0.44	-7.08
Physcomitrella patens subsp. patens	NC_005087.1	Streptophyta	Bryopsida	CCUCCUUUU	Classical	122890	84	37	0.44	-5.96

Carica papaya	NC_010323.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160100	84	37	0.44	-6.55
Cedrus deodara	NC_014575.1	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	119299	75	33	0.44	-6.80
Erycina pusilla	NC_018114.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	143164	73	32	0.44	-6.68
Pinus massoniana	NC_021439.1	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	119739	73	32	0.44	-7.09
Glycine tomentella	NC_021636.1	Streptophyta	no_rank	CCUCCUUUU	Classical	152728	80	35	0.44	-6.45
Glycine canescens	NC_021647.1	Streptophyta	no_rank	CCUCCUUUU	Classical	152518	80	35	0.44	-6.43
Glycine dolichocarpa	NC_021648.1	Streptophyta	no_rank	CCUCCUUUU	Classical	152804	80	35	0.44	-6.43
Glycine falcata	NC_021649.1	Streptophyta	no_rank	CCUCCUUUU	Classical	153023	80	35	0.44	-6.45
Glycine syndetika	NC_021650.1	Streptophyta	no_rank	CCUCCUUUU	Classical	152783	80	35	0.44	-6.43
Datura stramonium	NC_018117.1	Streptophyta	no_rank	CCUCCUUUU	Classical	155871	87	38	0.44	-6.77
Salvia miltiorrhiza	NC_020431.1	Streptophyta	no_rank	CCUCCUUUU	Classical	151328	87	38	0.44	-6.60
Pinus taeda	NC_021440.1	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	121530	71	31	0.44	-7.21
Parthenium argentatum	NC_013553.1	Streptophyta	no_rank	CCUCCUUUU	Classical	152803	55	24	0.44	-6.95
Eucalyptus globulus subsp. globulus	NC_008115.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160286	85	37	0.44	-6.11
Olimarabidopsis pumila	NC_009267.1	Streptophyta	no_rank	CCUCCUUUU	Classical	154737	85	37	0.44	-6.67
Capsella bursa-pastoris	NC_009270.1	Streptophyta	no_rank	CCUCCUUUU	Classical	154490	85	37	0.44	-6.66
Lepidium virginicum	NC_009273.1	Streptophyta	no_rank	CCUCCUUUU	Classical	154743	85	37	0.44	-6.44
Eucalyptus obliqua	NC_022378.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159527	85	37	0.44	-6.21
Eucalyptus radiata	NC_022379.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159529	85	37	0.44	-6.21
Eucalyptus delegatensis	NC_022380.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159724	85	37	0.44	-6.13
Eucalyptus verrucata	NC_022381.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160109	85	37	0.44	-6.11
Eucalyptus baxteri	NC_022382.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160032	85	37	0.44	-6.11
Eucalyptus diversifolia	NC_022383.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159954	85	37	0.44	-6.11
Eucalyptus sieberi	NC_022384.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159985	85	37	0.44	-6.18
Eucalyptus elata	NC_022385.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159899	85	37	0.44	-6.18
Eucalyptus regnans	NC_022386.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160031	85	37	0.44	-6.18
Eucalyptus umbra	NC_022387.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159576	85	37	0.44	-6.11
Eucalyptus cloeziana	NC_022388.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160015	85	37	0.44	-6.11
Eucalyptus patens	NC_022389.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160187	85	37	0.44	-6.11

Eucalyptus marginata	NC_022390.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160076	85	37	0.44	-6.09
Eucalyptus melliodora	NC_022392.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160386	85	37	0.44	-6.18
Eucalyptus polybractea	NC_022393.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160268	85	37	0.44	-6.18
Eucalyptus cladocalyx	NC_022394.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160213	85	37	0.44	-6.18
Eucalyptus nitens	NC_022395.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160271	85	37	0.44	-6.11
Eucalyptus aromaphloia	NC_022396.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160149	85	37	0.44	-6.11
Eucalyptus saligna	NC_022397.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160015	85	37	0.44	-6.11
Eucalyptus camaldulensis	NC_022398.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160164	85	37	0.44	-6.13
Eucalyptus deglupta	NC_022399.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160177	85	37	0.44	-6.11
Eucalyptus torquata	NC_022401.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160223	85	37	0.44	-6.11
Eucalyptus diversicolor	NC_022402.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160214	85	37	0.44	-6.11
Eucalyptus salmonophloia	NC_022403.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160413	85	37	0.44	-6.15
Eucalyptus microcorys	NC_022404.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160225	85	37	0.44	-6.12
Eucalyptus guilfoylei	NC_022405.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160520	85	37	0.44	-6.19
Eucalyptus erythrocorys	NC_022406.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159742	85	37	0.44	-6.11
Corymbia maculata	NC_022408.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160045	85	37	0.44	-6.10
Corymbia eximia	NC_022409.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160012	85	37	0.44	-6.10
Corymbia tessellaris	NC_022410.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160127	85	37	0.44	-6.10
Angophora floribunda	NC_022411.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160245	85	37	0.44	-6.10
Angophora costata	NC_022412.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160326	85	37	0.44	-6.10
Allosyncarpia ternata	NC_022413.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159593	85	37	0.44	-6.19
Stockwellia quadrifida	NC_022414.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159561	85	37	0.44	-6.17
Pinus gerardiana	NC_011154.4	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	117618	69	30	0.43	-7.10
Eucalyptus grandis	NC_014570.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160137	74	32	0.43	-6.10
Picea abies	NC_021456.1	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	124084	74	32	0.43	-6.72
Larix decidua	NC_016058.1	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	122474	72	31	0.43	-6.86
Typha latifolia	NC_013823.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	161572	86	37	0.43	-6.89
Capsicum annuum	NC_018552.1	Streptophyta	no_rank	CCUCCUUUU	Classical	156781	86	37	0.43	-6.55
Anthoceros formosae	NC_004543.1	Streptophyta	Anthocerotopsida	CCUCCUUUU	Classical	161162	84	36	0.43	-6.52

Ranunculus macranthus	NC_008796.1	Streptophyta	no_rank	CCUCCUUUU	Classical	155129	84	36	0.43	-6.31
Draba nemorosa	NC_009272.1	Streptophyta	no_rank	CCUCCUUUU	Classical	153289	84	36	0.43	-6.50
Lolium perenne	NC_009950.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	135282	84	36	0.43	-7.06
Pinus lambertiana	NC_011156.4	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	117239	70	30	0.43	-7.10
Cathaya argyrophylla	NC_014589.1	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	107122	70	30	0.43	-7.30
Oryza rufipogon	NC_017835.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	134544	77	33	0.43	-6.83
Pachycladon enysii	NC_018565.1	Streptophyta	no_rank	CCUCCUUUU	Classical	154896	84	36	0.43	-6.50
Fragaria vesca subsp. bracteata	NC_018766.1	Streptophyta	no_rank	CCUCCUUUU	Classical	129788	77	33	0.43	-6.30
Bambusa oldhamii	NC_012927.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	139350	82	35	0.43	-7.03
Silene noctiflora	NC_016728.1	Streptophyta	no_rank	CCUCCUUUU	Classical	151639	82	35	0.43	-6.57
Oryza meridionalis	NC_016927.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	134558	75	32	0.43	-6.85
Pinus krempfii	NC_011155.4	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	116989	68	29	0.43	-7.03
Cuscuta obtusiflora	NC_009949.1	Streptophyta	no_rank	CCUCCUUUU	Classical	85286	61	26	0.43	-6.64
Chara vulgaris	NC_008097.1	Streptophyta	Charophyceae	CCUCCUUUU	Classical	184933	94	40	0.43	-5.96
Huperzia lucidula	NC_006861.1	Streptophyta	Lycopodiopsida	CCUCCUUUU	Classical	154373	87	37	0.43	-6.23
Solanum lycopersicum	NC_007898.3	Streptophyta	no_rank	CCUCCUUUU	Classical	155461	87	37	0.43	-6.35
Phalaenopsis equestris	NC_017609.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	148959	73	31	0.42	-6.72
Atropa belladonna	NC_004561.1	Streptophyta	no_rank	CCUCCUUUU	Classical	156687	85	36	0.42	-6.86
Cucumis sativus	NC_007144.1	Streptophyta	no_rank	CCUCCUUUU	Classical	155293	85	36	0.42	-6.58
Platanus occidentalis	NC_008335.1	Streptophyta	no_rank	CCUCCUUUU	Classical	161791	85	36	0.42	-6.76
Nandina domestica	NC_008336.1	Streptophyta	no_rank	CCUCCUUUU	Classical	156599	85	36	0.42	-6.94
Agrostis stolonifera	NC_008591.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	136584	85	36	0.42	-7.05
Barbarea verna	NC_009269.1	Streptophyta	no_rank	CCUCCUUUU	Classical	154532	85	36	0.42	-6.50
Nasturtium officinale	NC_009275.1	Streptophyta	no_rank	CCUCCUUUU	Classical	155105	85	36	0.42	-6.59
Megaleranthis saniculifolia	NC_012615.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159924	85	36	0.42	-6.68
Dendrocalamus latiflorus	NC_013088.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	139394	85	36	0.42	-7.04
Prunus persica	NC_014697.1	Streptophyta	no_rank	CCUCCUUUU	Classical	157790	85	36	0.42	-6.27
Fragaria vesca subsp. vesca	NC_015206.1	Streptophyta	no_rank	CCUCCUUUU	Classical	155691	85	36	0.42	-6.62
Fragaria chiloensis	NC_019601.1	Streptophyta	no_rank	CCUCCUUUU	Classical	155603	85	36	0.42	-6.62

Fragaria virginiana	NC_019602.1	Streptophyta	no_rank	CCUCCUUUU	Classical	155621	85	36	0.42	-6.62
Heliconia collinsiana	NC_020362.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	161907	85	36	0.42	-6.64
Pachycladon cheesemanii	NC_021102.1	Streptophyta	no_rank	CCUCCUUUU	Classical	154498	85	36	0.42	-6.69
Eucalyptus curtisii	NC_022391.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160038	85	36	0.42	-6.19
Eucalyptus spathulata	NC_022400.1	Streptophyta	no_rank	CCUCCUUUU	Classical	161071	85	36	0.42	-6.12
Corymbia gummifera	NC_022407.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160713	85	36	0.42	-6.14
Cocos nucifera	NC_022417.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	154731	85	36	0.42	-6.62
Hordeum vulgare subsp. vulgare	NC_008590.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	136462	83	35	0.42	-7.09
Illicium oligandrum	NC_009600.1	Streptophyta	no_rank	CCUCCUUUU	Classical	148553	83	35	0.42	-6.92
Pyrus pyrifolia	NC_015996.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159922	83	35	0.42	-6.31
Leersia tisserantii	NC_016677.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	136551	83	35	0.42	-7.13
Arundinaria gigantea	NC_020341.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	138935	83	35	0.42	-7.07
Pinus contorta	NC_011153.4	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	120438	69	29	0.42	-7.12
Zygnema circumcarinatum	NC_008117.1	Streptophyta	Zygnemophyceae	CCUCCUUCU	Classical	165372	93	39	0.42	-6.28
Oncidium hybrid cultivar	NC_014056.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	146484	74	31	0.42	-6.83
Calycanthus floridus var. glaucus	NC_004993.1	Streptophyta	no_rank	CCUCCUUUU	Classical	153337	86	36	0.42	-6.38
Festuca altissima	NC_019648.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	135272	86	36	0.42	-7.06
Festuca ovina	NC_019649.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	133165	86	36	0.42	-7.07
Festuca pratensis	NC_019650.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	135291	86	36	0.42	-7.06
Lolium multiflorum	NC_019651.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	135175	86	36	0.42	-7.06
Pseudophoenix vinifera	NC_020364.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	157829	86	36	0.42	-6.63
Calamus caryotoides	NC_020365.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	157270	86	36	0.42	-6.72
Bismarckia nobilis	NC_020366.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	158210	86	36	0.42	-6.65
Solanum tuberosum	NC_008096.2	Streptophyta	no_rank	CCUCCUUUU	Classical	155296	84	35	0.42	-6.63
Aethionema cordifolium	NC_009265.1	Streptophyta	no_rank	CCUCCUUUU	Classical	154168	84	35	0.42	-6.53
Aethionema grandiflorum	NC_009266.1	Streptophyta	no_rank	CCUCCUUUU	Classical	154243	84	35	0.42	-6.55
Lobularia maritima	NC_009274.1	Streptophyta	no_rank	CCUCCUUUU	Classical	152659	84	35	0.42	-6.52
Acorus americanus	NC_010093.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	153819	84	35	0.42	-6.81
Pentactina rupicola	NC_016921.1	Streptophyta	no_rank	CCUCCUUUU	Classical	156612	84	35	0.42	-6.26

Jacobaea vulgaris	NC_015543.1	Streptophyta	no_rank	CCUCCUUUU	Classical	150689	87	36	0.41	-6.74
Eleutherococcus senticosus	NC_016430.1	Streptophyta	no_rank	CCUCCUUUU	Classical	156768	87	36	0.41	-6.84
Sesamum indicum	NC_016433.2	Streptophyta	no_rank	CCUCCUUUU	Classical	153324	87	36	0.41	-6.58
Brassica napus	NC_016734.1	Streptophyta	no_rank	CCUCCUUUU	Classical	152860	87	36	0.41	-6.48
Camellia sinensis	NC_020019.1	Streptophyta	no_rank	CCUCCUUUU	Classical	157103	87	36	0.41	-6.68
Ardisia polysticta	NC_021121.1	Streptophyta	no_rank	CCUCCUUUU	Classical	156506	87	36	0.41	-6.99
Catharanthus roseus	NC_021423.1	Streptophyta	no_rank	CCUCCUUUU	Classical	154950	87	36	0.41	-6.73
Keteleeria davidiana	NC_011930.1	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	117720	75	31	0.41	-7.25
Tetracentron sinense	NC_021425.1	Streptophyta	no_rank	CCUCCUUUU	Classical	164467	92	38	0.41	-6.61
Picea sitchensis	NC_011152.3	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	120176	63	26	0.41	-7.08
Lotus japonicus	NC_002694.1	Streptophyta	no_rank	CCUCCUUUU	Classical	150519	80	33	0.41	-6.45
Festuca arundinacea	NC_011713.2	Streptophyta	Liliopsida	CCUCCUUUU	Classical	136048	80	33	0.41	-7.19
Glycine stenophita	NC_021646.1	Streptophyta	no_rank	CCUCCUUUU	Classical	152618	80	33	0.41	-6.57
Arabidopsis thaliana	NC_000932.1	Streptophyta	no_rank	CCUCCUUUU	Classical	154478	85	35	0.41	-6.61
Solanum bulbocastanum	NC_007943.1	Streptophyta	no_rank	CCUCCUUUU	Classical	155371	85	35	0.41	-6.47
Daucus carota	NC_008325.1	Streptophyta	no_rank	CCUCCUUUU	Classical	155911	85	35	0.41	-6.73
Crucihimalaya wallichii	NC_009271.1	Streptophyta	no_rank	CCUCCUUUU	Classical	155199	85	35	0.41	-6.53
Guizotia abyssinica	NC_010601.1	Streptophyta	no_rank	CCUCCUUUU	Classical	151762	85	35	0.41	-6.86
Olea europaea	NC_013707.2	Streptophyta	no_rank	CCUCCUUUU	Classical	155888	85	35	0.41	-7.04
Anthriscus cerefolium	NC_015113.1	Streptophyta	no_rank	CCUCCUUUU	Classical	154719	85	35	0.41	-6.75
Nelumbo lutea	NC_015605.1	Streptophyta	no_rank	CCUCCUUUU	Classical	163206	85	35	0.41	-6.66
Nelumbo nucifera	NC_015610.1	Streptophyta	no_rank	CCUCCUUUU	Classical	163307	85	35	0.41	-6.66
Zingiber spectabile	NC_020363.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	155890	85	35	0.41	-6.92
Asclepias nivea	NC_022431.1	Streptophyta	no_rank	CCUCCUUUU	Classical	161592	85	35	0.41	-6.72
Pseudotsuga sinensis var. wilsoniana	NC_016064.1	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	122513	73	30	0.41	-6.99
Phoenix dactylifera	NC_013991.2	Streptophyta	Liliopsida	CCUCCUUUU	Classical	158462	95	39	0.41	-6.84
Triticum aestivum	NC_002762.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	134545	83	34	0.41	-7.11
Phalaenopsis aphrodite subsp. formosana	NC_007499.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	148964	83	34	0.41	-6.66
Glycine max	NC_007942.1	Streptophyta	no_rank	CCUCCUUUU	Classical	152218	83	34	0.41	-6.56

Populus alba	NC_008235.1	Streptophyta	no_rank	CCUCCUUUU	Classical	156505	83	34	0.41	-6.65
Corynocarpus laevigata	NC_014807.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159202	83	34	0.41	-6.46
Phyllostachys propinqua	NC_016699.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	139704	83	34	0.41	-7.17
Chrysanthemum indicum	NC_020320.1	Streptophyta	no_rank	CCUCCUUUU	Classical	150972	83	34	0.41	-6.91
Asclepias syriaca	NC_022432.1	Streptophyta	no_rank	CCUCCUUUU	Classical	158719	83	34	0.41	-6.73
Cucumis melo subsp. melo	NC_015983.1	Streptophyta	no_rank	CCUCCUUUU	Classical	156017	88	36	0.41	-6.60
Medicago truncatula	NC_003119.6	Streptophyta	no_rank	CCUCCUUUU	Classical	124033	76	31	0.41	-6.45
Millettia pinnata	NC_016708.2	Streptophyta	no_rank	CCUCCUUUU	Classical	152968	81	33	0.41	-6.57
Silene latifolia	NC_016730.1	Streptophyta	no_rank	CCUCCUUUU	Classical	151736	81	33	0.41	-6.75
Chloranthus spicatus	NC_009598.1	Streptophyta	no_rank	CCUCCUUUU	Classical	157772	86	35	0.41	-6.81
Buxus microphylla	NC_009599.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159010	86	35	0.41	-6.97
Ricinus communis	NC_016736.1	Streptophyta	no_rank	CCUCCUUUU	Classical	163161	86	35	0.41	-6.62
Trochodendron aralioides	NC_021426.1	Streptophyta	no_rank	CCUCCUUUU	Classical	165945	91	37	0.41	-6.59
Glycine cyrtoloba	NC_021645.1	Streptophyta	no_rank	CCUCCUUUU	Classical	152381	79	32	0.41	-6.52
Triticum monococcum	NC_021760.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	116399	79	32	0.41	-6.82
Acorus calamus	NC_007407.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	153821	84	34	0.40	-6.66
Morus indica	NC_008359.1	Streptophyta	no_rank	CCUCCUUUU	Classical	158484	84	34	0.40	-6.43
Gossypium barbadense	NC_008641.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160317	84	34	0.40	-6.43
Arabis hirsuta	NC_009268.1	Streptophyta	no_rank	CCUCCUUUU	Classical	153689	84	34	0.40	-6.61
Bambusa emeiensis	NC_015830.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	139493	84	34	0.40	-7.06
Prinsepia utilis	NC_021455.1	Streptophyta	no_rank	CCUCCUUUU	Classical	156328	84	34	0.40	-6.49
Staurastrum punctulatum	NC_008116.1	Streptophyta	Zygnemophyceae	CCUCCUUCC	Classical	157089	94	38	0.40	-6.67
Trithuria inconspicua	NC_020372.1	Streptophyta	no_rank	CCUCCUUUU	Classical	165389	94	38	0.40	-6.63
Geranium palmatum	NC_014573.1	Streptophyta	no_rank	CCUCCUUUU	Classical	155794	77	31	0.40	-6.08
Acidosasa purpurea	NC_015820.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	139697	82	33	0.40	-7.07
Dasypogon bromeliifolius	NC_020367.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	157858	87	35	0.40	-6.49
Utricularia gibba	NC_021449.1	Streptophyta	no_rank	CCUCCUUUU	Classical	152113	87	35	0.40	-6.85
Nymphaea alba	NC_006050.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159930	85	34	0.40	-6.94
Panax ginseng	NC_006290.1	Streptophyta	no_rank	CCUCCUUUU	Classical	156318	85	34	0.40	-6.91

Helianthus annuus	NC_007977.1	Streptophyta	no_rank	CCUCCUUUU	Classical	151104	85	34	0.40	-6.89
Silene conica	NC_016729.1	Streptophyta	no_rank	CCUCCUUUU	Classical	147208	80	32	0.40	-7.19
Vaccinium macrocarpon	NC_019616.1	Streptophyta	no_rank	CCUCCUUUU	Classical	176045	75	30	0.40	-6.36
Triticum urartu	NC_021762.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	115773	60	24	0.40	-6.80
Aegilops tauschii	NC_022133.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	114112	80	32	0.40	-7.01
Nicotiana tabacum	NC_001879.2	Streptophyta	no_rank	CCUCCUUUU	Classical	155943	98	39	0.40	-6.67
Oenothera elata subsp. hookeri	NC_002693.2	Streptophyta	no_rank	CCUCCUUUU	Classical	165728	83	33	0.40	-6.83
Gossypium hirsutum	NC_007944.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160301	83	33	0.40	-6.29
Spirodela polyrhiza	NC_015891.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	168788	83	33	0.40	-7.04
Wolffiella lingulata	NC_015894.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	169337	83	33	0.40	-6.89
Gossypium raimondii	NC_016668.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160161	83	33	0.40	-6.29
Gossypium darwinii	NC_016670.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160378	83	33	0.40	-6.29
Gossypium tomentosum	NC_016690.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160433	83	33	0.40	-6.29
Gossypium herbaceum subsp. africanum	NC_016692.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160315	83	33	0.40	-6.29
Gossypium mustelinum	NC_016711.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160313	83	33	0.40	-6.29
Gossypium arboreum	NC_016712.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160230	83	33	0.40	-6.29
Gossypium gossypioides	NC_017894.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159959	83	33	0.40	-6.29
Gossypium incanum	NC_018109.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159205	83	33	0.40	-6.29
Gossypium somalense	NC_018110.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159539	83	33	0.40	-6.29
Gossypium capitis-viridis	NC_018111.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159467	83	33	0.40	-6.32
Gossypium areysianum	NC_018112.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159572	83	33	0.40	-6.29
Gossypium robinsonii	NC_018113.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159849	83	33	0.40	-6.29
Nicotiana sylvestris	NC_007500.1	Streptophyta	no_rank	CCUCCUUUU	Classical	155941	101	40	0.40	-6.68
Populus trichocarpa	NC_009143.1	Streptophyta	no_rank	CCUCCUUUU	Classical	157033	96	38	0.40	-6.64
Brachypodium distachyon	NC_011032.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	135199	81	32	0.40	-6.85
Cyanophora paradoxa	NC_001675.1	Glaucophyta	Glaucocystophyceae	CCUCCUUUA	Classical	135599	142	56	0.39	-5.37
Chlorokybus atmophyticus	NC_008822.1	Streptophyta	Chlorokybophyceae	CCUCCUUUA	Classical	152254	104	41	0.39	-5.84
Camellia taliensis	NC_022264.1	Streptophyta	no_rank	CCUCCUUUU	Classical	156974	89	35	0.39	-6.65
Camellia impressinervis	NC_022461.1	Streptophyta	no_rank	CCUCCUUUU	Classical	156892	89	35	0.39	-6.64

Camellia pitardii	NC_022462.1	Streptophyta	no_rank	CCUCCUUUU	Classical	156585	89	35	0.39	-6.64
Camellia yunnanensis	NC_022463.1	Streptophyta	no_rank	CCUCCUUUU	Classical	156592	89	35	0.39	-6.64
Amborella trichopoda	NC_005086.1	Streptophyta	no_rank	CCUCCUUUU	Classical	162686	84	33	0.39	-7.16
Vitis vinifera	NC_007957.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160928	84	33	0.39	-6.29
Sorghum bicolor	NC_008602.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	140754	84	33	0.39	-7.18
Oenothera argillicola	NC_010358.1	Streptophyta	no_rank	CCUCCUUUU	Classical	165055	84	33	0.39	-6.82
Oenothera glazioviana	NC_010360.1	Streptophyta	no_rank	CCUCCUUUU	Classical	165225	84	33	0.39	-6.82
Oenothera biennis	NC_010361.1	Streptophyta	no_rank	CCUCCUUUU	Classical	164807	84	33	0.39	-6.83
Oenothera parviflora	NC_010362.1	Streptophyta	no_rank	CCUCCUUUU	Classical	163365	84	33	0.39	-6.82
Hevea brasiliensis	NC_015308.1	Streptophyta	no_rank	CCUCCUUUU	Classical	161191	84	33	0.39	-6.61
Phyllostachys edulis	NC_015817.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	139679	84	33	0.39	-7.07
Ferrocalamus rimosivaginus	NC_015831.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	139467	84	33	0.39	-7.10
Magnolia kwangsiensis	NC_015892.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159667	84	33	0.39	-6.96
Magnolia denudata	NC_018357.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160053	84	33	0.39	-6.93
Magnolia officinalis subsp. biloba	NC_020317.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160105	84	33	0.39	-6.96
Magnolia grandiflora	NC_020318.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159623	84	33	0.39	-6.77
Cuscuta reflexa	NC_009766.1	Streptophyta	no_rank	CCUCCUUUU	Classical	121521	69	27	0.39	-7.09
Andrographis paniculata	NC_022451.1	Streptophyta	no_rank	CCUCCUUUU	Classical	150249	87	34	0.39	-6.89
Cryptomeria japonica	NC_010548.1	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	131810	82	32	0.39	-6.42
Indocalamus longiauritus	NC_015803.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	139668	82	32	0.39	-7.17
Magnolia officinalis	NC_020316.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160183	82	32	0.39	-7.03
Cephalotaxus oliveri	NC_021110.1	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	134337	82	32	0.39	-6.15
Jasminum nudiflorum	NC_008407.1	Streptophyta	no_rank	CCUCCUUUU	Classical	165121	85	33	0.39	-7.03
Piper cenocladum	NC_008457.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160624	85	33	0.39	-6.67
Coffea arabica	NC_008535.1	Streptophyta	no_rank	CCUCCUUUU	Classical	155189	85	33	0.39	-6.88
Nuphar advena	NC_008788.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160866	85	33	0.39	-6.87
Ceratophyllum demersum	NC_009962.1	Streptophyta	no_rank	CCUCCUUUU	Classical	156252	85	33	0.39	-6.95
Lemna minor	NC_010109.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	165955	85	33	0.39	-6.82
Gossypium thurberi	NC_015204.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160264	85	33	0.39	-6.29

Olea europaea subsp. europaea	NC_015401.1	Streptophyta	no_rank	CCUCCUUUU	Classical	155875	85	33	0.39	-6.76
Olea europaea subsp. cuspidata	NC_015604.1	Streptophyta	no_rank	CCUCCUUUU	Classical	155862	85	33	0.39	-6.76
Olea woodiana subsp. woodiana	NC_015608.1	Streptophyta	no_rank	CCUCCUUUU	Classical	155942	85	33	0.39	-6.68
Olea europaea subsp. maroccana	NC_015623.1	Streptophyta	no_rank	CCUCCUUUU	Classical	155896	85	33	0.39	-6.76
Panicum virgatum	NC_015990.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	139619	85	33	0.39	-7.05
Boea hygrometrica	NC_016468.1	Streptophyta	no_rank	CCUCCUUUU	Classical	153493	85	33	0.39	-6.90
Chrysanthemum x morifolium	NC_020092.1	Streptophyta	no_rank	CCUCCUUUU	Classical	151033	85	33	0.39	-6.95
Cuscuta exaltata	NC_009963.1	Streptophyta	no_rank	CCUCCUUUU	Classical	125373	67	26	0.39	-7.57
Trebouxiophyceae sp. MX-AZ01	NC_018569.1	Chlorophyta	Trebouxiophyceae	CCUCCUUUA	Classical	149707	80	31	0.39	-6.45
Neottia nidus-avis	NC_016471.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	92060	31	12	0.39	-7.44
Cistanche deserticola	NC_021111.1	Streptophyta	no_rank	CCUCCUGUU	Classical	102657	31	12	0.39	-6.97
Nicotiana tomentosiformis	NC_007602.1	Streptophyta	no_rank	CCUCCUUUU	Classical	155745	101	39	0.39	-6.71
Welwitschia mirabilis	NC_010654.1	Streptophyta	Gnetopsida	CCUCCUUUU	Classical	119726	70	27	0.39	-6.43
Wolffia australiana	NC_015899.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	168704	83	32	0.39	-6.68
Taiwania cryptomerioides	NC_016065.1	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	132588	83	32	0.39	-6.02
Taiwania flousiana	NC_021441.1	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	131413	83	32	0.39	-6.02
Colocasia esculenta	NC_016753.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	162424	86	33	0.38	-6.50
Tectona grandis	NC_020098.1	Streptophyta	no_rank	CCUCCUUUU	Classical	153953	86	33	0.38	-6.40
Silene vulgaris	NC_016727.1	Streptophyta	no_rank	CCUCCUUUU	Classical	151583	81	31	0.38	-6.74
Francoa sonchifolia	NC_021101.1	Streptophyta	no_rank	CCUCCUUUU	Classical	157312	81	31	0.38	-6.13
Camellia cuspidata	NC_022459.1	Streptophyta	no_rank	CCUCCUUUU	Classical	156618	89	34	0.38	-6.73
Lactuca sativa	NC_007578.1	Streptophyta	no_rank	CCUCCUUUU	Classical	152765	84	32	0.38	-6.87
Liriodendron tulipifera	NC_008326.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159886	84	32	0.38	-6.77
Dioscorea elephantipes	NC_009601.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	152609	84	32	0.38	-6.51
Fagopyrum esculentum subsp. ancestrale	NC_010776.1	Streptophyta	no_rank	CCUCCUUUU	Classical	159599	84	32	0.38	-6.49
Phyllostachys nigra var. henonis	NC_015826.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	139839	84	32	0.38	-7.17
Ginkgo biloba	NC_016986.1	Streptophyta	Ginkgoopsida	CCUCCUUUU	Classical	156988	84	32	0.38	-6.53
Elodea canadensis	NC_018541.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	156700	84	32	0.38	-6.94
Oryza sativa Japonica Group	NC_001320.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	134525	108	41	0.38	-6.92

Artemisia frigida	NC_020607.1	Streptophyta	no_rank	CCUCCUUUU	Classical	151076	87	33	0.38	-7.00
Camellia danzaiensis	NC_022460.1	Streptophyta	no_rank	CCUCCUUUU	Classical	156576	87	33	0.38	-6.44
Cephalotaxus wilsoniana	NC_016063.1	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	136196	82	31	0.38	-6.05
Mesostigma viride	NC_002186.1	Streptophyta	Mesostigmatophyceae	CCUCCUUUC	Classical	118360	98	37	0.38	-5.75
Nephroselmis olivacea	NC_000927.1	Chlorophyta	Nephroselmidophyceae	CCUCCUUUG	Classical	200799	151	57	0.38	-6.63
Oryza sativa Indica Group	NC_008155.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	134496	64	24	0.38	-7.07
Berberis bealei	NC_022457.1	Streptophyta	no_rank	CCUCCUUUU	Classical	164792	104	39	0.38	-6.45
Phaseolus vulgaris	NC_009259.1	Streptophyta	no_rank	CCUCCUUUU	Classical	150285	83	31	0.37	-7.07
Manihot esculenta	NC_010433.1	Streptophyta	no_rank	CCUCCUUUU	Classical	161453	83	31	0.37	-6.65
Castanea mollissima	NC_014674.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160799	83	31	0.37	-6.86
Veratrum patulum	NC_022715.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	153699	83	31	0.37	-6.33
Saccharum hybrid cultivar SP80-3280	NC_005878.2	Streptophyta	Liliopsida	CCUCCUUUU	Classical	141182	97	36	0.37	-7.14
Theobroma cacao	NC_014676.2	Streptophyta	no_rank	CCUCCUUUU	Classical	160619	81	30	0.37	-6.23
Vigna angularis	NC_021091.1	Streptophyta	no_rank	CCUCCUUUU	Classical	151683	81	30	0.37	-6.95
Cunninghamia lanceolata	NC_021437.1	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	135334	81	30	0.37	-6.06
Gnetum montanum	NC_021438.1	Streptophyta	Gnetopsida	CCUCCUUUU	Classical	115019	65	24	0.37	-7.18
Jatropha curcas	NC_012224.1	Streptophyta	no_rank	CCUCCUUUU	Classical	163856	84	31	0.37	-6.50
Pharus latifolius	NC_021372.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	142077	76	28	0.37	-7.00
Oltmannsiellopsis viridis	NC_008099.1	Chlorophyta	Ulvophyceae	CCUCCUUUA	Classical	151933	90	33	0.37	-5.38
Vigna radiata	NC_013843.1	Streptophyta	no_rank	CCUCCUUUU	Classical	151271	82	30	0.37	-6.96
Najas flexilis	NC_021936.1	Streptophyta	Liliopsida	ACCUCCUUU	Classical	156366	74	27	0.36	-7.31
Anomochloa marantoidea	NC_014062.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	138412	85	31	0.36	-6.82
Nicotiana undulata	NC_016068.1	Streptophyta	no_rank	CCUCCUUUU	Classical	155863	110	40	0.36	-6.53
Secale cereale	NC_021761.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	114843	77	28	0.36	-6.88
Zea mays	NC_001666.2	Streptophyta	Liliopsida	CCUCCUUUU	Classical	140384	111	40	0.36	-7.12
Epifagus virginiana	NC_001568.1	Streptophyta	no_rank	CCUCCUUUU	Classical	70028	25	9	0.36	-8.10
Taxus mairei	NC_020321.1	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	127665	78	28	0.36	-6.16
Pelargonium x hortorum	NC_008454.1	Streptophyta	no_rank	CCUCCUCUU	Classical	217942	131	47	0.36	-7.32
Equisetum arvense	NC_014699.1	Streptophyta	Equisetopsida	CCUCCUUUU	Classical	133309	84	30	0.36	-6.20

Chlorella vulgaris	NC_001865.1	Chlorophyta	Trebouxiophyceae	CCUCCUUUU	Classical	150613	82	29	0.35	-6.16
Angiopteris evecta	NC_008829.1	Streptophyta	Marattiopsida	CCUCCUUUU	Classical	153901	88	31	0.35	-6.26
Spinacia oleracea	NC_002202.1	Streptophyta	no_rank	CCUCCUUUU	Classical	150725	94	33	0.35	-6.90
Oryza rufipogon	NC_022668.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	134557	114	40	0.35	-6.98
Aegilops speltoides	NC_022135.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	113536	77	27	0.35	-7.09
Psilotum nudum	NC_003386.1	Streptophyta	no_rank	CCUCCUUUU	Classical	138829	100	35	0.35	-6.02
Gnetum parvifolium	NC_011942.1	Streptophyta	Gnetopsida	CCUCCUUUU	Classical	114914	66	23	0.35	-7.14
Quercus rubra	NC_020152.1	Streptophyta	no_rank	CCUCCUUUU	Classical	161304	89	31	0.35	-6.97
Nannochloropsis gaditana	NC_020014.1	Heterokontophyta	no_rank	CCUCCUUUA	Classical	114989	115	40	0.35	-5.63
Ephedra equisetina	NC_011954.1	Streptophyta	Gnetopsida	CCUCCUUUU	Classical	109518	72	25	0.35	-6.12
Vigna unguiculata	NC_018051.1	Streptophyta	no_rank	CCUCCUUUU	Classical	152415	84	29	0.35	-7.03
Cycas taitungensis	NC_009618.1	Streptophyta	Cycadopsida	CCUCCUUUU	Classical	163403	113	39	0.35	-5.94
Oryza nivara	NC_005973.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	134494	119	41	0.34	-7.00
Chaetosphaeridium globosum	NC_004115.1	Streptophyta	Coleochaetophyceae	CCUCCUUUU	Classical	131183	96	33	0.34	-6.24
Saccharum hybrid cultivar NCo 310	NC_006084.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	141182	117	40	0.34	-7.15
Nannochloropsis salina	NC_022261.1	Heterokontophyta	no_rank	CCUCCUUUA	Classical	114882	123	42	0.34	-5.75
Drimys granadensis	NC_008456.1	Streptophyta	no_rank	CCUCCUUUU	Classical	160604	85	29	0.34	-6.71
Pyramimonas parkeae	NC_012099.1	Chlorophyta	Prasinophyceae	CCUCCUUUU	Classical	101605	91	31	0.34	-6.49
Coix lacryma-jobi	NC_013273.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	140745	103	35	0.34	-7.16
Cycas revoluta	NC_020319.1	Streptophyta	Cycadopsida	CCUCCUUUU	Classical	162489	109	37	0.34	-5.92
Chlorella variabilis	NC_015359.1	Chlorophyta	Trebouxiophyceae	CCUCCUUUA	Classical	124579	80	27	0.34	-5.96
Equisetum hyemale	NC_020146.1	Streptophyta	Equisetopsida	CCUCCUUUU	Classical	131760	83	28	0.34	-6.24
Micromonas sp. RCC299	NC_012575.1	Chlorophyta	Mamiellophyceae	CCUCCUUUU	Classical	72585	57	19	0.33	-5.27
Saccharina japonica	NC_018523.1	Heterokontophyta	Phaeophyceae	CCUCCUUAA	Classical	130584	139	46	0.33	-5.47
Ostreococcus tauri	NC_008289.1	Chlorophyta	Mamiellophyceae	CCUCCUUUU	Classical	71666	61	20	0.33	-5.67
Ageratina adenophora	NC_015621.1	Streptophyta	no_rank	CCUCCUUCC	Classical	150698	86	28	0.33	-7.02
Nannochloropsis limnetica	NC_022262.1	Heterokontophyta	no_rank	CCUCCUUUA	Classical	117806	124	40	0.32	-5.69
Ectocarpus siliculosus	NC_013498.1	Heterokontophyta	Phaeophyceae	CCUCCUUAA	Classical	139954	147	47	0.32	-5.59
Parachlorella kessleri	NC_012978.1	Chlorophyta	Trebouxiophyceae	CCUCCUUAA	Classical	123994	83	26	0.31	-5.69

Pedinomonas minor	NC_016733.1	Chlorophyta	Trebouxiophyceae	CCUCCUUUU	Classical	98340	80	25	0.31	-6.46
Nannochloropsis granulata	NC_022259.1	Heterokontophyta	no_rank	CCUCCUUUA	Classical	117672	125	39	0.31	-5.66
Rhizanthella gardneri	NC_014874.1	Streptophyta	Liliopsida	CCUCCUUUU	Classical	59190	23	7	0.30	-6.66
Pinus thunbergii	NC_001631.1	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	119707	126	38	0.30	-7.01
Nannochloropsis oculata	NC_022260.1	Heterokontophyta	no_rank	CCUCCUUUA	Classical	117463	126	38	0.30	-5.73
Nannochloropsis oceanica	NC_022263.1	Heterokontophyta	no_rank	CCUCCUUUA	Classical	117557	126	37	0.29	-5.72
Leptosira terrestris	NC_009681.1	Chlorophyta	Trebouxiophyceae	CCUCCUGGA	Classical	195081	77	21	0.27	-5.62
Pseudendoclonium akinetum	NC_008114.1	Chlorophyta	Ulvophyceae	CCUCCUUCA	Classical	195867	97	26	0.27	-6.28
Odontella sinensis	NC_001713.1	Heterokontophyta	Mediophyceae	CCUCCUUUU	Classical	119704	131	35	0.27	-5.87
Coccomyxa subellipsoidea C-169	NC_015084.1	Chlorophyta	Trebouxiophyceae	CCUCCUUAA	Classical	175731	80	21	0.26	-5.30
Rhodomonas salina	NC_009573.1	Cryptophyta	Cryptophyceae	CCUCCUUUU	Classical	135854	144	37	0.26	-5.80
Chondrus crispus	NC_020795.1	Rhodophyta	Florideophyceae	CCUCCUUAA	Classical	180086	204	52	0.25	-5.58
Heterosigma akashiwo	NC_010772.1	Heterokontophyta	Raphidophyceae	CCUCCUUAU	Classical	159370	155	38	0.25	-6.16
Grateloupia taiwanensis	NC_021618.1	Rhodophyta	Florideophyceae	CCUCCUUAA	Classical	191270	202	49	0.24	-5.62
Thalassiosira pseudonana	NC_008589.1	Heterokontophyta	Coscinodiscophyceae	CCUCCUUUU	Classical	128814	141	34	0.24	-5.76
Pyropia yezoensis	NC_007932.1	Rhodophyta	Bangiophyceae	CCUCCUUUU	Classical	191952	206	49	0.24	-5.82
Pavlova lutheri	NC_020371.1	Haptophyta	Prymnesiophyceae	CCUCCUUUA	Classical	95281	102	24	0.24	-5.63
Porphyra purpurea	NC_000925.1	Rhodophyta	Bangiophyceae	CCUCCUUUU	Classical	191028	209	49	0.23	-5.44
Pyropia haitanensis	NC_021189.1	Rhodophyta	Bangiophyceae	CCUCCUUUU	Classical	195597	210	49	0.23	-5.57
Guillardia theta	NC_000926.1	Cryptophyta	Cryptophyceae	CCUCCUUAA	Classical	121524	142	33	0.23	-5.64
Pinus koraiensis	NC_004677.2	Streptophyta	Coniferopsida	CCUCCUUUU	Classical	117190	168	39	0.23	-6.50
Fucus vesiculosus	NC_016735.1	Heterokontophyta	Phaeophyceae	CCUCCUUAA	Classical	124986	139	31	0.22	-5.37
Eutreptiella gymnastica	NC_017754.2	Euglenophyta	Euglenida	CCUCCUAAU	Classical	67623	59	13	0.22	-5.18
Phaeodactylum tricornutum	NC_008588.1	Heterokontophyta	Bacillariophyceae	CCUCCUUUU	Classical	117369	132	29	0.22	-5.86
Durinskia baltica	NC_014287.1	Dinoflagellata	Dinophyceae	CCUCCUUUU	Classical	116470	129	28	0.22	-5.90
Gracilaria tenuistipitata var. liui	NC_006137.1	Rhodophyta	Florideophyceae	CCUCCUUAU	Classical	183883	201	43	0.21	-5.79
Kryptoperidinium foliaceum	NC_014267.1	Dinoflagellata	Dinophyceae	CCUCCUUUU	Classical	140426	134	28	0.21	-5.98
Monomastix sp. OKE-1	NC_012101.1	Chlorophyta	Prasinophyceae	CCUCCUUAA	Classical	114528	73	15	0.21	-5.14
Fistulifera sp. JPCC DA0580	NC_015403.1	Heterokontophyta	Bacillariophyceae	CCUCCUUUU	Classical	134918	135	27	0.20	-5.70

Calliarthron tuberculosum	NC_021075.1	Rhodophyta	Florideophyceae	CCUCCUUUU	Classical	178981	199	39	0.20	-5.73
Vaucheria litorea	NC_011600.1	Heterokontophyta	Xanthophyceae	CCUCCUUAU	Classical	115341	138	27	0.20	-5.93
Synedra acus	NC_016731.1	Heterokontophyta	Fragilariophyceae	CCUCCUUUU	Classical	116251	130	25	0.19	-5.78
Thalassiosira oceanica CCMP1005	NC_014808.1	Heterokontophyta	Coscinodiscophyceae	CCUCCUUUU	Classical	141790	141	25	0.18	-5.67
Schizomeris leibleinii	NC_015645.1	Chlorophyta	Chlorophyceae	CUCCUUCUU	Reduced	182759	75	13	0.17	-7.66
Cyanidium caldarium	NC_001840.1	Rhodophyta	Bangiophyceae	CCUCCUUUA	Classical	164921	184	30	0.16	-5.93
Cyanidioschyzon merolae strain 10D	NC_004799.1	Rhodophyta	Bangiophyceae	CCUCCUUAA	Classical	149987	187	30	0.16	-6.29
Cryptomonas paramecium	NC_013703.1	Cryptophyta	Cryptophyceae	CCUCCUUUU	Classical	77717	80	12	0.15	-7.05
Phaeocystis antarctica	NC_016703.2	Haptophyta	Prymnesiophyceae	CCUCCUUUU	Classical	105512	108	16	0.15	-5.56
Dunaliella salina	NC_016732.1	Chlorophyta	Chlorophyceae	CCUCCUUCU	Classical	269044	82	12	0.15	-5.90
Monomorphina aenigmatica	NC_020018.1	Euglenophyta	Euglenida	ACUCCAUUA	Reduced	74746	62	9	0.15	-5.00
Aureococcus anophagefferens	NC_012898.1	Heterokontophyta	Pelagophyceae	CCUCCUUAA	Classical	89599	105	15	0.14	-6.19
Phaeocystis globosa	NC_021637.1	Haptophyta	Prymnesiophyceae	CCUCCUUUU	Classical	107461	108	15	0.14	-5.55
Pycnococcus provasolii	NC_012097.1	Chlorophyta	Prasinophyceae	CCUCCUAUA	Classical	80211	65	9	0.14	-4.96
Floydiella terrestris	NC_014346.1	Chlorophyta	Chlorophyceae	UUCUCCUAU	Reduced	521168	74	10	0.14	-7.21
Stigeoclonium helveticum	NC_008372.1	Chlorophyta	Chlorophyceae	UCCUUCUUA	Reduced	223902	76	10	0.13	-8.01
Chlamydomonas reinhardtii	NC_005353.1	Chlorophyta	Chlorophyceae	CCUCCUUCA	Classical	203828	69	8	0.12	-6.28
Euglena gracilis	NC_001603.2	Euglenophyta	Euglenida	ACUCCC	Reduced	143171	62	7	0.11	-5.09
Aureoumbra lagunensis	NC_012903.1	Heterokontophyta	Pelagophyceae	CCUCCUUAA	Classical	94346	110	12	0.11	-5.99
Gonium pectorale	NC_020438.1	Chlorophyta	Chlorophyceae	CCUCCUUCA	Classical	222582	68	7	0.10	-6.79
Pleodorina starrii	NC_021109.1	Chlorophyta	Chlorophyceae	CCUCCUUCA	Classical	269857	84	8	0.10	-7.18
Euglena viridis	NC_020460.2	Euglenophyta	Euglenida	ACUCCAAUG	Reduced	91616	64	6	0.09	-4.87
Oedogonium cardiacum	NC_011031.1	Chlorophyta	Chlorophyceae	CCUCCUUCA	Classical	196547	99	9	0.09	-7.29
Bryopsis hypnoides	NC_013359.1	Chlorophyta	Ulvophyceae	CCUCCUUAU	Classical	153429	69	5	0.07	-6.40
Emiliania huxleyi	NC_007288.1	Haptophyta	Prymnesiophyceae	CCUCCUUUU	Classical	105309	112	8	0.07	-5.85
Theileria parva strain Muguga	NC_007758.1	Apicomplexa	Aconoidasida	UAAUUGUAG	Lost	39579	36	1	0.03	-4.70
Chromerida sp. RM11	NC_014345.1	Chromerida	no_rank	CAUGUAUCA	Lost	85535	78	1	0.01	-5.60
Toxoplasma gondii RH	NC_001799.1	Apicomplexa	Coccidia	UUUACUAAA	Lost	34996	21	0	0.00	
Euglena longa	NC_002652.1	Euglenophyta	Euglenida	AUGUUUUAA	Lost	73345	39	0	0.00	

Eimeria tenella strain Penn State	NC_004823.1	Apicomplexa	Coccidia	UAUACUAUC	Lost	34750	20	0	0.00
Helicosporidium sp. ex Simulium jonesi	NC_008100.1	Chlorophyta	Trebouxiophyceae	AUAAGAACU	Lost	37454	25	0	0.00
Acutodesmus obliquus	NC_008101.1	Chlorophyta	Chlorophyceae	AUCUUUUUA	Lost	161452	77	0	0.00
Babesia bovis T2Bo	NC_011395.1	Apicomplexa	Aconoidasida	UAUCAAUAU	Lost	35107	30	0	0.00
Chromera velia	NC_014340.2	Chromerida	no_rank	UUUUAUUUU	Lost	120426	63	0	0.00
Plasmodium falciparum HB3	NC_017928.1	Apicomplexa	Aconoidasida	UAAAAUAAA	Lost	29529	30	0	0.00
Leucocytozoon caulleryi	NC_022667.1	Apicomplexa	Aconoidasida	UAUAAAAUA	Lost	34779	29	0	0.00

This table is adapted from Lim et al. (submitted)

Table S3. Orthologous groups of putative genes in plastids

Orthologous group	Product	Locus tag							
		Monomorphina aenigmatica	*Euglena viridis*	*Euglena gracilis*	*Dunaliella salina*	*Oedogonium cardiacum*	*Floydiella terrestris*	*Schizomeris leibleinii*	*Stigeoclonium helveticum*
		NC_020018.1	NC_020460.2	NC_001603.2	NC_016732.1	NC_011031.1	NC_014346.1	NC_015645.1	NC_008372.1
pcog01	hypothetical protein		I642_p013	EugrCp050					
pcog02	putative LAGLIDADG homing endonuclease							ScleC_p001	StheCp002
pcog03	putative site-specific DNA endonuclease							ScleC_p012	StheCp072
pcog04	putative GIY-YIG homing endonuclease							ScleC_p054	
pcog05	hypothetical protein						FlteC_p024, FlteC_p025		
pcog06	putative HNH homing endonucleases						DUSAC_p024, DUSAC_p070	ScleC_p053	StheCp029
pcog07	putative reverse transcriptase					OecaC_p031			
pcog08	putative HNH homing endonucleases				DUSAC_p071	OecaC_p050		ScleC_p074	StheCp074
pcog09	putative reverse transcriptase	G259_p61	I642_p006	EugrCp005					

This table is adapted from Lim et al. (submitted).

Table S4. Annotated proteins-coding genes in *P. falciparum* HB3 plastid genome (NC_017928.1)

start	end	gene
394	1020	*rps4*
2060	2632	*rpl4*
2636	2863	*rpl23*
2860	3597	*rpl2*
3619	3891	*rps19*
3900	4544	*rps3*
4571	4960	*rpl16*
4974	5198	*rps17*
5195	5551	*rpl14*
5555	5941	*rps8*
5962	6468	*rpl6*
6465	7184	*rps5*
7192	7467	orf
7540	7938	*rps11*
7952	8320	*rps12*
8336	8764	*rps7*
8810	10042	*tufA*
10053	10289	orf
10616	11005	orf
10989	13289	*clpC*
13395	13634	orf
13699	14064	orf
14734	14051	*rps2*
16054	14747	orf
17628	16051	*rpoD*
19364	17637	*rpoC*
22441	19367	*rpoB*
22747	22442	orf
22911	22756	orf
24330	22918	*sufB*
28319	28122	orf

This table is adapted from Lim et al. (submitted)

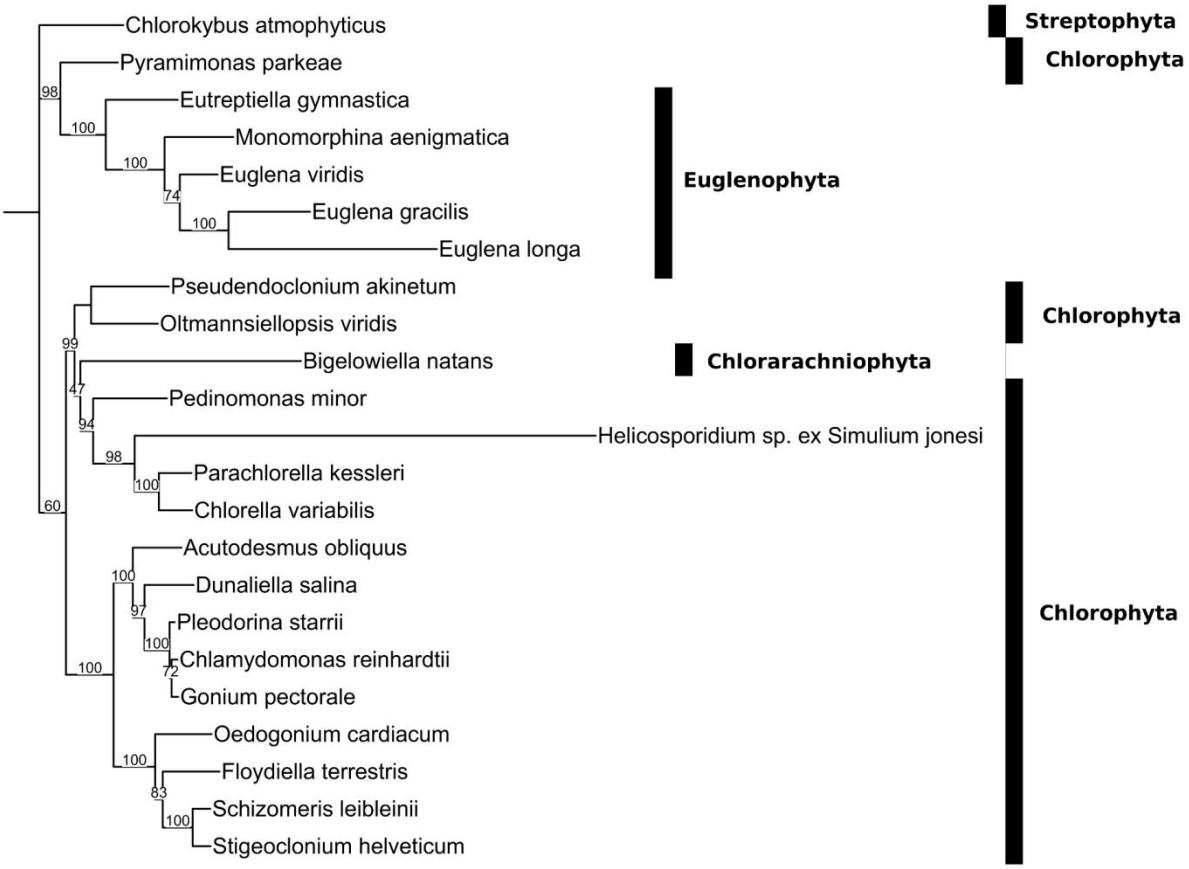


Fig. S1. Best maximum likelihood tree for plastid multi-genes in selected species belonging to Chlorophyta, Euglenophyta, and Chlorarachniophyta. Bootstrap support values for tree nodes are shown. A Streptophyta plastid were used as an outgroup. This figure is adapted from Lim et al.(submitted)