

of *CmACS7* as well as the development of the carpel through expression of *CmACSII*. This is likely due to a tight control of the kinetics of the production of this hormone during sex determination. Because ethylene seems to be a major hormone in sex determination in angiosperms (18), it is likely that our model of sex determination in a monoecious plant can be used as a framework for investigations of sex determination in other plant families. Furthermore, this work may allow easier breeding and optimization of the synchronization of male and female flower development on the same plant so as to improve fruit yields in nonmodel, cultivated *Cucurbitaceae* species.

REFERENCES AND NOTES

1. S. S. Renner, *Am. J. Bot.* **101**, 1588–1596 (2014).
2. B. Charlesworth, D. Charlesworth, *Am. Nat.* **112**, 975 (1978).
3. S. S. Renner, R. E. Ricklefs, *Am. J. Bot.* **82**, 596 (1995).
4. S. C. Barrett, *Nat. Rev. Genet.* **3**, 274–284 (2002).
5. D. Lewis, *Biol. Rev. Camb. Philos. Soc.* **17**, 46–67 (1942).
6. D. Charlesworth, B. Charlesworth, *Heredity* **41**, 137–153 (1978).
7. A. Kocyan, L. B. Zhang, H. Schaefer, S. S. Renner, *Mol. Phylogenet. Evol.* **44**, 553–577 (2007).
8. A. Martin et al., *Nature* **461**, 1135–1138 (2009).
9. A. Boualem et al., *Science* **321**, 836–838 (2008).
10. P. Sebastian, H. Schaefer, I. R. Telford, S. S. Renner, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 14269–14273 (2010).
11. Supplementary text and materials and methods are available as supplementary materials on Science Online.
12. B. Kubicki, *Genet. Pol.* **10**, 101–122 (1969).
13. P. W. Oeller, M. W. Lu, L. P. Taylor, D. A. Pike, *A. Theologis, Science* **254**, 437–439 (1991).
14. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. In the mutants, other amino acids were substituted at certain locations; for example, G39R describes a mutant in which glycine at position 39 is replaced by arginine.
15. S. L. Bai et al., *Planta* **220**, 230–240 (2004).
16. R. Turgeon, K. Oparka, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 13201–13202 (2010).
17. M. L. Tang, Y. B. Pei, W. K. Wong, J. L. Li, *Stat. Methods Med. Res.* **21**, 331–345 (2012).
18. T. F. Law, S. Lebel-Hardenack, S. R. Grant, *Am. J. Bot.* **89**, 1014–1020 (2002).

ACKNOWLEDGMENTS

We thank P. Audiger, F. Vion, N. Giovinazzo, and V. Sarnette for plant care; S. Renner, S. Grant, and G. Marais for critical comments on the manuscript; and V. Gomez-Roldan and C. Clepet for helpful discussions on the manuscript. This work was supported by the Plant Biology and Breeding department in INRA, the grants Program Sacly Plant Sciences (SPS, ANR-10-LABX-40), Initiative d'Excellence Paris-Saclay (Lidex-3P, ANR-11-IDEX-0003-02), L'Agence Nationale de la Recherche MELODY (ANR-11-BSV7-0024), and the European Research Council (ERC-SEXPARTH). We declare no financial conflicts of interest in relation to this work. Gene sequences of *CsACS11* and *CmACS11* have been deposited into GenBank (accession nos. KT715743 and KT715744). A provisional patent (WO/2013/107632) has been deposited by A.Be., A.Bo., and C.T. that covers the use of the androcyte gene for cucurbit breeding. The described material can be obtained from A.Be. under a materials transfer agreement with INRA. The supplementary materials contain additional data.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/350/6261/688/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S13
Tables S1 to S3
References (19–28)
20 June 2015; accepted 22 September 2015
10.1126/science.aac8370

NONHUMAN GENOMICS

The *Symbiodinium kawagutii* genome illuminates dinoflagellate gene expression and coral symbiosis

Senjie Lin,^{1,2*} Shifeng Cheng,^{3,4,5†} Bo Song,^{3†} Xiao Zhong,^{3†} Xin Lin,^{1†} Wujiiao Li,³ Ling Li,¹ Yaqun Zhang,¹ Huan Zhang,² Zhiiliang Ji,⁶ Meichun Cai,⁶ Yunyun Zhuang,^{2†} Xinguo Shi,¹ Lingxiao Lin,¹ Lu Wang,¹ Zhaobao Wang,³ Xin Liu,³ Sheng Yu,³ Peng Zeng,³ Han Hao,⁷ Quan Zou,⁶ Chengxuan Chen,³ Yanjun Li,³ Ying Wang,³ Chunyan Xu,³ Shanshan Meng,¹ Xun Xu,³ Jun Wang,^{3,8,9} Huamming Yang,^{3,9,10} David A. Campbell,¹¹ Nancy R. Sturm,¹¹ Steve Dagenais-Bellefeuille,¹² David Morse¹²

Dinoflagellates are important components of marine ecosystems and essential coral symbionts, yet little is known about their genomes. We report here on the analysis of a high-quality assembly from the 1180-megabase genome of *Symbiodinium kawagutii*. We annotated protein-coding genes and identified *Symbiodinium*-specific gene families. No whole-genome duplication was observed, but instead we found active (retro) transposition and gene family expansion, especially in processes important for successful symbiosis with corals. We also documented genes potentially governing sexual reproduction and cyst formation, novel promoter elements, and a microRNA system potentially regulating gene expression in both symbiont and coral. We found biochemical complementarity between genomes of *S. kawagutii* and the anthozoan *Acropora*, indicative of host-symbiont coevolution, providing a resource for studying the molecular basis and evolution of coral symbiosis.

Dinoflagellates are alveolates, with the most likely parasitic apicomplexans as their closest relatives (fig. S1A). Members of the genus *Symbiodinium* are essential photosynthetic endosymbionts in coral reefs (1). Dinoflagellates show enigmatic genetic and cytological characteristics, including permanently condensed chromosomes and a high proportion of diverse methylated nucleotides, and often feature large nuclear genomes (up to 250 Gb) (2). We report a 0.935-Gbp assembly of the 1.18-Gbp genome of

Symbiodinium kawagutii (figs. S1B and S2), a Clade F strain originally isolated from a Hawaiian reef ecosystem (3). A high-quality *S. kawagutii* genome assembly corresponding to ~80% of the genome was achieved from ~151-Gbp Illumina genome shotgun sequence (~130x genome coverage) (tables S1 to S4 and fig. S3). Genome annotation revealed 36,850 nuclear genes, with 68% occurring in families (1.69 genes per family) (table S5). Only ~9% (3280) of *S. kawagutii* genes were in tandem arrays (1279 clusters) (table S6), with 2 to 10 repeats (76% being ≤4 repeats) per array. The genome encodes the common metabolic pathways expected for typical photosynthetic eukaryotes (fig. S4 and table S7), and we found genes involved in sexual reproduction, cyst formation and germination, and telomere synthesis (table S8). The telomeric motif (TTTAGGG)_n was identified at the ends of scaffolds and was also detected by fluorescence *in situ* hybridization (fig. S1B).

Globally, our analysis revealed extensive genomic innovation in dinoflagellates. A total of 25,112 gene families were clustered from the genomes of *S. kawagutii* and eight other species representing higher plants, chlorophytes, rhodophytes, diatoms, phaeophytes, alveolates, and cnidarians. *S. kawagutii* has 12,516 gene families, of which 7663 were gained in the ancestor of *Symbiodinium* (Fig. 1A and table S9). These genes were enriched in 62 metabolic gene ontologies (table S10). When the gene families were normalized to *z* scores to balance the effect of different total gene numbers, 96 gene families had shrunk (table S11) and 265 gene families had expanded in *Symbiodinium* (table S12). The LINE-1 reverse transcriptase (a retroelement) is the most highly expanded family.

State Key Laboratory of Marine Environmental Science and Marine Biodiversity and Global Change Research Center, Xiamen University, Xiamen 361101, China. [†]Department of Marine Sciences, University of Connecticut, Groton, CT 06340, USA. [‡]Beijing Genomics Institute (BGI)-Shenzhen, Shenzhen, 518083, China. [§]Hong Kong University (HKU)-BGI Bioinformatics Algorithms and Core Technology Research Laboratory, The Computer Science Department, The University of Hong Kong, Hong Kong, China. [¶]School of Biological Sciences, The University of Hong Kong, Pokfulam, Hong Kong, China. [¶]State Key Laboratory of Stress Cell Biology, School of Life Sciences, Xiamen University, Xiamen 361101, China. [¶]Bioinformatics Institute, Agency for Science, Technology and Research, Singapore. [¶]Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Denmark. [¶]Princess Al Jawhara Center of Excellence in the Research of Hereditary Disorders, King Abdulaziz University, Jeddah, Saudi Arabia. [¶]James D. Watson Institute of Genome Science, Hangzhou, China. [¶]Department of Microbiology, Immunology, and Molecular Genetics, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA 90095, USA. [¶]Institut de Recherche en Biologie Végétale, Département de Sciences Biologiques, Université de Montréal, Montréal, Canada. ^{}Corresponding author. E-mail: senjie.lin@uconn.edu. [†]These authors contributed equally to this work. [‡]Present address: College of Environmental Science and Engineering, Ocean University of China, Qingdao, Shandong 266100, China.

Our synteny and homology analysis showed no evidence of whole-genome duplication, because little collinearity within *S. kawagutii* genome was observed (table S13). Instead, the *S. kawagutii* genome shows evidence of transposon propagation, in particular long terminal repeat (LTR) retrotransposons and DNA transposons (table S14), which contributes to differences between *S. kawagutii* and *S. minutum* (Fig. 1B). Furthermore, protein domains linked to transposons (table S15) may lead to proliferation of these protein domains in the genome; in the case of cytosine methyltransferase, the resulting expansion of the gene family may help explain the extensive DNA methylation seen in dinoflagellates; in the case of retroelements such as reverse transcriptase and integrase, the expanded families may increase the frequency of transcript retrotransposition into the genome. In keeping with this latter, we found numerous genes with a full (62 genes) or partial (5506 genes) dinoflagellate spliced leader (DinoSL) (4) in their 5' untranslated region (table S16). The 22-nucleotide (nt) dinoSL is *trans*-spliced to the 5' end of all mRNAs, and its presence in the genome is a signature of retrotranscript insertion (5); thus this represents an efficient mechanism for gene family expansion. Last, horizontal gene transfer (HGT) may also contribute to *S. kawagutii* genome innovation. Conservatively, we found 56 potential HGT genes, 41 of which had best Basic Local Alignment Search Tool (BLAST) hits to marine bacteria (table S17 and fig. S5).

The partial genome (~41%) of *S. minutum* (6) allowed some comparative genomic studies. *S. minutum* and *S. kawagutii* have similar genome sizes and gene numbers (table S5) and show some genomic collinearity (Fig. 1B and table S18) and gene ontology profiles (table S19 and fig. S6A). MUMmer (Maximal Unique Matches) alignment data showed that 2.17% (20.4 Mb) of the *S. kawagutii* genome matched to *S. minutum*, and only 5.92% (36.5 Mb) of the *S. minutum* genome matched to *S. kawagutii*. This divergence was confirmed by the reciprocal mapping of their raw reads (fig. S6B and tables S5 and S20), implying that these two species are more diverged than usually assumed. Yet both genomes showed expansion of gene families involving cargo transport and stress responses (fig. S6C and tables S21 and S22), which may reflect the shared symbiotic lifestyles.

The transcriptional machinery of dinoflagellates does not contain the typical eukaryotic TATA box promoter element (2). Instead of a TATA-box binding protein (TBP), dinoflagellates express a TBP-like factor that has a stronger affinity to TTTT than to TATA (7). A global search of the *S. kawagutii* genome 1000-bp (base pair) region upstream of putative start codons revealed 564 conserved motifs, which were grouped into 108 clusters on the basis of sequence similarities (table S23). About 92% of these were located within 100 bp upstream of the start codon. The motifs with the most conserved positions are remnants of SL (Fig. 1C). Motifs TTTG and TTTT were found in the upstream regions of 34,524 and 35,348 genes, respectively (94% and 96% of the gene repertoire). Curiously, although both are part

of the SL, the TTTG motif has a position consistent with that of the SL, whereas TTTT tends to be further upstream (Fig. 1D). This suggests that the TTTT may serve as a core promoter motif replacing the TATA box used by other eukaryotes. The TTTT is typically 30 bp upstream from a potential transcriptional start site (fig. S7). The next most highly enriched motif, (TATG)2, was associated with only 257 promoters and is thus more likely to be a binding target of specific regulators.

Sequences from the genome and purified small RNAs predicted 367 and 354 mature microRNAs (miRNAs), respectively (3), with 102 of the latter (table S24) retained after stringent filtering and structural analysis (fig. S8A). We matched 255 of the genome-predicted miRNAs to 99 of the small RNA-based miRNAs. The mature miRNA candi-

dates varied in length from 21 to 24 nt, with most (91.5%) containing 22 nt. Northern blot analysis revealed a decreased expression level for some miRNA in cultures grown at 35°C instead of 25°C (Fig. 2A); of these, scaffold270_6017 is predicted to target heat shock proteins 90 and 70 (table S25), consistent with an expected up-regulation in translation of these thermal stress proteins. A bias toward uridine (U) (47.9%) and against guanine (G) (5.0%) was observed at the 5'-end ultimate nucleotide, as was in *Arabidopsis thaliana* (8). Interestingly, of the 102 mature miRNA sequences, 49 were similar to animal miRNAs, 11 to plant miRNAs, and 1 to viral miRNAs (3). *S. kawagutii* thus has a miRNA reservoir dominated by miRNAs with considerable sequence identity to those found in animals (fig. S8B).

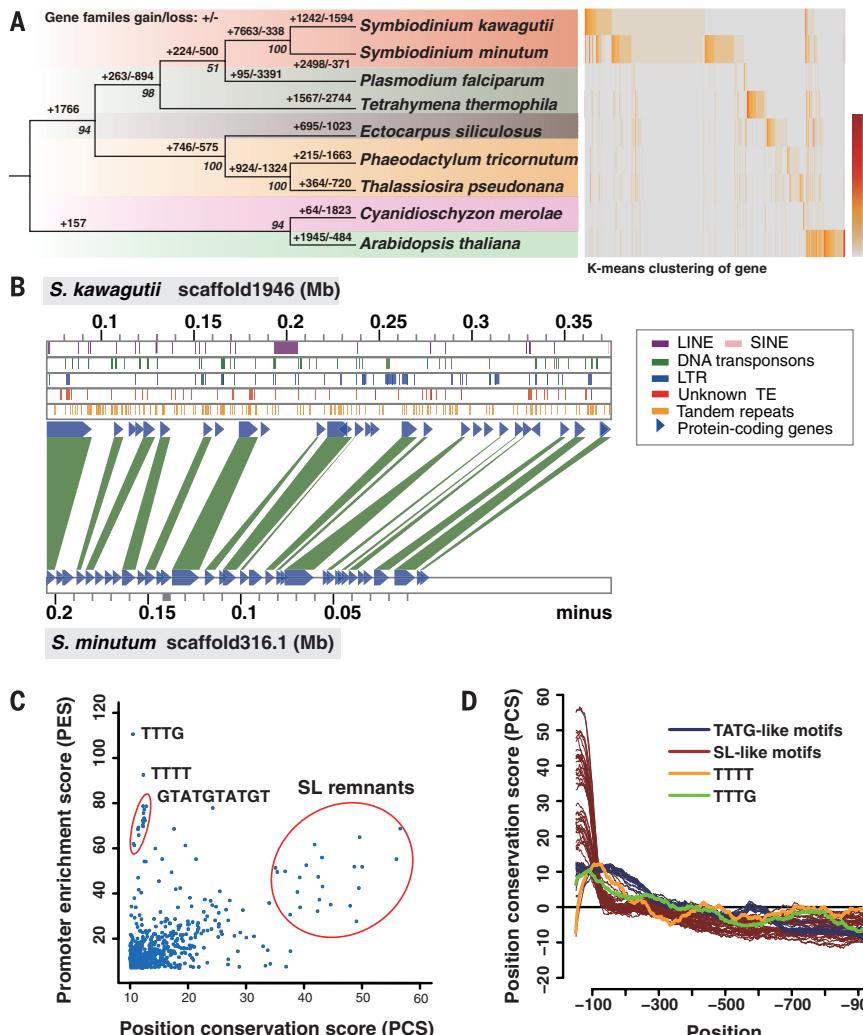


Fig. 1. Comparative genomic analysis between *S. kawagutii* and other eukaryotes. (A) (Left) Predicted pattern of gain or loss of gene families across eukaryotes shown on a phylogenetic tree inferred from genome data. Numbers on branches indicate the number of gene families gained (+) or lost (-); those at the left of the nodes are bootstrap values supporting the tree topology. (Right) K-means clustering of gene families based on number of members. Columns represent gene families, and rows are species of eukaryote. (B) Synteny between regions of the genomes of *S. kawagutii* and *S. minutum*. (C) TTTG, TTTT, and (TATG)2 are the top three motifs enriched in the *S. kawagutii* upstream regions and are potential novel promoter elements. (D) Promoter enrichment scores for selected motifs as a function of distance upstream from the start codon.

We identified 1 perfect (plant-type) and 6026 partial (animal-type) complementarity miRNA targets in the *S. kawagutii* genome. Among the genes with partial complementarity, 381 were potentially targets of miRNAs with higher expression levels (read counts > 1000), suggesting that these genes are more likely to be regulated by miRNAs. In total, 2557 of all the potential target genes were annotated with known functions (table S25) with enrichments in biological processes of carbohydrate metabolism, transcription regulation, and biosynthesis of amino acids and antibiotics (tables S26 and S27). miRNA targets are often clustered in networks of interacting proteins (3), in which some genes are targeted by many miRNAs and some miRNAs target multiple genes (Fig. 2B, fig. S9, and table S28). In addition, the *S. kawagutii* genome harbors small RNA-degrading nucleases 1 and 3, one of which is itself a miRNA target (table S25). Thus, the evidence for miRNA-based gene regulatory machinery is robust and extensive, complementary to the limited transcriptional regulation documented in dinoflagellates (2).

We identified a double-stranded RNA (dsRNA)-gated channel protein Systemic RNA Interference Deficiency-1 (SID-1) (9) required for systemic RNA interference in animals (10). SID-1 sequences in *Symbiodinium* and Cnidaria are similar, suggestive of horizontal gene transfer (fig. S10) (3). Pathogen-to-host miRNA transfer has been shown to silence host immunity genes in plants (11). We identified 1514 coral genes (table S29) (6.4% of the total number) as potential targets of *S. kawagutii* miRNAs; these had similar molecular functions [Gene Ontology (GO) slim category] as targets in

S. kawagutii (Fig. 2C), which were enriched in GO categories related to protein modification and regulation of transcription and cell growth (table S30). This suggests that transferred miRNAs might regulate similar processes in symbiont and host.

Recognition of *Symbiodinium* by the host cells is mediated primarily through binding of *Symbiodinium* high-mannose glycans by lectins on the coral cell surface (12, 13) (Fig. 3). We found a glycan biosynthesis pathway in *S. kawagutii* lacking several enzymes catalyzing the final steps of the common glycan biosynthesis pathway (fig. S11). This altered pathway is predicted to produce a (GlcNAc)5(Man)5(Asn)1 glycan that carries abundant free mannose branches and terminal mannose-mannose units available for lectin binding, consistent with previous findings (14). However, the enzymes involved in mannose-rich glycan biosynthesis differ between *S. kawagutii* and *S. minutum* (table S31), suggesting that variations in the glycoprotein structure may tune the host recognition specificity.

Other *Symbiodinium* genes may also be related to symbiosis (table S32). These include homologs of nodulation factors involved in establishing the symbiosis between legume and the nitrogen-fixing bacteria rhizobia, as well as cell surface proteins with a role in pathogen infection or host recognition. Furthermore, some *S. kawagutii* and *S. minutum* proteins share homology with *Plasmodium falciparum* proteins involved in the interaction between the parasite and its host (15).

To assess their role in symbiosis, we compared genes encoding transporters in *S. kawagutii* with those in *Acropora digitifera*, the only sequenced

coral species (16) (Fig. 3 and table S33). Remarkably, nearly half of the numerous transporters (>300) in *S. kawagutii* are shared by this coral (table S34). Both dinoflagellate and coral genomes encode transporters of C (bicarbonate), N, P, and trace metals, as well as carbon-concentrating mechanism enzymes, and many of these are lacking in the two nonsymbiotic cnidarians *Hydra magnipapillata* and *Nematostella vectensis* (table S35). Most of the 49 carbonic anhydrases (CA) genes in the *S. kawagutii* genome are cytoplasmic, suggesting that cytoplasmic CA is critical for CO₂ acquisition; only two δ-CAs are predicted to be localized at the plasma membrane and one β-CA in the thylakoids.

The *S. kawagutii* genome also contains the complete biosynthesis pathways of all the standard amino acids except lysine and histidine and could potentially supply nine of the amino acids that *A. digitifera* cannot produce (fig. S12 and table S35). Interestingly, the majority of the *S. kawagutii* transporters, as well as some of their coral cognates, are potential miRNA targets (red circles, Fig. 3).

Symbiodinium-to-coral translocation of photosynthates is critical for reef growth, with either glycerol (17) or glucose (18) translocated. The *S. kawagutii* genome contains 12 genes encoding glycerol-3-phosphate dehydrogenase, an enzyme essential for glycerol production in yeast (19), as well as a plasma membrane aquaporin with glycerol transport ability and low- and high-affinity glucose transporters (Fig. 3 and table S35). However, the *A. digitifera* genome (16) contains only the transporter for glucose and not for glycerol.

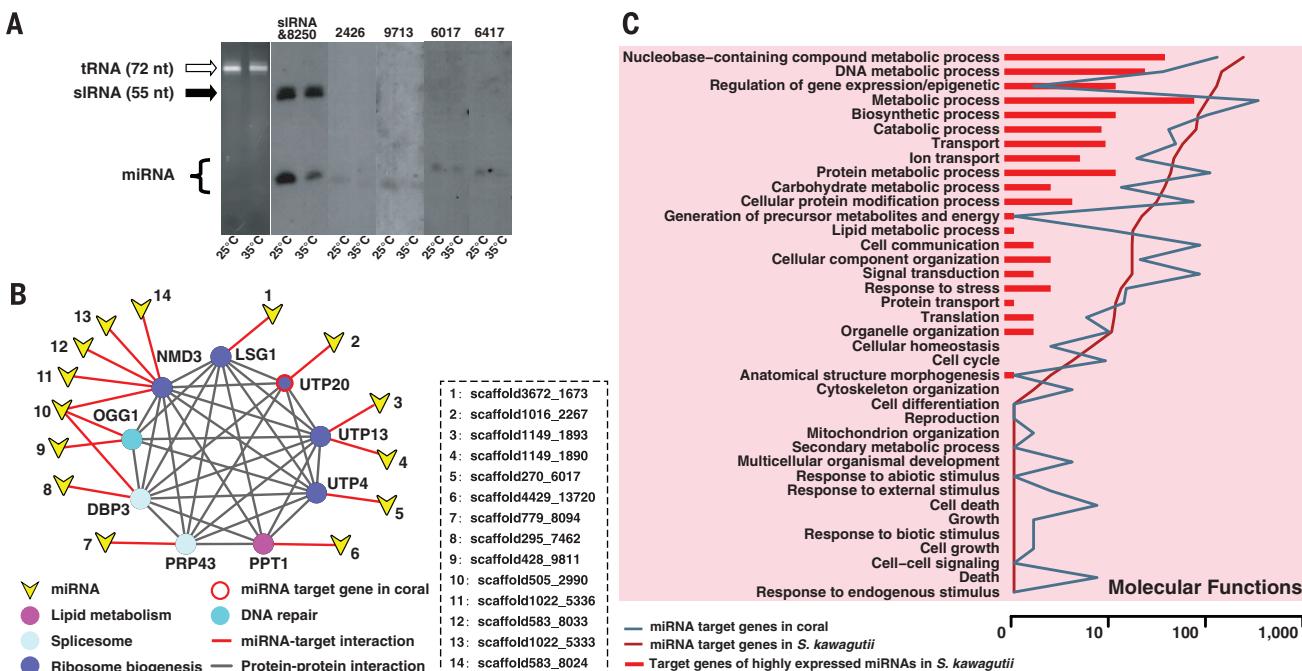
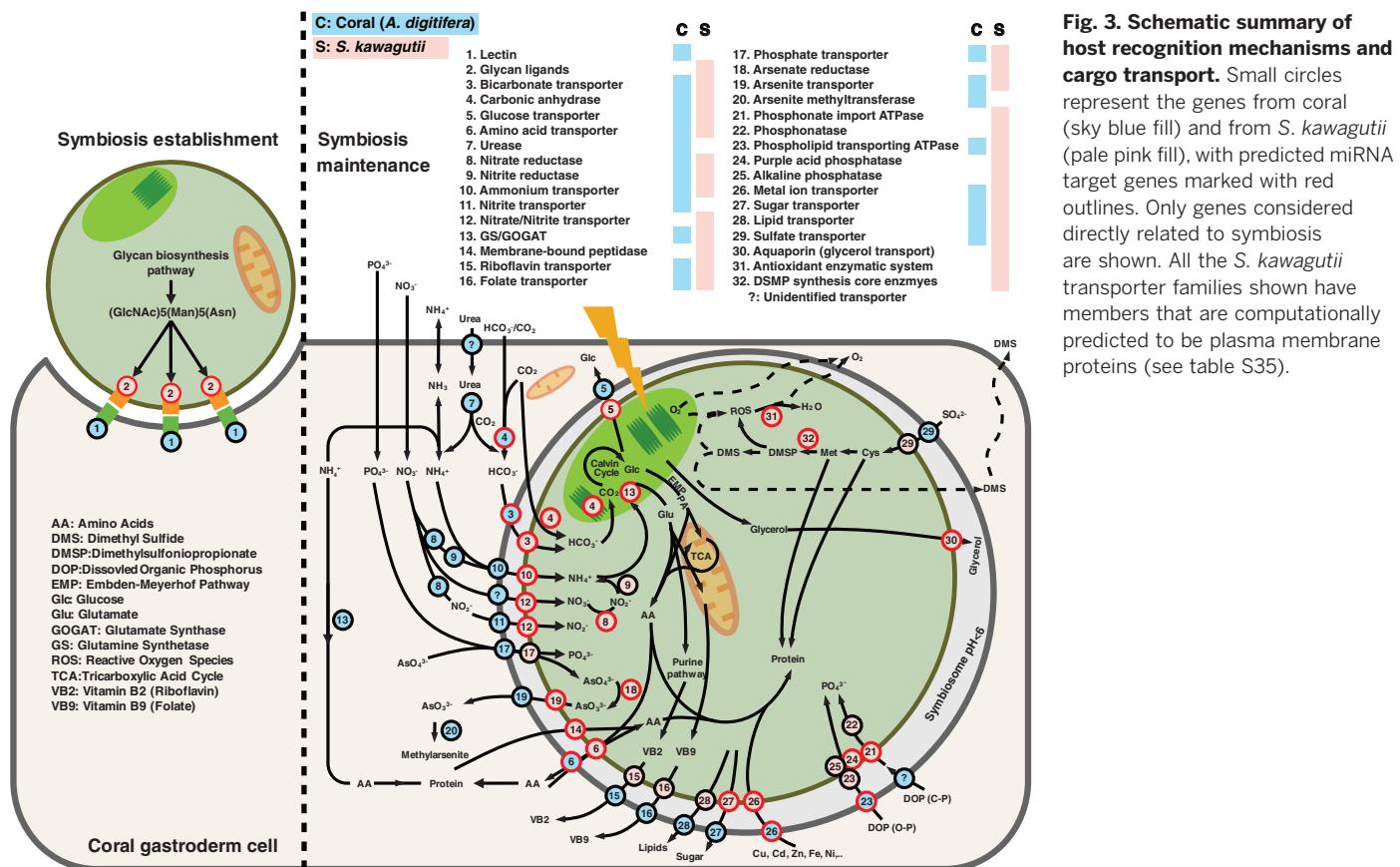


Fig. 2. miRNA in *S. kawagutii* and potential target genes in *S. kawagutii* and coral. (A) Northern blot analysis of *Symbiodinium* miRNAs. Most miRNAs are 22 nucleotides as assessed by the migration of DNA oligonucleotides. (B) A putative protein-protein interaction network of miRNA target genes. The network is a cluster of highly connected nodes (genes) showing miRNA-target interactions (red lines) and protein-protein interactions (gray lines). Node colors represent different KEGG (Kyoto Encyclopedia of Genes and Genomes) and GO functional annotations (see table S28). (C) GO categorization of predicted miRNA target genes in both *S. kawagutii* and coral.



This suggests that glucose can be exported to the coral cells, whereas glycerol is exported only to the symbiosome, possibly as an osmolyte (20).

S. kawagutii possesses a large ensemble of genes potentially conferring tolerance to thermal stress and ultraviolet irradiation, including expanded gene families encoding heat shock proteins and DNA repair/recombination proteins (table S22). There is also a large set of antioxidant genes, including the large thioredoxin gene family, a diverse set of genes for (Cu/Zn-, Mn/Fe-, Ni-dependent) superoxide dismutases (SOD), and ascorbate peroxidases (APx). We found a Ni-dependent SOD, rarely reported for marine algae, consistent with the abundant high-affinity nickel transporter genes in this species (fig. S6C), as well as six genes encoding xanthine dehydrogenase/oxidase, which catalyzes the oxidation of xanthine to uric acid in purine metabolism (table S35). Uric acid forms crystalline deposits in *Symbiodinium* that function as an N reserve (21), and it is a potent antioxidant.

Unexpectedly, the *S. kawagutii* genome lacks the genes of the four major photoprotector mycosporine-like amino acids (MAA) biosynthesis enzymes: dehydroquinate synthase (DHQS), O-methyltransferase (O-MT), ATP-grasp, and non-ribosomal peptide synthetase (NRPS). Their loss may thus represent a coevolution of *S. kawagutii* with its host, because all four genes are found in *A. digitifera*, one of which shows a close relationship with that in other dinoflagellates (fig. S13).

This study provides a portrait of a symbiotic dinoflagellate genome, with insights into genome evolution and regulation of gene expression in dinoflagellates and the molecular basis of coral-*Symbiodinium* symbiosis. Our results are a stepping-stone to understanding how the genetic complementarity between anthozoans and *Symbiodinium* can explain host specificity (1, 22) and to determining the molecular mechanisms responsible for coral bleaching.

REFERENCES AND NOTES

1. A. C. Baker, *Annu. Rev. Ecol. Evol. Syst.* **34**, 661–689 (2003).
2. S. Lin, *Res. Microbiol.* **162**, 551–569 (2011).
3. Materials and methods are available as supplementary materials on *Science Online*.
4. H. Zhang et al., *Proc. Natl. Acad. Sci. U.S.A.* **104**, 4618–4623 (2007).
5. C. H. Slamovits, P. J. Keeling, *Curr. Biol.* **18**, R550–R552 (2008).
6. E. Shoguchi et al., *Curr. Biol.* **23**, 1399–1408 (2013).
7. D. Guillebault et al., *J. Biol. Chem.* **277**, 40881–40886 (2002).
8. S. Mi et al., *Cell* **133**, 116–127 (2008).
9. J. D. Shih, C. P. Hunter, *RNA* **17**, 1057–1065 (2011).
10. W. M. Winston, C. Molodowitch, C. P. Hunter, *Science* **295**, 2456–2459 (2002).
11. A. Weisberg et al., *Science* **342**, 118–123 (2013).
12. L. K. Bay et al., *Diversity* **3**, 356–374 (2011).
13. E. M. Wood-Charlson, L. L. Hollingsworth, D. A. Krupp, V. M. Weis, *Cell. Microbiol.* **8**, 1985–1993 (2006).
14. K.-L. Lin, J.-T. Wang, L.-S. Fang, *Zool. Stud.* **39**, 172–178 (2000).
15. V. K. Goel et al., *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5164–5169 (2003).
16. C. Shinzato et al., *Nature* **476**, 320–323 (2011).
17. L. Muscatine, *Science* **156**, 516–519 (1967).
18. M. S. Burriesci, T. K. Raab, J. R. Pringle, *J. Exp. Biol.* **215**, 3467–3477 (2012).
19. J. Albertyn, S. Hohmann, B. A. Prior, *Curr. Genet.* **25**, 12–18 (1994).
20. A. B. Mayfield, R. D. Gates, *Comparative Biochem. Physiol. A: Mol. Integr. Physiol.* **147**, 1–10 (2007).
21. P. L. Clode, M. Saunders, G. Maker, M. Ludwig, C. A. Atkins, *Plant Cell Environ.* **32**, 170–177 (2009).
22. S. K. Davy, D. Allemand, V. M. Weis, *Microbiol. Mol. Biol. Rev.* **76**, 229–261 (2012).

ACKNOWLEDGMENTS

This work was supported by Natural Science Foundation of China grants K16110 and K16044 (to S.L.), U.S. National Science Foundation grant OCE-0854719 (to S.L. and H.Z.), U.S. NIH awards AI056034 and AI073806 (to D.A.C. and N.R.S.), National Science and Engineering Research Council of Canada grant 171382-03 (to D.M.), and various funds to BGI-Shenzhen, Shenzhen [State Key Laboratory of Agricultural Genomics, Guangdong Provincial Key Laboratory of core collection of crop genetic resources research and application (2011A091000047), Shenzhen Engineering Laboratory of Crop Molecular design breeding, and China National GeneBank-Shenzhen]. The genomic sequences and the annotated genes of *S. kawagutii*, as well as RNA sequencing data (unigenes), are available in our Symka Genome Database in Xiamen University: http://bioinf.xmu.edu.cn/symka_new/; these data have also been deposited into the National Center for Biotechnology Information Short Read Archive (SRA) under accession number SRA148697.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/350/6261/691/suppl/DC1
Materials and Methods
Figs. S1 to S13
Tables S1 to S35
References (23–56)

17 July 2015; accepted 25 September 2015
10.1126/science.aad0408

Supplementary Materials for

The *Symbiodinium kawagutii* genome illuminates dinoflagellate gene expression and coral symbiosis

Senjie Lin,* Shifeng Cheng, Bo Song, Xiao Zhong, Xin Lin, Wujiao Li, Ling Li, Yaqun Zhang, Huan Zhang, Zhiliang Ji, Meichun Cai, Yunyun Zhuang, Xinguo Shi, Lingxiao Lin, Lu Wang, Zhaobao Wang, Xin Liu, Sheng Yu, Peng Zeng, Han Hao, Quan Zou, Chengxuan Chen, Yanjun Li, Ying Wang, Chunyan Xu, Shanshan Meng, Xun Xu, Jun Wang, Huanming Yang, David A. Campbell, Nancy R. Sturm, Steve Dagenais-Bellefeuille, David Morse

*Corresponding author. E-mail: senjie.lin@uconn.edu

Published 6 November 2015, *Science* **350**, 691 (2015)

DOI: 10.1126/science.aad0408

This PDF file includes:

Materials and Methods
Figs. S1 to S13
References

Other Supplementary Material for this manuscript includes the following: (available at www.sciencemag.org/cgi/content/full/350/6261/691/DC1)

Tables S1 to S35 as Excel files

Materials and Methods

Accession codes of the genome dataset

The genomic sequences and the annotated genes of *S. kawagutii*, as well as RNA sequencing data (unigenes), are available in our Symka Genome Database in Xiamen University: http://bioinf.xmu.edu.cn/symka_new/; these data have also been deposited into NCBI Short Read Archive (SRA) under accession number SRA148697.

Algal culture

Symbiodinium kawagutii strain CCMP2468 (Clade F) was originally isolated from a Hawaiian coral reef (host originally reported as *Montipora verrucosa* [<https://ncma.bigelow.org/ccmp2468>] but subsequently questioned [23]). The culture was grown in L1-medium containing ampicillin (200 mg/L), kanamycin (100 mg/L) and streptomycin (100 mg/L) to minimize the growth of bacteria. Temperature was controlled at 25°C and illumination was 200 $\mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ under a 14:10 light:dark cycle (these and L1-medium combined are referred to hereafter as normal conditions). The culture was replenished with fresh antibiotic-containing medium weekly, and handled strictly aseptically. Cells were harvested at exponential stage.

DNA sample preparation for Illumina sequencing

Cells were first washed with TEN buffer (100 mM Tris·Cl pH 8, 100 mM EDTA pH 8, 1.5 M NaCl) containing 0.5 mg/mL proteinase K and 7% SDS at 65°C for 2 hours, a condition expected to lyse contaminating bacteria that might have survived the antibiotics treatment. After centrifugation, the pelleted *Symbiodinium* cells were homogenized in 0.1 mM EDTA pH 8 using bead beating with 0.5 mm-zirconia/silica beads (Biospec Products, INC, Bartlesville, OK 74005) on a MP Fast Prep-24 Tissue and Cell Homogenizer (MP Biomedicals; Solon, OH 44139). DNA was extracted using a CTAB protocol (24) and precipitated using isopropanol.

Sequencing and assembly

Paired-end libraries with insert sizes of 170 bp, 200 bp, 500 bp, 800 bp, 2 kb, 5 kb, 10 kb and 20 kb were constructed following standard Illumina protocols. The libraries were sequenced on an Illumina HiSeq 2000 platform. A total of 266 Gb (~229X genome depth) paired-end data were generated. After filtering off duplicated, low quality reads and reads with adaptor sequences, ~151 Gb high-quality clean reads remained, which were then subjected to a SOAPdenovo pipeline for genome assembly (25). K-mer analysis (26) was performed to survey the genome size, heterozygosity, and repeat content before genome assembly. The peak of K-mer frequency (M) is determined by the real sequencing depth of the genome (N), read length (L), and the length of K-mer (K) following the formula: $M=N \times (L-K+1)/L$. This formula enables accurate estimation of N, and hence an estimation of the genome size for homozygous diploid or haploid genomes. K-mer analyses were repeated 12 times with different amounts of data and K-mer sizes (15-, 17-, 19- and 21-mer) (table S2). All these analyses indicated homozygosity of the genome and gave similar estimations of the genome size. The smallest estimate (1.12 Gb) was obtained with 15-mer analysis with ~44 G bases while 17-mer analysis using ~14 G

bases gave the largest estimate (1.29 Gb) (table S2). The final genome size estimate (1.18 Gb) was obtained by averaging all 12 estimates.

Fosmid library construction and assembly evaluation

About 100 µg *S. kawagutii* genomic DNA was extracted with care to minimize fragmentation, sheared, and subsequently separated by pulsed field electrophoresis. DNA fragments of ~40 kb were purified and ligated into a fosmid vector (CopyControlTM Fosmid Library Production Kit, Epicentre, USA) according to the manufacturer's protocol; this was performed by Amplicon Express (ampliconexpress.com). Ten of the resulting clones were randomly picked and the DNA isolated, pooled, sheared and sequenced using the Sanger method and assembled as reported (27). The genomic DNA scaffolds were aligned to the fosmid sequences using BLASTn with an E value of 1e⁻²⁰ to evaluate the continuity of the assembly.

Transcriptome sequencing

Cell samples were collected at various growth stages of the *S. kawagutii* culture. Total RNA/ mRNA extraction and Sanger library construction and sequencing were performed as described (24), generating 12,990 high quality reads clustered into 3036 unigenes. For 454 sequencing, *S. kawagutii* mRNA extracted from different samples were pooled, and a regular 454 cDNA library was constructed using GS Titanium SV emPCR Kit (Lib-L) (Roche, Germany) following the manufacturer's protocol, while a 5'-end enriched 454 cDNA library was constructed using DinoSL (28). These 454 cDNA libraries were sequenced on the GS FLX System, generating a total of 834,546 high quality reads (653,560 and 180,986 respectively, read length 170-500 bp). These were assembled into 40,052 and 57,324 unigenes as described (29). Next, all the unigenes from the three transcriptome datasets were pooled and used as input to run a clustering at 97% identity cutoff, yielding a final set of 70,985 unigenes. This set of unigenes (named *S. kawagutii* ESTs) were used for assessing the accuracy of the genome assembly and for gene annotation.

Identification of potential microbial contaminant sequences

To detect sequences of any remaining antibiotic-resistant microbial contaminants in the genome dataset, we conducted GC-content depth analysis to identify possible microbe-derived reads. These potential contaminant sequences were further examined by BLAST comparisons against both the NCBI NT database (Version 20110911) and the Fungi genome database (Version 20101213).

Gene annotation

We used a combination of *de novo* prediction, homology searches and RNA-aided annotation. *De novo* prediction was performed using AUGUSTUS (30). For homologous annotation, we queried the *S. kawagutii* genome scaffolds against a database containing protein sequences from 21 organisms (*S. minutum*, *Perkinsus marinus*, *Plasmodium falciparum*, *Theileria parva*, *Cryptosporidium parvum*, *Paramecium tetraurelia*, *Phaeodactylum tricornutum*, *Ectocarpus siliculosus*, *Guillardia theta* (nuclear), *Emiliania huxleyi*, *Bigelowiella natans*, *Cyanidioschyzon merolae*, *Chlamydomonas reinhardtii*, *Arabidopsis thaliana*, *Trypanosoma brucei*, *Saccharomyces cerevisiae*,

Caenorhabditis elegans, *Homo sapiens*, *Synechocystis* sp., *Escherichia coli* and *Aciduliprofundum boonei*) with an E-value cutoff of 10^{-4} , followed by a Genewise annotation. Unigene sequences of *S. kawagutii* were also mapped back to the genome using BLAST (31) and then assembled by PASA (32) according to the genomic locations. The final consensus gene sets were generated by combining all the predictions using GLEAN (33).

Analysis of repetitive sequences in the genome

We used the software packages RepeatMasker (34) (Ver 3.2.7) and RepeatProteinMask (<http://www.repeatmasker.org/>, Ver 3.2.2) to search homologous repeat elements in the database Repbase (35). A *de novo* repeat sequences database of *S. kawagutii* was built using LTR-FINDER (36) and RepeatModeler (37) and subsequently searched with RepeatMasker. Tandem repeats were predicted using the TRF program (38).

Identification of horizontal gene transfer (HGT)

BLAST analysis of predicted *S. kawagutii* genes (after stringent contaminant filtering) was performed against the GenBank nr database, and genes showing a best hit to bacterial sequences (with E value cutoff -5) were collected. Of these genes found only in bacteria and dinoflagellates were catalogued as HGT candidates, which were further verified by EST mapping and analysis of their flanking genes. Candidates flanked by at least one eukaryotic gene at either or both sides and those with EST support were then classified as putative HGT genes.

Comparative genomics analysis for genome differences, synteny and symbiosis-related transporters

Paralogous duplicates and collinear orthologous blocks within the *S. kawagutii* genome and between *S. kawagutii* and closely related species were identified by MCScan (39) (using genes as anchors, at least 5 gene pairs/block) and MUMmer (40) (whole genome sequence alignment). The similarity between these two assemblies was evaluated using MUMmer (41) with default parameters. To further compare the two genomes, 39.1 Gbp of Illumina HiSeq IIx *S. minutum* reads was downloaded and mapped to the *S. kawagutii* assembly, while 226 Gbp of paired-end reads from *S. kawagutii* was mapped to the *S. minutum* assembly using SOAPdenovo2 package (25) allowing 10 mismatches. Genes annotated as transporter, channel, carrier, pump, exchanger, permease and facilitator were collected from functional annotation files generated from KEGG, SwissProt, InterProt and TrEMBLE databases for each of the *A. digitifera*, *S. kawagutii*, *S. minutum*, *P. tricornutum* and *Thalassiosira pseudonana* genomes and transcriptomes of the free living phytoplankton species *Karlodinium veneficum* and *Amphidinium carterae*. Transporters shared by *S. kawagutii*, *S. minutum* and coral (*A. digitifera*) were catalogued as genes potentially involved in cargo trafficking between coral and *Symbiodinium*. These genes were also used as queries to search in the genomes of the non-symbiotic cnidarians *Hydra magnipapillata* and *Nematostella vectensis* to determine if they are absent.

Gene family analysis

The genome of the higher plant *A. thaliana*, four algal genomes (*C. merolae*, *T. peseudonana*, *P. tricornutum*, *C. crispus*, *E. siliculosus*), and the genomes of the apicomplexan *P. falciparum* and the ciliate *Tetrahymena thermophila*, which are closely related to dinoflagellates, were compared with the two *Symbiodinium* spp genomes. An all-against-all BLASTP analysis was performed for all genes of the 20 species mentioned above with an e-value cutoff of $1e^{-5}$. Homologous proteins were clustered using OrthoMCL (43) with default parameters.

OrthoMCL-clustered gene family data were then used to identify *S. kawagutii*-specific and expanded gene families, as well as dinoflagellate-specific families. Since dinoflagellate genomes are much larger than the others used for comparison, gene family sizes were normalized using a z-score :

$$z = \frac{x - \mu}{\sigma}$$

where x is the number of genes in a family for a given species, μ is the population mean and σ is the population standard deviation. Unique genes and species-specific genes were excluded from the analysis. Expanded and contracted gene families were identified using student's t test with a cutoff of P-value <0.001 . Gene ontology categorization was conducted by the map2slim program (<http://search.cpan.org/~cmungall/go-perl/scripts/map2slim>). The numbers of gained and lost gene families in each node or branch in the phylogenetic tree were inferred using the Dollop algorithm in the PHYLIP package (Felsenstein, 2002; PHYLIP version 36a3, distributed by the author, Department of Genome Sciences, University of Washington, Seattle, USA). Gene families were clustered into 9 groups using the k-means algorithm in the R package (<http://www.r-project.org/>) with K set at nine.

Detection of SL-containing genes and analysis of their genomic features

Due to the limited power of BLAST in searching short sequences, we artificially generated a set of DinoSL variants by replacing each nucleotide with A, T, C, G, as well as with an empty gap. These variants together with the canonical DinoSL sequences were then queried against the 500 bp upstream sequences of genes with the parameters were set as “-e 100,000 -W 4 -G 0 -E 0 -q 0” for BLAST. This detected a total of 5,568 SL-containing genes. Sixty-two genes contain full length DinoSL, of which 28 are canonical DinoSL sequences while the others are variants with one to three substitutions. These sequences are likely to represent retrotransposition of SL-containing mRNA. To search for genomic features enabling retrotransposition we extracted 50 bp upstream of SL remnants. The frequencies for all motifs (2 to 10 nucleotides in length) in the upstream sequences as well as the GC content were computed and compared with the 5'UTR of genes lacking SL remnants. No conserved motif was observed although the GC content was significantly lower in the 5' flanking region of the SL-containing genes than in the other genes (46.64% to 51.58%, $P=2.2e-16$).

miRNA prediction, sequencing, and Northern blot analysis

To sequence *S. kawagutii* small RNA, total RNA was extracted from cells grown in exponential stage under the normal condition described above. Small RNA was selectively fractioned and purified using Zymo's Quick-RNA™ MiniPrep kit (Zymo Research Corporation, Irvine, CA 92614) according to the manufacturer's instructions. Libraries were constructed and sequenced using an Illumina HiSeq2000. Raw reads were

filtered by removing low-quality, 3' adapter null reads and those smaller than 18 nt, followed by the trimming of adapters. Reads were aligned to *S. kawagutii* genome using SOAP2 package (25) and those mapping to rRNA and tRNA were removed. A total of 354 *S. kawagutii* miRNA candidates were predicted from clean reads by miRDeep2 (44). The minimum free energy (MFE) of precursors were predicted using RNAfold (45) with 1000 bootstraps and a cutoff of P=0.05. Candidates whose precursors aligned to repeated regions in the genome were removed and the *bona fide* miRNA were selected following the three criteria proposed by Baumgarten et al. (2013) (46), (1) both the guide and passenger strands of the precursor were mapped by clean reads; (2) at least 2 nt dangling bases were present at the 3' terminus of the hairpin duplex; and (3) the MFE of the hairpin structure is less than -25 kcal mol⁻¹. As a result, 102 miRNAs were eventually retained. The predicted miRNAs were searched against miRBase (47) using the default settings of its web tool to identify potential homologs.

For RNA blot analysis, in one experiment, cells were harvested from the culture grown under the normal condition. In another, the cultures previously maintained under the normal condition were split and one group stayed at 25°C while the other transferred to 35°C (other conditions unchanged) for two weeks, and about 1 g of cell pellets from each condition were collected. Cell pellets were homogenized thoroughly for RNA isolation using Zymo Quick-RNA™ MiniPrep (Zymo Research) or mirVana kit that to isolate small RNAs. Cell pellets or RNA samples were delivered on dry ice to destination laboratories via FEDEX overnight service. Either 5 or 10 µg total RNA was loaded to each lane, and electrophoresis and blot hybridization were conducted as reported (28) in University of California at Los Angeles, University of Montreal, and Xiamen University. Six oligonucleotide sequences complementary to miRNA were used as probes, five of which were randomly picked from the list of highly expressed miRNA and another predicted to target heat shock proteins and many other targets (table S25, 26). For size estimation, tRNA (72 nt) or oligonucleotides were loaded in the gel, and a probe against dinoflagellate spliced leader precursor RNA (slrRNA; 55 nt) was used in hybridization.

Prediction and functional analyses of miRNA target genes

Plant type target genes of *S. kawagutii* were predicted by aligning mature miRNA sequences to the coding sequences of genes using the criteria suggested by Allen et al (2005) (48): (1) less than four mismatches between miRNA and target; (2) less than two adjacent mismatches in the miRNA/target duplex; (3) less than 2.5 mismatches between position 1-12 of the duplex, no adjacent mismatches in this window allowed; (4) no mismatch in position 10-11 of the duplex. The 3' UTR sequences of *S. kawagutii* and coral genes were used to predict animal type target genes by using the PITA package (49), and *Symbiodinium* candidates with $\Delta\Delta G$ less than -20 kcal mol⁻¹ were kept as possible targets, while a less stringent threshold ($\Delta\Delta G < -15$ kcal mol⁻¹) was used to predict potential target genes in the coral genome.

Networks of protein-protein interaction involving these target genes were constructed using Cytoscape (50) and clusters that focused on highly connected nodes were obtained using the molecular complex detection (MCODE) clustering algorithm (51). The networks of interacting genes were based on interactions identified in *Arabidopsis thaliana* and *Saccharomyces cerevisiae* using the STRING resource (52). The interactions include direct (physical) and indirect (functional) associations, which

were derived from i) known experimental interactions, ii) pathway knowledge, iii) automated text-mining to uncover statistical and/or semantic links between proteins, iv) interactions predicted by co-expression analysis, and v) interactions that are observed in one organism and systematically transferred to other organisms via pre-computed orthology relations.

Promoter motif identification

Promoter regions, defined to be either 1000 bp upstream of the translation start site, or to be the upstream intergenic sequences if it was shorter than 1000 bp, were used to discover motifs (4 to 10 bp in length) significantly enriched in this region. Two scores, Promoter Enrichment Score (PES) and Position Conservation Score (PCS), were used together to evaluate the conservation of promoter motifs. Assuming the distribution of any random oligomer in the promoter a binomial model, the PES, which was used to identify motifs enriched in the promoter regions, was calculated as:

$$PES_i = \frac{K_i - N_i P_0}{\sqrt{N_i P_0 (1 - P_0)}}$$

where N_i and K_i denote the occurrences of that oligomer in the *S. kawagutii* genome and promoter regions, respectively, and P_0 is the observed average probability of any tested oligomers locating at the promoter regions. Simple tandem repeat sequences longer than 1000 bp in the genome were masked because the real promoter motifs will be obscured if they were accidentally found in the tandem repeat regions. Oligomers with PES higher than 7 were retained and were subjected to an evaluation of the degree of positional conservation in the promoter.

The number of the occurrences (K_j) of each oligomer in a sliding window of 100 bp from the translation start site and the occurrence (N_j) in the whole promoter region were computed, and PCS was calculated as:

$$PCS_j = \frac{K_j - N_j P_k}{\sqrt{N_j P_k (1 - P_k)}}$$

where P_k is the average probability of any oligomer being located in each window (note that it is close but not equal to 0.1 since not all the promoters were 1000 bp long). Oligomers with consistent promoter positions ($PCS \geq 10$) were retained as conserved promoter motifs.

The selected promoter motifs were clustered according to their sequence similarities using the method described by Xie (53), and degenerate motifs were generated from the resulted clusters.

Gene Ontology categorization and enrichment analyses

Gene Ontology (GO) terms for each gene were obtained from its corresponding InterPro entry and perl script map2slim (<http://search.cpan.org/~cmungall/go-perl/scripts/map2slim>) was used to categorize them into GO slims. GO terms in which miRNA target genes or genes harboring different promoter motifs were significantly enriched were identified using a hypergeometric test with a P-value cutoff of 0.05. The

number of genes in a given GO term was counted and compared to the genome background by calculating P as:

$$P = 1 - \sum_{i=0}^{M-1} \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{n}}$$

where N denotes the number of all the genes with GO annotations in the genome, and M denotes the number of genes in a given GO term. The resulted P values were subjected to a Bonferroni correction to filter the potential false enrichments.

Telomere fluorescence in situ hybridization (FISH) and microscopic observation

A telomeric peptide nucleic acid probe ($[ccctaaa]_3$) labeled with Cy5 was synthesized by Panagene (Daejeon, Korea). Harvested cells were fixed in EtOH/acetic acid (3:1) overnight. Hybridization was carried out essentially as previously reported (54) with, as a slight modification in the denaturation step, the incubation of slides at 80°C for 10 min. Chromosomal DNA was stained with 4,6-diamidino-2-phenylindole (DAPI). Optical observations were performed using Zeiss MultiPhoto Laser Scanning Micro LSM780 NLO (Zeiss, Oberkochen, Germany), and bright-field and fluorescent images were processed with Zen2011 software. General morphology was observed using scanning electron microscopy, for which living cells were fixed in 1% OsO₄ (final concentration) for 2 hours and processed as described (55). Coated specimens were scanned under LEO 1530 Gemini scanning electron microscope (Zeiss, Oberkochen, Germany).

Subcellular localization prediction

Transporters in *S. kawagutii* potentially involved in cargo transport to or from its coral host were subjected to computational prediction of subcellular localization using YLoc+ model (56), under the version for plants. The genes predicted to be localized at plasma membrane, relevant to function in the host-symbiont exchange, are shown in table S35.

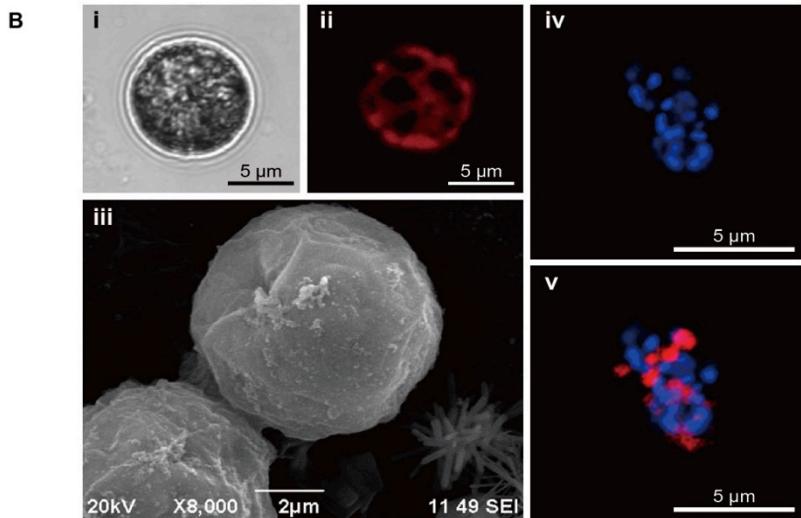
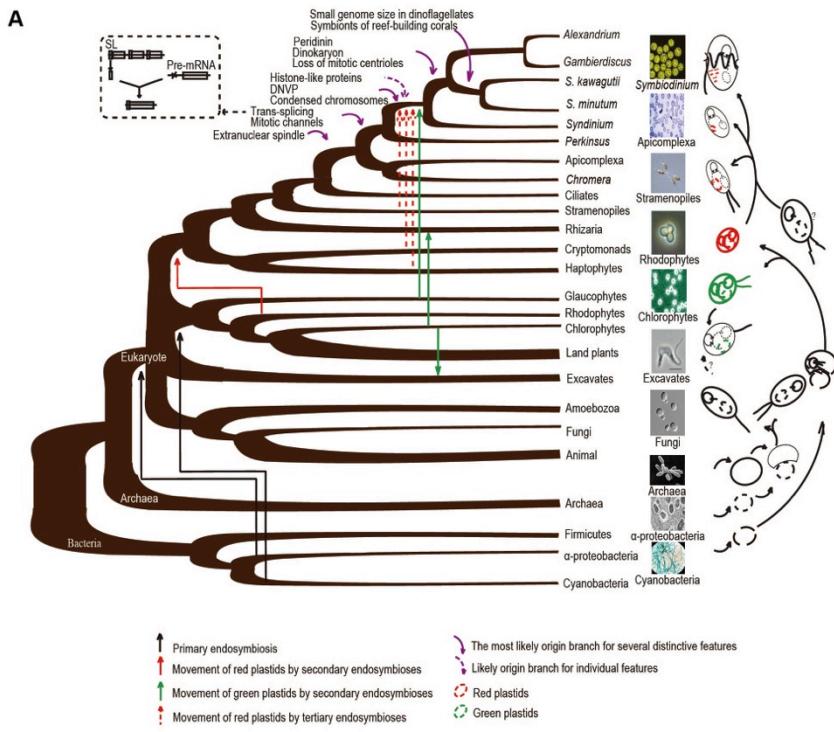


Fig. S1. Phylogenetic position and microscopic images of *S. kawagutii*. (A) A schematic diagram showing the phylogenetic position of *S. kawagutii* and innovative evolutionary events in dinoflagellates. The topology of the tree and the labeled unique characteristics were based on surveys of the literature and molecular evidence from our data analysis. (B) Images from bright-field light microscopy (i), fluorescence microscopy showing fluorescence of chlorophyll in red (ii), scanning electron microscopy (iii), fluorescence microscopy showing blue fluorescence of DAPI-stained condensed chromosomes in the nucleus (iv), and telomere hybridization (red) superimposed on DNA counterstain (blue; v).

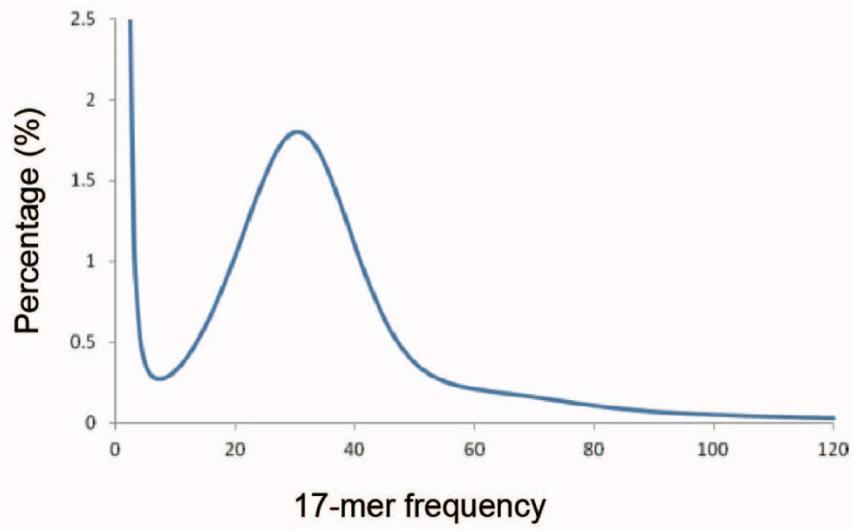


Fig. S2. Illumina 17-mer percentage histograms. The percentage of 17-mers is plotted against the frequency at which they occur. The truncated peak at low frequency and high percentage represents 17-mers containing essentially random sequencing errors, while the distribution to the right represents proper (putatively error-free) data. The genome size was estimated as (total K-mer number)/(the peak). Data can be found in table S2.

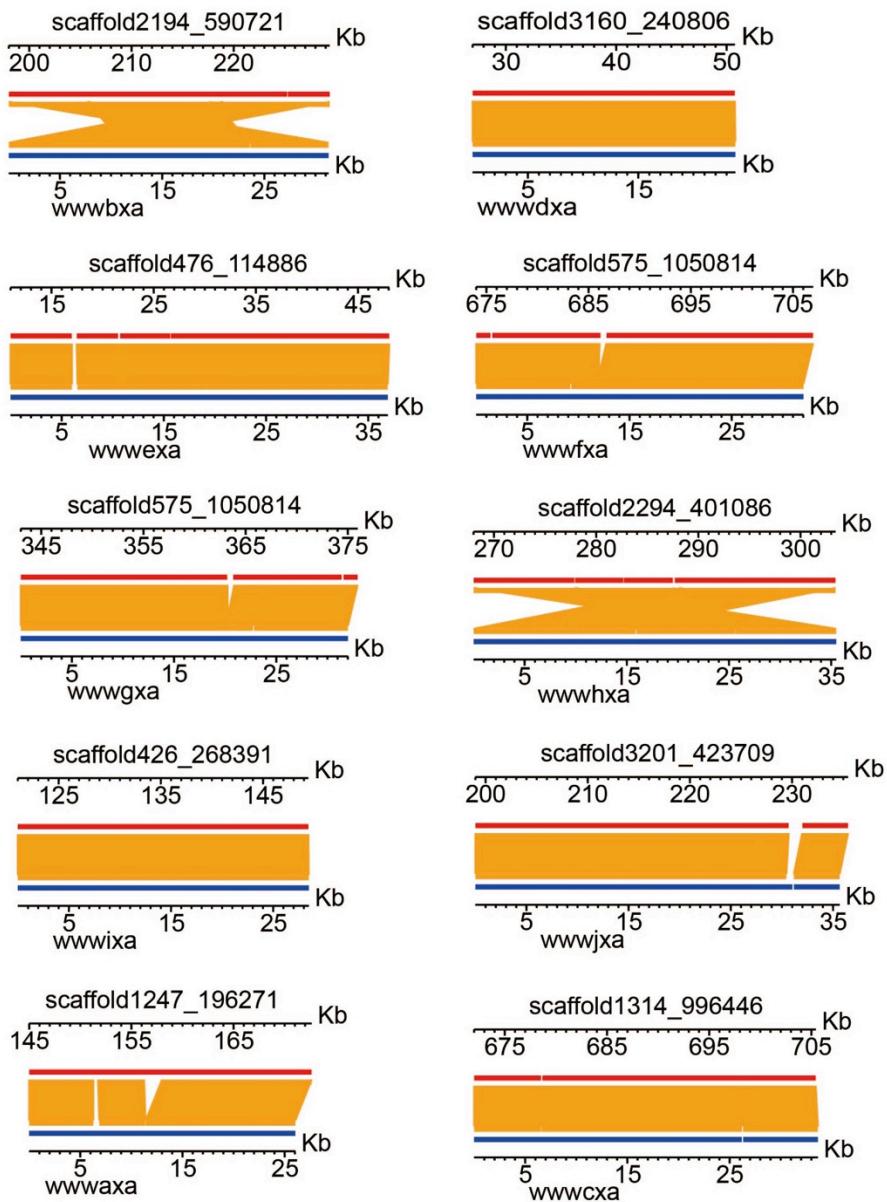


Fig. S3. Evaluation of the integrity of our genome assembly integrity by fosmid mapping. To assess the continuity and completeness of the assembly, 10 Sanger-derived fosmid sequences (blue lines, bottom) were mapped against the assembled scaffolds (red lines, top). Relative positions are indicated by the top and bottom scale marks. Matching positions in the scaffold and fosmid tracks are connected by yellow lines, with white gaps indicating the positions of nucleotide differences.

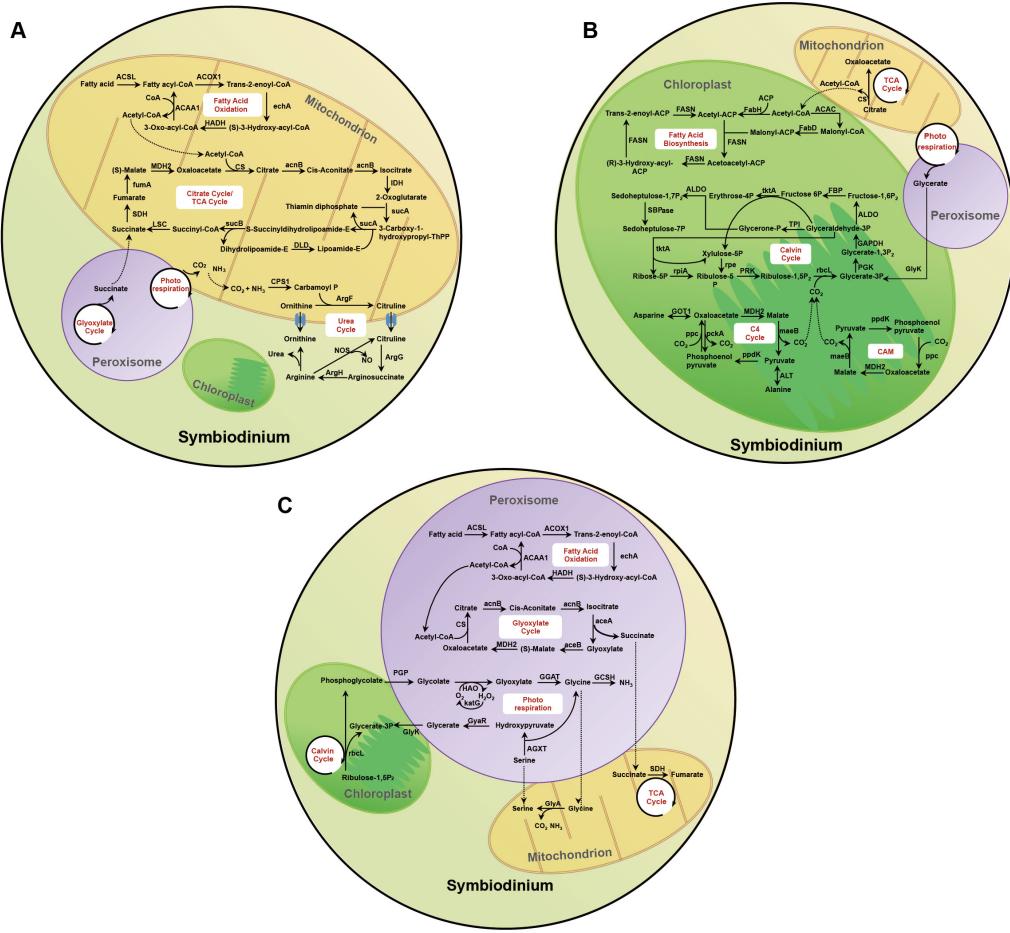


Fig. S4. Major metabolic pathways in the *S. kawagutii* genome. Metabolic pathways occurring in (A) mitochondria, (B) chloroplasts, and (C) peroxisomes are shown separately. Complete pathways for the mitochondrial TCA cycle, fatty acid oxidation and the urea cycle, the chloroplastic Calvin cycle, C4 cycle, CAM, and fatty acid biosynthesis and peroxisomal fatty acid oxidation, glyoxylate cycle and photorespiration are found in the *S. kawagutii* genome. *S. kawagutii* genes involved in these pathways are supplied in table S7.

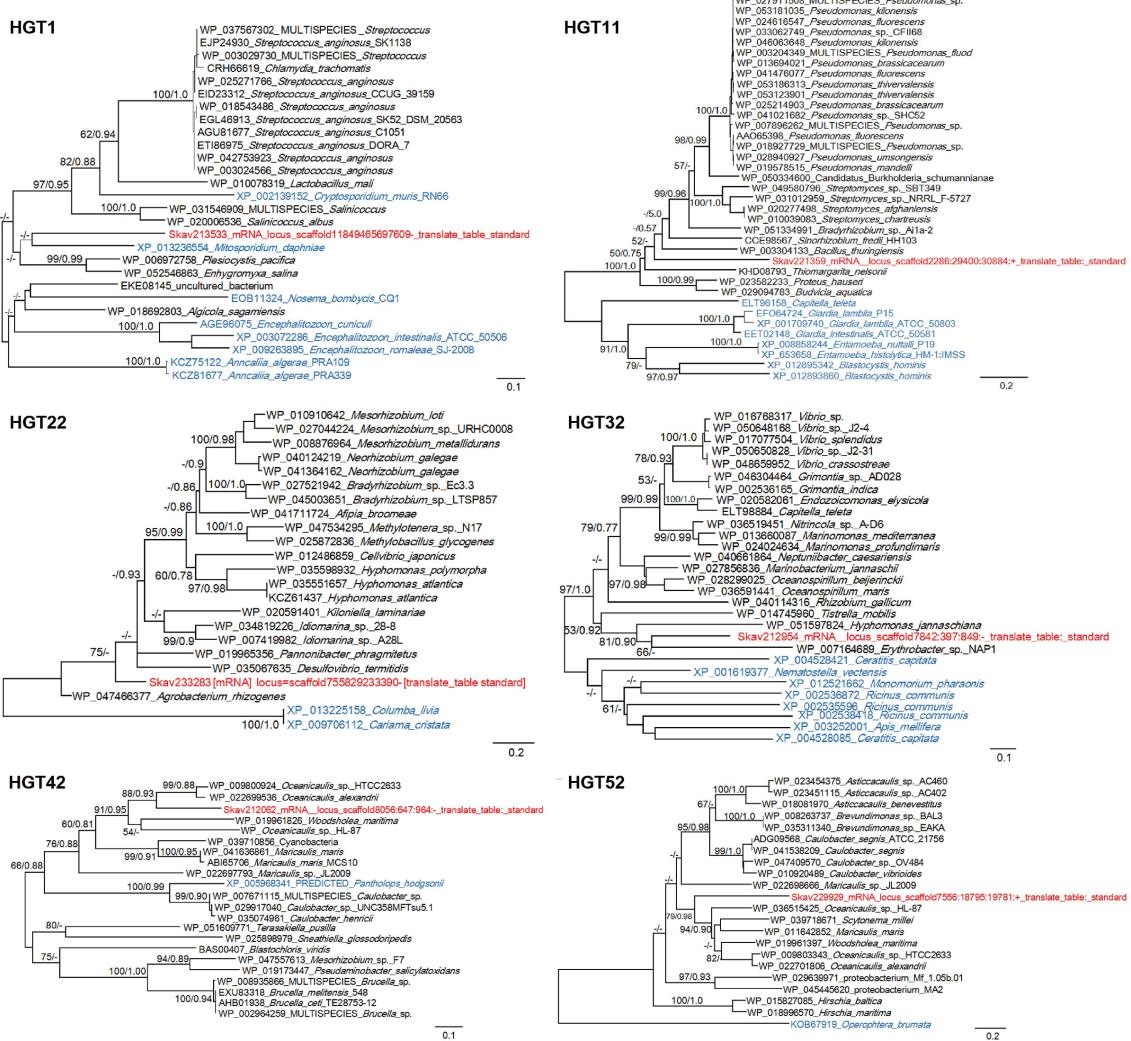


Fig. S5. Phylogenetic trees of putative horizontally transferred (HGT) genes inferred from amino sequences. Shown are a random selection of phylogenetic reconstructions (every 10th sequence in table S17). Support values for nodes in these Neighbor-Joining (NJ) tree topologies are from both NJ (left) and Maximum Likelihood analyses (right); "-" or blank indicates values below 50% (NJ) or 0.50 (ML). Scale bars depict substitution rates per amino acid residue. Besides the putative HGT sequences (in red), included in the analysis were about 30 top hit prokaryote sequences (in black) identified from BLASTP analysis, and one or more best hit eukaryotic sequences if present (in blue). Note that the HGT sequences are affiliated with bacteria in all cases except HGT1, which is in a bacteria-dominated clade but clustered most closely with *Mitosporidium daphiae*, a fungus-like parasite that lacks mitochondrial respiration. See more data information in table S17.

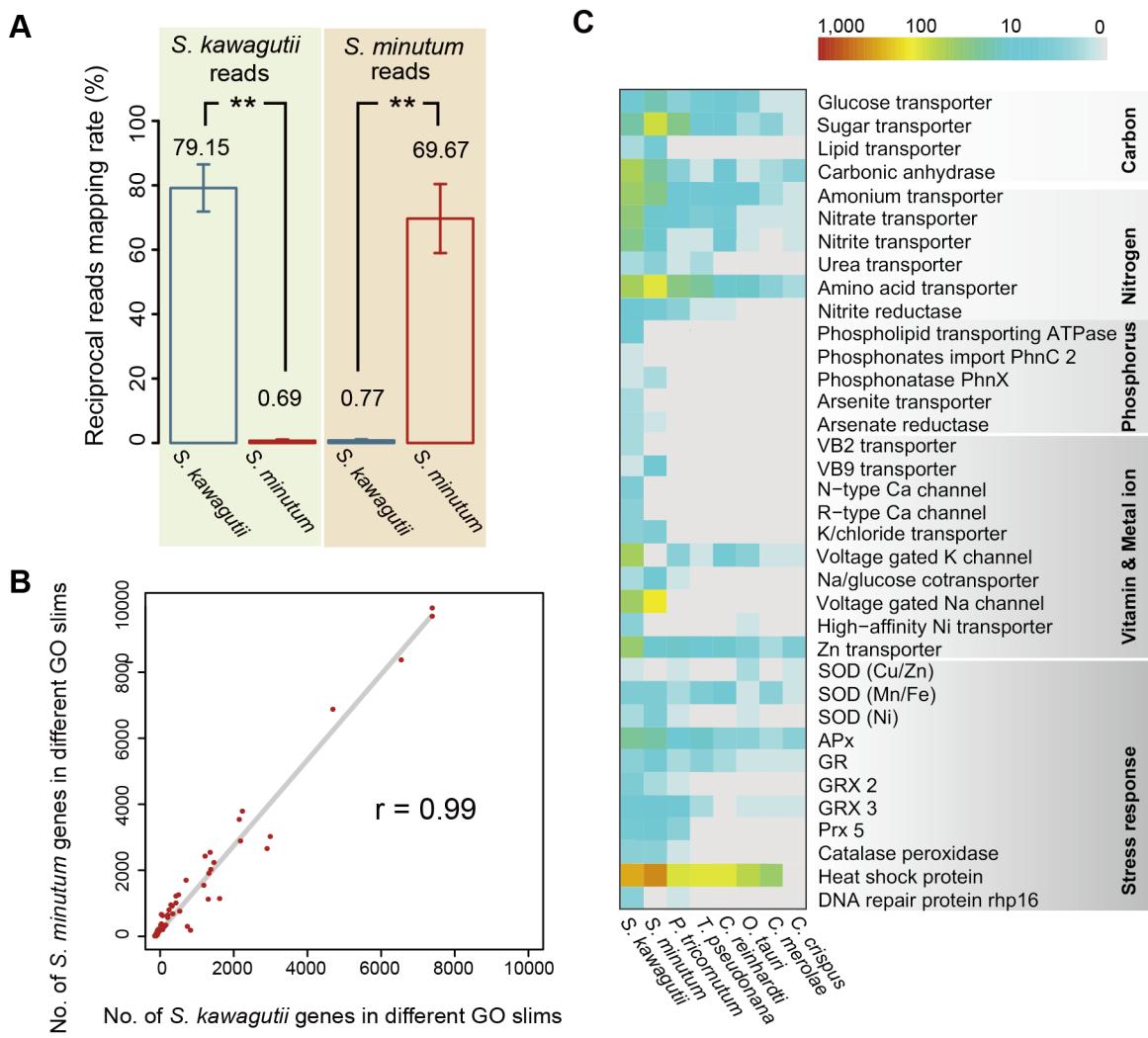


Fig. S6. Comparative genomic analysis between *S. kawagutii*, *S. minutum* and other eukaryotic alage. **(A)** Stark difference between the genomes of *S. kawagutii* and *S. minutum* revealed by reciprocal mapping of genome sequence reads to both assemblies. Values shown are percentages of the reads that have good match to the assembly. (**, p<0.01). **(B)** Similar gene coding regions between the two genomes indicated by the strong correlation between the number of genes in a given GO slim category for *S. kawagutii* (X axis) and *S. minutum* (Y axis). **(C)** Representative key gene families involved in material exchange and stress response expanded in *S. kawagutii* and *S. minutum*. Data details are shown in table S19, 20, 21, and 22.

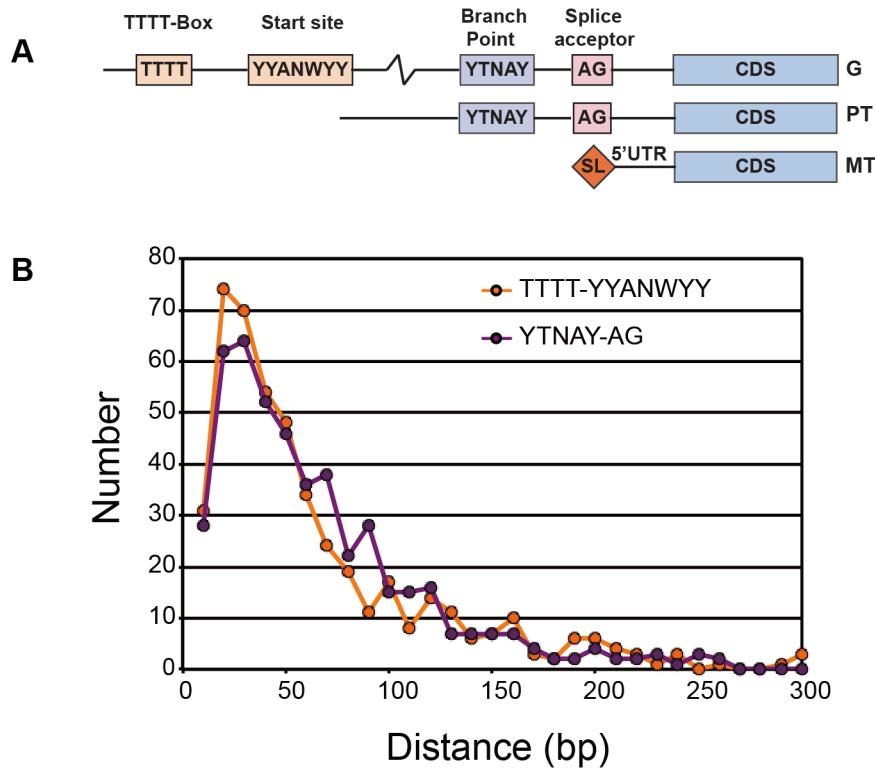


Fig. S7. Unique promoter architecture in *S. kawagutii*. (A) A schematic view of the proposed promoter (TTTT-box) relative to the putative transcription start site, splice branch point and acceptor site, upstream of the coding region (CDS) in the genomic sequence (G). For comparison, premature transcript (PT) and mature transcript (MT) are also shown. (B) The measured distances on the sense strand between the TTTT and the putative start site, as well as between the branch point and the splice acceptor for 494 genes whose AG splice site was unambiguously mapped to the genome.

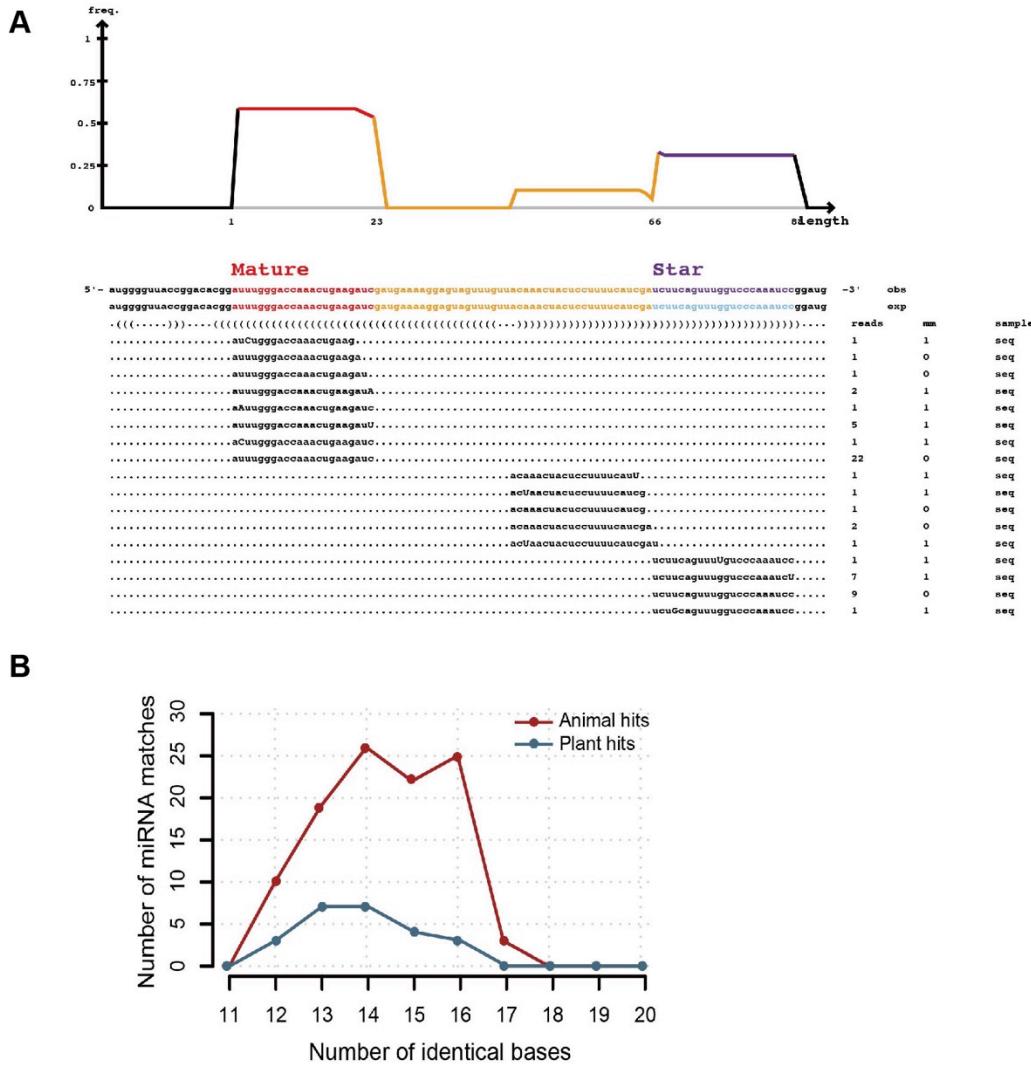


Fig. S8. (A) Alignment of sequenced small RNAs to the pre-miRNA sequence together with a graphical display of the number of small RNA reads corresponding to both miRNA and miRNA* sequences. Parentheses "(" or ")" indicate the folding direction of the given base in the hairpin structure; sequences below the bracket-containing line are the sequenced reads used to predict this miRNA. (B) Number of miRNA BLAST hits with a given number of bases identical to plant and animal miRNAs for *S. kawagutii* miRNAs.

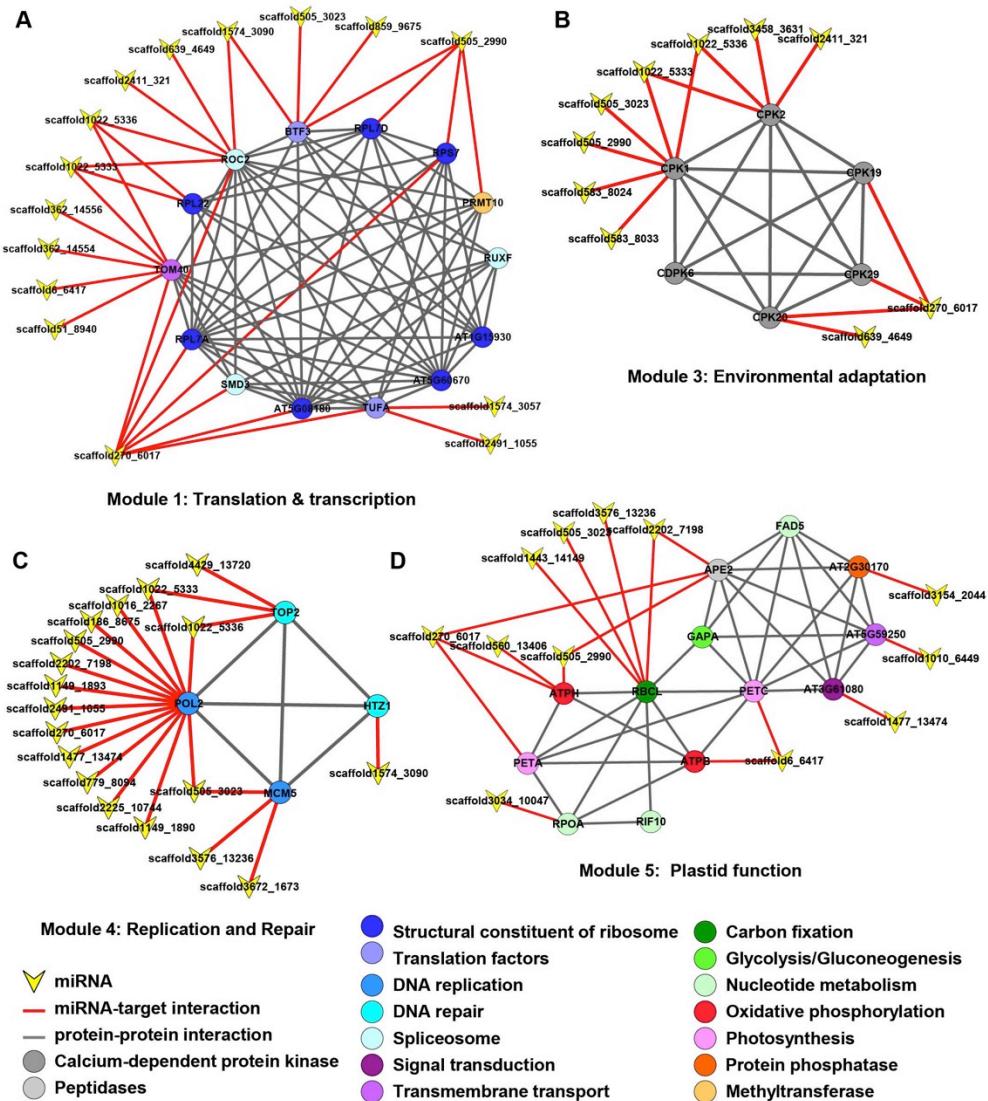


Fig. S9. Examples of networks of potential miRNA targets related to (A) translation and transcription (module 1); (B) environmental adaptation (module 3); (C) replication and repair (module 4); (D) plastid function (module 5). The networks were constructed using Cytoscape and clusters that focused on highly connected nodes obtained using the molecular complex detection (MCODE) clustering algorithm. The associated genes are displayed as nodes in a network, with connections between nodes representing miRNA-target interactions (red lines) or protein–protein interactions (grey lines). Node colors correspond to KEGG and GO functional annotations. More detailed information is available in table S28.

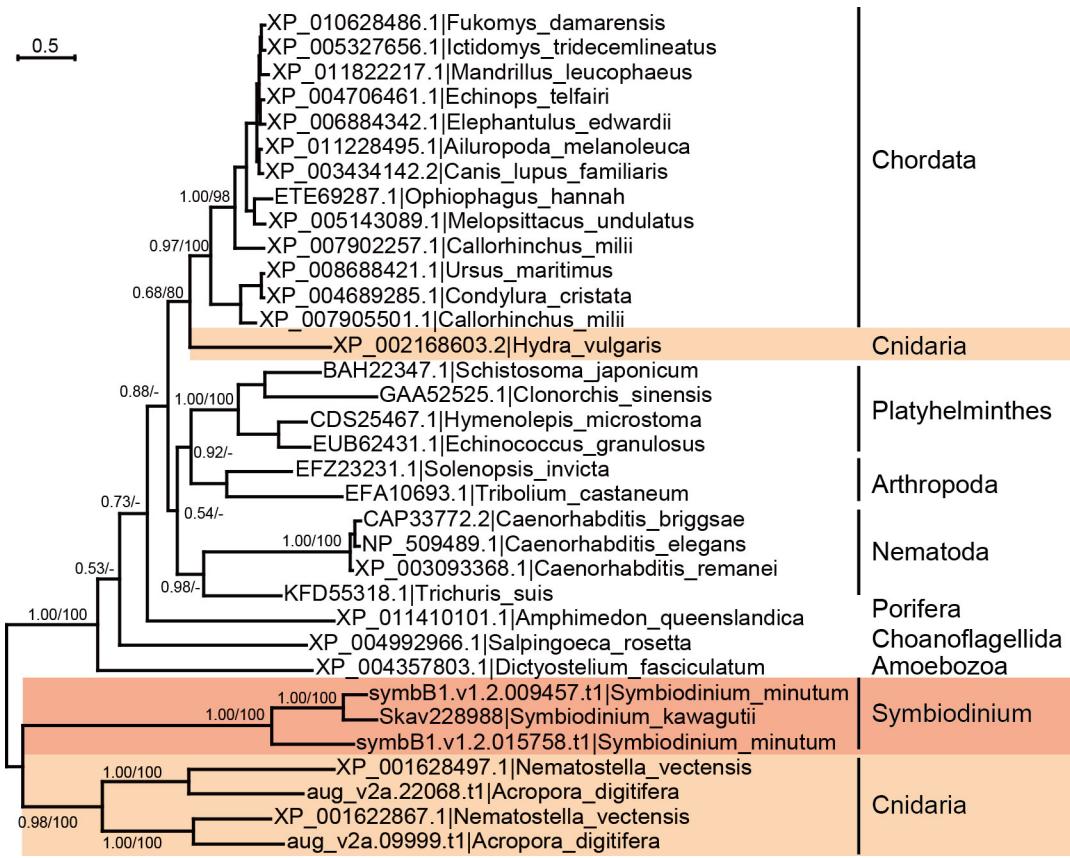


Fig. S10. Phylogenetic tree of double-stranded RNA channel (SID-1). Support values for nodes in the Maximum Likelihood (ML) tree topology are bootstrap values from ML (left) and Neighbor-Joining (NJ; right) analyses; only values above 0.50 (ML) or 50% (NJ) are shown. The close relationship between the homologs in *Symbiodinium* and the cnidarians suggests horizontal gene transfer between the two lineages. Scale bar depicts substitution rate per amino acid residue.

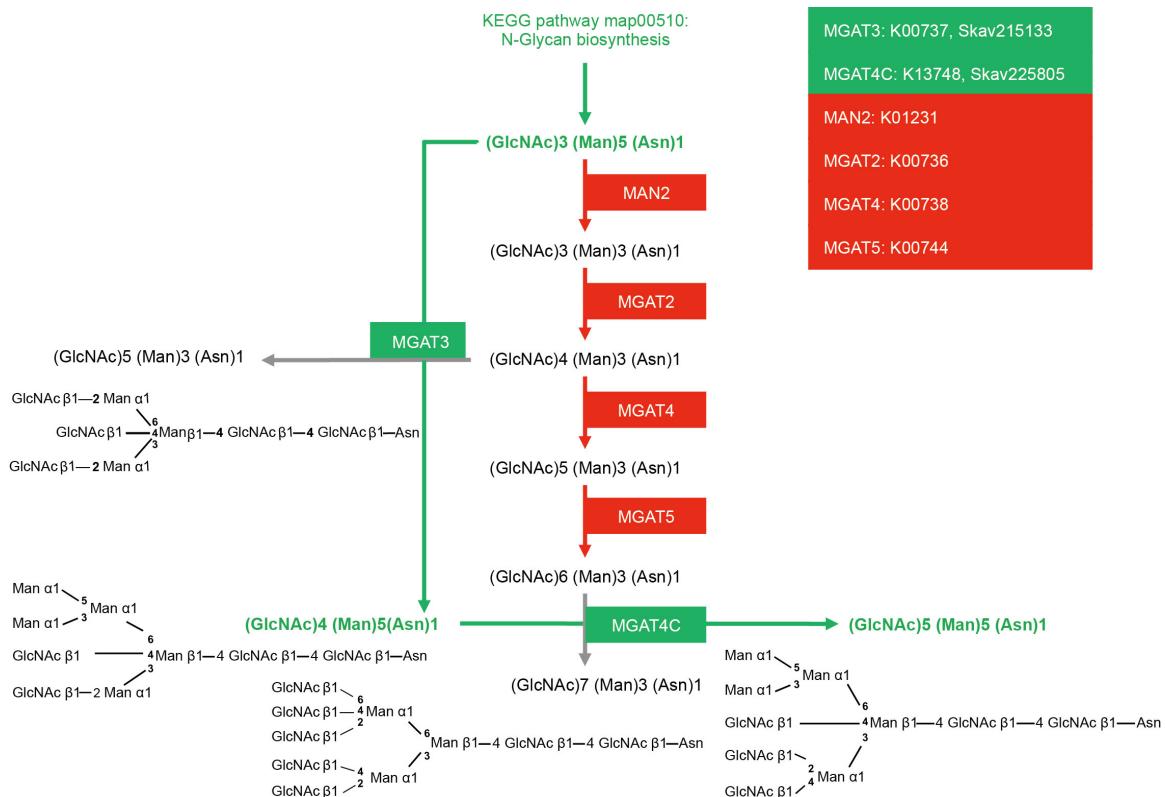


Fig. S11. Pathway of N-Glycan biosynthesis in *S. kawagutii*. Compared to the pathway in typical eukaryotes, four enzymes (MAN2, MFAT2, MGAT4, MGAT5) participating in the final steps of glycan biosynthesis (red) are missing in the *S. kawagutii* genome. As a result of catalysis by MGAT3 (Skav215133) and MGAT4C (Skav225805), the final product of the pathway will be (GlcNAc)5 (Man)5 (Asn)1, which carries mannose branches and man-man terminal linkages. The upstream pathway can be seen at the KEGG website (pathway: map00510). Detailed data sources are shown in table S31.

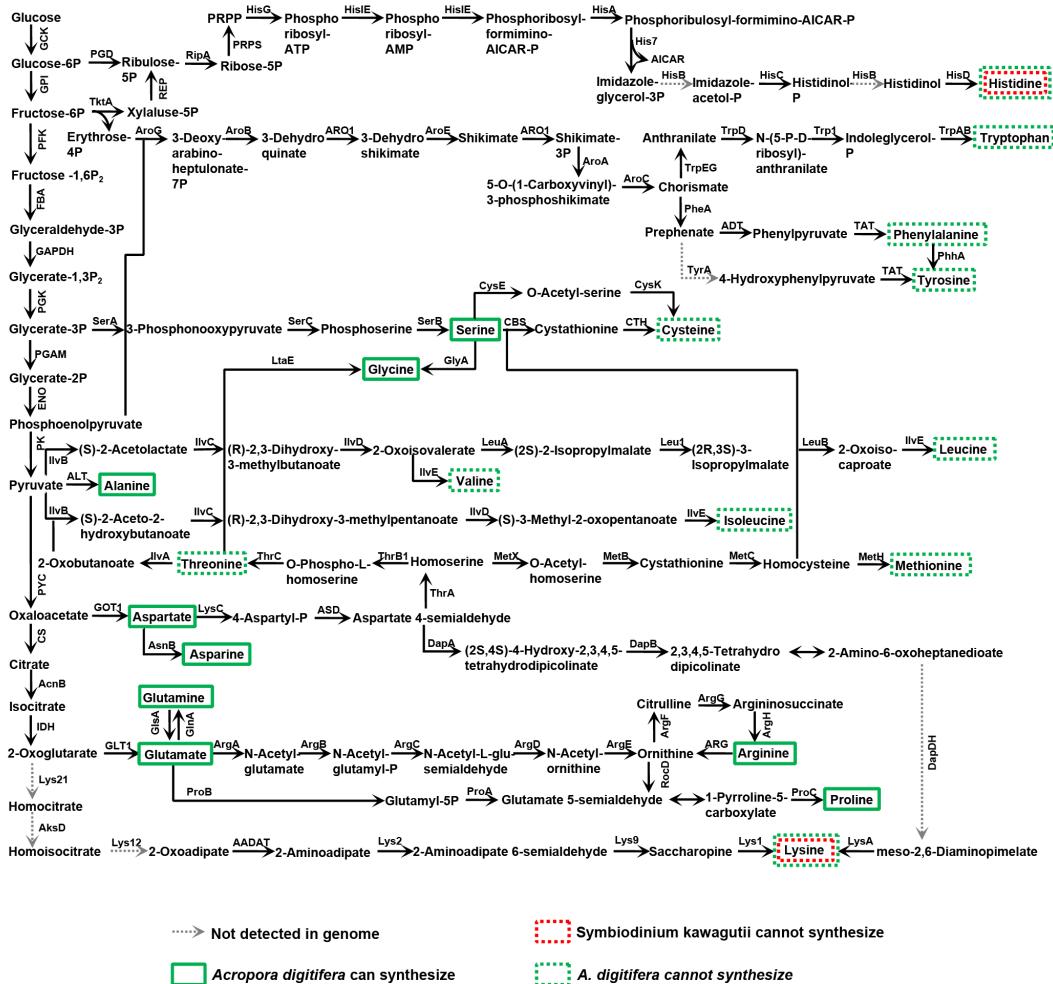


Fig. S12. *S. kawagutii* amino acid biosynthesis pathways. *S. kawagutii* can synthesize all 20 amino acids except for histidine and lysine. Gray dashed arrows indicate genes absent in both *S. kawagutii* and coral (*A. digitifera*) genomes. Predicted from the genome sequence, the coral *A. digitifera* can synthesize 9 (green solid line boxes) of the 20 amino acids and cannot synthesize the other 11 (green dashed line boxes). Only lysine and histidine cannot be synthesized by *S. kawagutii* (red dashed line boxes). *S. kawagutii* genes involved in these pathways can be found in table S7.

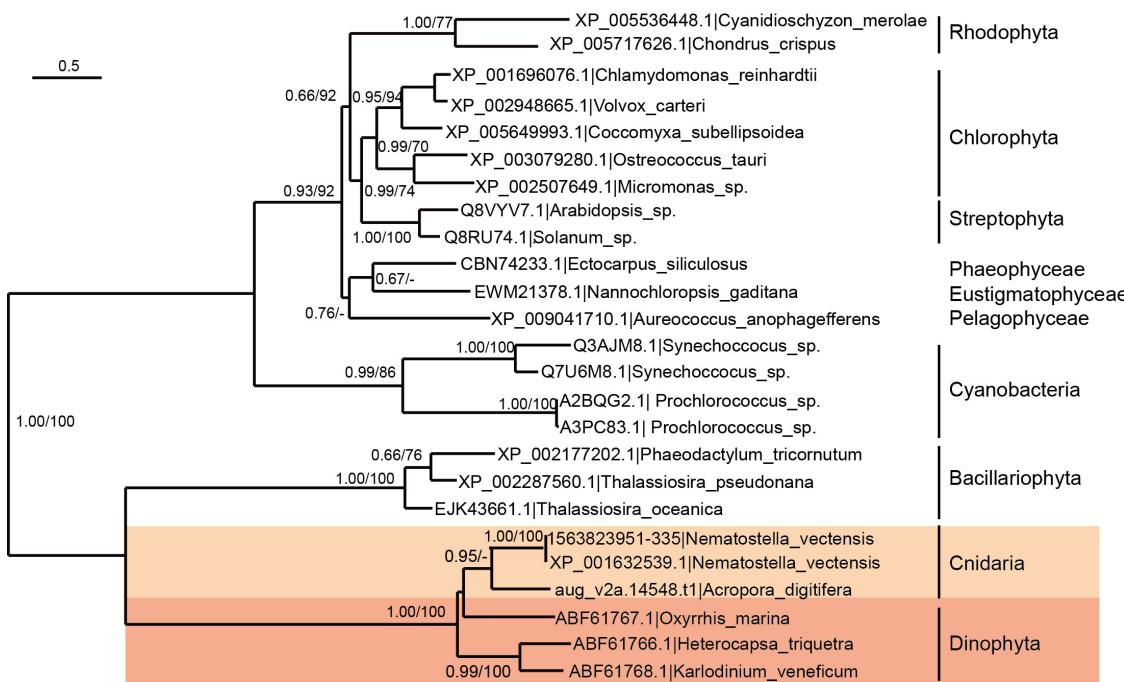


Fig. S13. Phylogenetic tree of 2-epi-5-epi-valiolone synthase gene. This is one of the four major MAA biosynthesis genes, dehydroquinate synthase-like. The close relationship between the cnidian and dinoflagellate sequences suggests horizontal gene transfer. Support values for nodes in the Maximum Likelihood (ML) tree topology are bootstrap values from ML (left) and Neighbor-Joining (NJ; right) analyses; only values above 0.50 (ML) or 50% (NJ) are shown. Scale bar depicts substitution rate per amino acid residue.

Supplementary Tables 1-35 (each in a separate Excel file)

Table S1. Features of sequencing libraries (insert size between 170bp and ~20Kb) and statistics of the sequencing data (post-filtering clean data). In total, 151 Gb clean data was produced with a resulting sequencing depth of 130X.

Table S2. Summary statistics of k-mer analysis and resulting genome size estimates in *S. kawagutii*.

Table S3. Genome assembly evaluation by mapping gene regions to transcriptome sequences. Transcriptome data (EST sequences) were collected from three cDNA libraries sequenced using Sanger (12,990 high-quality long reads) and 454 (834,546 high-quality reads) sequencing techniques, which were merged and clustered into a set of 70,985 unique genes. This table describes the alignment results between the 70,985 unigenes and the gene regions of our genome assembly.

Table S4. Genome assembly evaluation by mapping fosmid clones to scaffolds.

Table S5. General genome profile in *S. kawagutii* in comparison to the eukaryotes *S. minutum*, *P. falciparum*, *P. tricornutum* and *A. thaliana*.

Table S6. Tandem array genes and their InterPro functional annotation in the *S. kawagutii* genome.

Table S7. Enzymes and their corresponding gene IDs for the major metabolic pathways in the assembled *S. kawagutii* genome.

Table S8. Genes annotated as potentially involved in sexual reproduction, sporulation (cyst formation and germination) and telomere synthesis in the assembled *S. kawagutii* and *S. minutum* genomes (numbers indicate copy numbers).

Table S9. Gene families gained in the ancestor of *Symbiodinium*.

Table S10. The enriched ontologies of gene families gained in the ancestral *Symbiodinium*.

Table S11. Gene families that have shrunk in *Symbiodinium* spp.

Table S12. Gene families expanded in *Symbiodinium* spp.

Table S13. An example of duplicated genomic blocks within *S. kawagutii* genome and the associated functional annotation. Overall, duplicated segments within *S. kawagutii* are rare; here we give the most notable syntenic blocks in the genome detected by Mescan method (<http://chibba.agtec.uga.edu/duplication/mescan/>) using genes as anchors (-e_value: 1e-5; -match_size=5).

Table S14. Statistics of repeat elements annotated in *S. kawagutii* genome.

Table S15. Domains linked to transposable elements in *S. kawagutii* and six other species. All domain information was annotated by the Pfam package; here we focus on "Reverse transcriptase", "Integrase", "C-5 cytosine methyltransferase", "Zinc finger (Znf)", and their connections (links) within genes. *S. minutum*, *Symbiodinium minutum*; *P. marinus*, *Perkinsus marinus* (perkinsid); *P. falciparum*, *Plasmodium falciparum* (apicomplexan); *P. tetraurelia*, *Paramecium tetraurelia* (ciliate); *P. tricornutum*, *Phaeodactylum tricornutum* (diatom); *A. thaliana*, *Arabidopsis thaliana* (land plant).

Table S16. *S. kawagutii* genes containing a DinoSL (dinoflagellate-specific spliced leader) in their 5' UTR. In total, we identified 5,568 genes with partial or full SL. Here, we list the 62 genes with full length DinoSL at their 5' UTR, of which 28 (shaded by orange) contain canonical SL sequence while the others have SLs with 1-3 nucleotides variant.

Table S17. Putative horizontal gene transfer (HGT)-derived genes detected in the *S. kawagutii* genome. See Supplementary material for methodology used.

Table S18. Collinear blocks between *S. kawagutii* and *S. minutum* genomes identified by McScan (<http://chibba.agtec.uga.edu/duplication/mcscan/>) (-e_value: 1e-5; -match_size=5).

Table S19. Statistics of clean *S. kawagutii* reads aligned to two *Symbiodinium* genomes. PE, paired end; SE, single end.

Table S20. GO slim categorization of *S. kawagutii* and *S. minutum* genes.

Table S21. Summary of copy numbers for genes related to nutrient acquisition and utilization, and metabolite transport in *S. kawagutii* and other organisms. *P. tricornutum*, *Phaeodactylum tricornutum* (diatom); *T. pseudonana*, *Thalassiosira pseudonana* (diatom); *C. reinhardtii*, *Chlamydomonas reinhardtii* (green alga); *O. tauri*, *Ostreococcus tauri* (green alga); *C. merolae*, *Cyanidioschyzon merolae* (red alga); *C. crispus*, *Chondrus crispus* (red alga); *E. siliculosus*, *Ectocarpus siliculosus* (brown alga); *P. falciparum*, *Plasmodium falciparum* (apicomplexan); *A. thaliana*, *Arabidopsis thaliana* (land plant).

Table S22. Summary of copy numbers for genes related to antioxidant and anti-stress in *S. kawagutii* and other organisms. *P. tricornutum*, *Phaeodactylum tricornutum* (diatom); *T. pseudonana*, *Thalassiosira pseudonana* (diatom); *C. reinhardtii*, *Chlamydomonas reinhardtii* (green alga); *O. tauri*, *Ostreococcus tauri* (green alga); *C. merolae*, *Cyanidioschyzon merolae* (red alga); *C. crispus*, *Chondrus crispus* (red alga); *E. siliculosus*, *Ectocarpus siliculosus* (brown alga); *P. falciparum*, *Plasmodium falciparum* (apicomplexan); *A. thaliana*, *Arabidopsis thaliana* (land plant).

Table S23. Clusters of conserved promoter motifs. Symbols for degenerate positions are shown at right.

Table S24. Pre- and mature miRNA sequences (those highlighted in green were used to design probes for RNA blot hybridization).

Table S25. Target genes of miRNA in *S. kawagutii* genome (those targeted by highly expressed miRNAs are highlighted in yellow).

Table S26. The enriched ontologies of genes potentially targeted by miRNAs in the *S. kawagutii* genome.

Table S27. The enriched ontologies of genes potentially targeted by highly expressed miRNA in the *S. kawagutii* genome.

Table S28. Protein-protein interaction networks of *S. kawagutii* miRNA targets based on networks from *Arabidopsis thaliana* and *Saccharomyces cerevisiae*. In red are miRNA target genes detected in the coral *Acropora digitifera*.

Table S29. Genes in the coral *A. digitifera* (*Acropora digitifera*) computationally predicted to be targets of *S. kawagutii* miRNA.

Table S30. The enriched ontologies of genes potentially targeted by miRNAs in the *A. digitifera* genome.

Table S31. N-Glycan biosynthetic enzymes in the genomes of *Symbiodinium* and other algae as well as transcriptomes of two other dinoflagellates. *K. veneficum*, *Karlodinium veneficum* (dinoflagellate); *A. carterae*, *Amphidinium carterae* (dinoflagellate); *O. tauri*, *Ostreococcus tauri* (green alga); *P. tricornutum*, *Phaeodactylum tricornutum* (diatom); *M. pussila*, *Micromonas pussila* (green alga); *E. huxleyi*, *Emiliania huxleyi* (haptophyte).

Table S32. Genes possibly involved in symbiosis (numbers indicating copy numbers) unique to *Symbiodinium* (compared with nine other eukaryotes representing plants, algae, cnidian, and Apicomplexa) or only shared with the apicomplexan parasite *Plasmodium* (shaded in pink). Phosphoadenosine phosphosulfate reductase, calumenin and phosphatidylinositol-3-phosphate kinase are exceptions that are shared with several other organisms despite being previously proposed to be involved in symbiosis (Davy et al. 2012; ref 22). Species abbreviations: A_thaliana, *Arabidopsis thaliana* (land plant); C_crispus, *Chondrus crispus* (red macroalga); C_merolae, *Cyanidioschyzon merolae* (unicellular red alga); C_reinhardtii, *Chlamydomonas reinhardtii* (green alga); E_siliculosus, *Ectocarpus siliculosus* (brown alga); O_tauri, *Ostreococcus tauri* (green alga); P_falciparum, *Plasmodium falciparum* (apicomplexan parasite); P_tricornutum, *Phaeodactylum tricornutum* (diatom); T_pseudonana, *Thalassiosira pseudonana* (diatom); S_kawagutii, *Symbiodinium kawagutii*; S_minutum, *Symbiodinium minutum*.

Table S33. Major *S. kawagutii* and *A. digitifera* (coral) genes potentially involved in symbiosis and likely direction of action of the encoded products in the presumed pair of these two organisms.

Table S34. Transporters identified from *Symbiodium* genomes compared with those in related organisms. *S. kawagutii*, *Symbiodinium kawagutii*; *S. minutum*, *Symbiodinium minutum*; *A. digitifera*, *Acropora digitifera* (coral); *K. veneficum*, *Karlodinium veneficum* (dinoflagellate); *A. carterae*, *Amphidinium carterae* (dinoflagellate); *P. tricornutum*, *Phaeodactylum tricornutum* (diatom); *T. pseudonana*, *Thalassiosira pseudonana* (diatom).

Table S35. Genes in the genomes of both *Acropora digitifera* (coral) and *S. kawagutii* (symbiont) that can potentially be involved in symbiotic material exchange and collaborative combat against stressors. Comparison to the genomes of *Hydra magnipapillata* and *Nematostella vectensis* is also shown. The symbiont genes in red are computationally predicted to be plasma membrane proteins. Gene copies located next to one other in the genome are underlined. Light blue shading depicts absence of the gene.

References

1. A. C. Baker, Flexibility and specificity in coral-algal symbiosis: Diversity, ecology, and biogeography of *Symbiodinium*. *Annu. Rev. Ecol. Evol. Syst.* 34, 661–689 (2003). [doi:10.1146/annurev.ecolsys.34.011802.132417](https://doi.org/10.1146/annurev.ecolsys.34.011802.132417)
2. S. Lin, Genomic understanding of dinoflagellates. *Res. Microbiol.* 162, 551–569 (2011). [Medline doi:10.1016/j.resmic.2011.04.006](https://doi.org/10.1016/j.resmic.2011.04.006)
3. Materials and methods are available as supplementary materials on *Science Online*.
4. H. Zhang, Y. Hou, L. Miranda, D. A. Campbell, N. R. Sturm, T. Gaasterland, S. Lin, Spliced leader RNA trans-splicing in dinoflagellates. *Proc. Natl. Acad. Sci. U.S.A.* 104, 4618–4623 (2007). [Medline doi:10.1073/pnas.0700258104](https://doi.org/10.1073/pnas.0700258104)
5. C. H. Slamovits, P. J. Keeling, Widespread recycling of processed cDNAs in dinoflagellates. *Curr. Biol.* 18, R550–R552 (2008). [Medline doi:10.1016/j.cub.2008.04.054](https://doi.org/10.1016/j.cub.2008.04.054)
6. E. Shoguchi, C. Shizuno, T. Kawashima, F. Gyoja, S. Mungpakdee, R. Koyanagi, T. Takeuchi, K. Hisata, M. Tanaka, M. Fujiwara, M. Hamada, A. Seidi, M. Fujie, T. Usami, H. Goto, S. Yamasaki, N. Arakaki, Y. Suzuki, S. Sugano, A. Toyoda, Y. Kuroki, A. Fujiyama, M. Medina, M. A. Coffroth, D. Bhattacharya, N. Satoh, Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. *Curr. Biol.* 23, 1399–1408 (2013). [Medline doi:10.1016/j.cub.2013.05.062](https://doi.org/10.1016/j.cub.2013.05.062)
7. D. Guillebault, S. Sasorith, E. Derelle, J.-M. Wurtz, J.-C. Lozano, S. Bingham, L. Tora, H. Moreau, A new class of transcription initiation factors, intermediate between TATA box-binding proteins (TBPs) and TBP-like factors (TLFs), is present in the marine unicellular organism, the dinoflagellate *Cryptocodonium cohnii*. *J. Biol. Chem.* 277, 40881–40886 (2002). [Medline doi:10.1074/jbc.M205624200](https://doi.org/10.1074/jbc.M205624200)
8. S. Mi, T. Cai, Y. Hu, Y. Chen, E. Hodges, F. Ni, L. Wu, S. Li, H. Zhou, C. Long, S. Chen, G. J. Hannon, Y. Qi, Sorting of small RNAs into *Arabidopsis argonaute* complexes is directed by the 5' terminal nucleotide. *Cell* 133, 116–127 (2008). [Medline doi:10.1016/j.cell.2008.02.034](https://doi.org/10.1016/j.cell.2008.02.034)
9. J. D. Shih, C. P. Hunter, SID-1 is a dsRNA-selective dsRNA-gated channel. *RNA* 17, 1057–1065 (2011). [Medline doi:10.1261/rna.2596511](https://doi.org/10.1261/rna.2596511)
10. W. M. Winston, C. Molodowitch, C. P. Hunter, Systemic RNAi in *C. elegans* requires the putative transmembrane protein SID-1. *Science* 295, 2456–2459 (2002). [Medline doi:10.1126/science.1068836](https://doi.org/10.1126/science.1068836)
11. A. Weiberg, M. Wang, F.-M. Lin, H. Zhao, Z. Zhang, I. Kaloshian, H.-D. Huang, H. Jin, Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways. *Science* 342, 118–123 (2013). [Medline](https://doi.org/10.1126/science.1241036)
12. L. K. Bay, V. R. Cumbo, D. Abrego, J. T. Kool, T. D. Ainsworth, B. L. Willis, Infection dynamics vary between *Symbiodinium* types and cell surface treatments during establishment of endosymbiosis with coral larvae. *Diversity* 3, 356–374 (2011). [doi:10.3390/d3030356](https://doi.org/10.3390/d3030356)

13. E. M. Wood-Charlson, L. L. Hollingsworth, D. A. Krupp, V. M. Weis, Lectin/glycan interactions play a role in recognition in a coral/dinoflagellate symbiosis. *Cell. Microbiol.* 8, 1985–1993 (2006). [Medline](#) doi:10.1111/j.1462-5822.2006.00765.x
14. K.-L. Lin, J.-T. Wang, L.-S. Fang, Participation of glycoproteins on zooxanthellal cell walls in the establishment of a symbiotic relationship with the sea anemone, *Aiptasia pulchella*. *Zool. Stud.* 39, 172–178 (2000).
15. V. K. Goel, X. Li, H. Chen, S. C. Liu, A. H. Chishti, S. S. Oh, Band 3 is a host receptor binding merozoite surface protein 1 during the *Plasmodium falciparum* invasion of erythrocytes. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5164–5169 (2003). [Medline](#) doi:10.1073/pnas.0834959100
16. C. Shinzato, E. Shoguchi, T. Kawashima, M. Hamada, K. Hisata, M. Tanaka, M. Fujie, M. Fujiwara, R. Koyanagi, T. Ikuta, A. Fujiyama, D. J. Miller, N. Satoh, Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature* 476, 320–323 (2011). [Medline](#)
17. L. Muscatine, Glycerol excretion by symbiotic algae from corals and tridacna and its control by the host. *Science* 156, 516–519 (1967). [Medline](#) doi:10.1126/science.156.3774.516
18. M. S. Burriesci, T. K. Raab, J. R. Pringle, Evidence that glucose is the major transferred metabolite in dinoflagellate-cnidarian symbiosis. *J. Exp. Biol.* 215, 3467–3477 (2012). [Medline](#) doi:10.1242/jeb.070946
19. J. Albertyn, S. Hohmann, B. A. Prior, Characterization of the osmotic-stress response in *Saccharomyces cerevisiae*: Osmotic stress and glucose repression regulate glycerol-3-phosphate dehydrogenase independently. *Curr. Genet.* 25, 12–18 (1994). [Medline](#) doi:10.1007/BF00712960
20. A. B. Mayfield, R. D. Gates, Osmoregulation in anthozoan–dinoflagellate symbiosis. *Comparative Biochem. Physiol. A Mol. Integr. Physiol.* 147, 1–10 (2007). doi:10.1016/j.cbpa.2006.12.042
21. P. L. Clode, M. Saunders, G. Maker, M. Ludwig, C. A. Atkins, Uric acid deposits in symbiotic marine algae. *Plant Cell Environ.* 32, 170–177 (2009). [Medline](#) doi:10.1111/j.1365-3040.2008.01909.x
22. S. K. Davy, D. Allemand, V. M. Weis, Cell biology of cnidarian-dinoflagellate symbiosis. *Microbiol. Mol. Biol. Rev.* 76, 229–261 (2012). [Medline](#) doi:10.1128/MMBR.05014-11
23. T. C. LaJeunesse, D. J. Thornhill, E. F. Cox, F. G. Stanton, W. K. Fitt, G. W. Schmidt, High diversity and host specificity observed among symbiotic dinoflagellates in reef coral communities from Hawaii. *Coral Reefs* 23, 596–603 (2004). 10.1007/s00338-004-0428-4
24. H. Zhang, D. Bhattacharya, S. Lin, Phylogeny of dinoflagellates based on mitochondrial cytochrome b and nuclear small subunit rDNA sequence comparisons. *J. Phycol.* 41, 411–420 (2005). doi:10.1111/j.1529-8817.2005.04168.x
25. R. Luo, B. Liu, Y. Xie, Z. Li, W. Huang, J. Yuan, G. He, Y. Chen, Q. Pan, Y. Liu, J. Tang, G. Wu, H. Zhang, Y. Shi, Y. Liu, C. Yu, B. Wang, Y. Lu, C. Han, D. W. Cheung, S.-M. Yiu, S. Peng, Z. Xiaoqian, G. Liu, X. Liao, Y. Li, H. Yang, J. Wang, T.-W. Lam, J.

Wang, SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *GigaScience* 1, 18 (2012). [Medline](#) doi:10.1186/2047-217X-1-18

26. R. Li, W. Fan, G. Tian, H. Zhu, L. He, J. Cai, Q. Huang, Q. Cai, B. Li, Y. Bai, Z. Zhang, Y. Zhang, W. Wang, J. Li, F. Wei, H. Li, M. Jian, J. Li, Z. Zhang, R. Nielsen, D. Li, W. Gu, Z. Yang, Z. Xuan, O. A. Ryder, F. C.-C. Leung, Y. Zhou, J. Cao, X. Sun, Y. Fu, X. Fang, X. Guo, B. Wang, R. Hou, F. Shen, B. Mu, P. Ni, R. Lin, W. Qian, G. Wang, C. Yu, W. Nie, J. Wang, Z. Wu, H. Liang, J. Min, Q. Wu, S. Cheng, J. Ruan, M. Wang, Z. Shi, M. Wen, B. Liu, X. Ren, H. Zheng, D. Dong, K. Cook, G. Shan, H. Zhang, C. Kosiol, X. Xie, Z. Lu, H. Zheng, Y. Li, C. C. Steiner, T. T.-Y. Lam, S. Lin, Q. Zhang, G. Li, J. Tian, T. Gong, H. Liu, D. Zhang, L. Fang, C. Ye, J. Zhang, W. Hu, A. Xu, Y. Ren, G. Zhang, M. W. Bruford, Q. Li, L. Ma, Y. Guo, N. An, Y. Hu, Y. Zheng, Y. Shi, Z. Li, Q. Liu, Y. Chen, J. Zhao, N. Qu, S. Zhao, F. Tian, X. Wang, H. Wang, L. Xu, X. Liu, T. Vinar, Y. Wang, T.-W. Lam, S.-M. Yiu, S. Liu, H. Zhang, D. Li, Y. Huang, X. Wang, G. Yang, Z. Jiang, J. Wang, N. Qin, L. Li, J. Li, L. Bolund, K. Kristiansen, G. K.-S. Wong, M. Olson, X. Zhang, S. Li, H. Yang, J. Wang, J. Wang, The sequence and de novo assembly of the giant panda genome. *Nature* 463, 311–317 (2010). [Medline](#)
27. G. Zhang, X. Liu, Z. Quan, S. Cheng, X. Xu, S. Pan, M. Xie, P. Zeng, Z. Yue, W. Wang, Y. Tao, C. Bian, C. Han, Q. Xia, X. Peng, R. Cao, X. Yang, D. Zhan, J. Hu, Y. Zhang, H. Li, H. Li, N. Li, J. Wang, C. Wang, R. Wang, T. Guo, Y. Cai, C. Liu, H. Xiang, Q. Shi, P. Huang, Q. Chen, Y. Li, J. Wang, Z. Zhao, J. Wang, Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nat. Biotechnol.* 30, 549–554 (2012). [Medline](#) doi:10.1038/nbt.2195
28. H. Zhang, Y. Zhuang, J. Gill, S. Lin, Proof that dinoflagellate spliced leader (DinoSL) is a useful hook for fishing dinoflagellate transcripts from mixed microbial samples: *Symbiodinium kawagutii* as a case study. *Protist* 164, 510–527 (2013). [Medline](#) doi:10.1016/j.protis.2013.04.002
29. Y. Zhuang, H. Zhang, L. Hannick, S. Lin, Metatranscriptome profiling reveals versatile N-nutrient utilization, CO₂ limitation, oxidative stress, and active toxin production in an *Alexandrium fundyense* bloom. *Harmful Algae* 42, 60–70 (2015). doi:10.1016/j.hal.2014.12.006
30. M. Stanke, S. Waack, Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19 (Suppl 2), ii215–ii225 (2003). [Medline](#) doi:10.1093/bioinformatics/btg1080
31. W. J. Kent, BLAT—the BLAST-like alignment tool. *Genome Res.* 12, 656–664 (2002). [Medline](#) doi:10.1101/gr.229202. Article published online before March 2002
32. B. J. Haas, A. L. Delcher, S. M. Mount, J. R. Wortman, R. K. Smith Jr., L. I. Hannick, R. Maiti, C. M. Ronning, D. B. Rusch, C. D. Town, S. L. Salzberg, O. White, Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* 31, 5654–5666 (2003). [Medline](#) doi:10.1093/nar/gkg770
33. C. G. Elsik, A. J. Mackey, J. T. Reese, N. V. Milshina, D. S. Roos, G. M. Weinstock, Creating a honey bee consensus gene set. *Genome Biol.* 8, R13 (2007). 10.1186/gb-2007-8-1-r13 [Medline](#) doi:10.1186/gb-2007-8-1-r13

34. M. Tarailo-Graovac, N. Chen, Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* 25, 4.10.1–4.10.14 (2009). [Medline](#)
35. J. Jurka, V. V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany, J. Walichiewicz, Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* 110, 462–467 (2005). [Medline](#) doi:10.1159/000084979
36. Z. Xu, H. Wang, LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35 (Web Server), W265–W268 (2007). [Medline](#) doi:10.1093/nar/gkm286
37. A. Smit, R. Hubley, RepeatModeler-1.0. 5. Institute for Systems Biology. <http://www.repeatmasker.org/RepeatModeler.html> (6 June 2011, date last accessed) (2012).
38. G. Benson, Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580 (1999). [Medline](#) doi:10.1093/nar/27.2.573
39. H. Tang, J. E. Bowers, X. Wang, R. Ming, M. Alam, A. H. Paterson, Synteny and collinearity in plant genomes. *Science* 320, 486–488 (2008). [Medline](#) doi:10.1126/science.1153917
40. A. L. Delcher, S. Kasif, R. D. Fleischmann, J. Peterson, O. White, S. L. Salzberg, Alignment of whole genomes. *Nucleic Acids Res.* 27, 2369–2376 (1999). [Medline](#) doi:10.1093/nar/27.11.2369
41. A. L. Delcher, A. Phillippy, J. Carlton, S. L. Salzberg, Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res.* 30, 2478–2483 (2002). [Medline](#) doi:10.1093/nar/30.11.2478
42. Y. Zheng, L. Zhao, J. Gao, Z. Fei, iAssembler: A package for de novo assembly of Roche-454/Sanger transcriptome sequences. *BMC Bioinformatics* 12, 453 (2011). [Medline](#) doi:10.1186/1471-2105-12-453
43. L. Li, C. J. Stoeckert, D. S. Roos, OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189 (2003). doi:10.1101/gr.1224503
44. M. R. Friedländer, S. D. Mackowiak, N. Li, W. Chen, N. Rajewsky, miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40, 37–52 (2012). [Medline](#)
45. R. Lorenz, S. H. Bernhart, C. Höner Zu Siederdissen, H. Tafer, C. Flamm, P. F. Stadler, I. L. Hofacker, ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 6, 26 (2011). [Medline](#)
46. S. Baumgarten, T. Bayer, M. Aranda, Y. J. Liew, A. Carr, G. Micklem, C. R. Voolstra, Integrating microRNA and mRNA expression profiling in *Symbiodinium microadriaticum*, a dinoflagellate symbiont of reef-building corals. *BMC Genomics* 14, 704 (2013). [Medline](#) doi:10.1186/1471-2164-14-704
47. A. Kozomara, S. Griffiths-Jones, miRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42 (D1), D68–D73 (2014). [Medline](#) doi:10.1093/nar/gkt1181

48. E. Allen, Z. Xie, A. M. Gustafson, J. C. Carrington, microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* 121, 207–221 (2005). [Medline](#)
[doi:10.1016/j.cell.2005.04.004](https://doi.org/10.1016/j.cell.2005.04.004)
49. M. Kertesz, N. Iovino, U. Unnerstall, U. Gaul, E. Segal, The role of site accessibility in microRNA target recognition. *Nat. Genet.* 39, 1278–1284 (2007). [Medline](#)
[doi:10.1038/ng2135](https://doi.org/10.1038/ng2135)
50. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504 (2003). [Medline](#)
[doi:10.1101/gr.1239303](https://doi.org/10.1101/gr.1239303)
51. G. D. Bader, C. W. Hogue, An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4, 2 (2003). [Medline](#)
[doi:10.1186/1471-2105-4-2](https://doi.org/10.1186/1471-2105-4-2)
52. D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K. P. Tsafou, M. Kuhn, P. Bork, L. J. Jensen, C. von Mering, STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43 (D1), D447–D452 (2015). [Medline](#) doi:[10.1093/nar/gku1003](https://doi.org/10.1093/nar/gku1003)
53. X. Xie, J. Lu, E. J. Kulkos, T. R. Golub, V. Mootha, K. Lindblad-Toh, E. S. Lander, M. Kellis, Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* 434, 338–345 (2005). [Medline](#)
[doi:10.1038/nature03441](https://doi.org/10.1038/nature03441)
54. M. Fojtová, J. T. Wong, M. Dvorácková, K. T. Yan, E. Sýkorová, J. Fajkus, Telomere maintenance in liquid crystalline chromosomes of dinoflagellates. *Chromosoma* 119, 485–493 (2010). [Medline](#) doi:[10.1007/s00412-010-0272-y](https://doi.org/10.1007/s00412-010-0272-y)
55. Y. Baud, D. Guillebault, J. Lennon, H. Defacque, M. O. Soyer-Gobillard, H. Moreau, Morphology and behaviour of dinoflagellate chromosomes during the cell cycle and mitosis. *J. Cell Sci.* 113, 1231–1239 (2000). [Medline](#)
56. S. Briesemeister, J. Rahnenführer, O. Kohlbacher, YLoc—an interpretable web server for predicting subcellular localization. *Nucleic Acids Res.* 38 (Web Server), W497–W502 (2010). [Medline](#) doi:[10.1093/nar/gkq477](https://doi.org/10.1093/nar/gkq477)