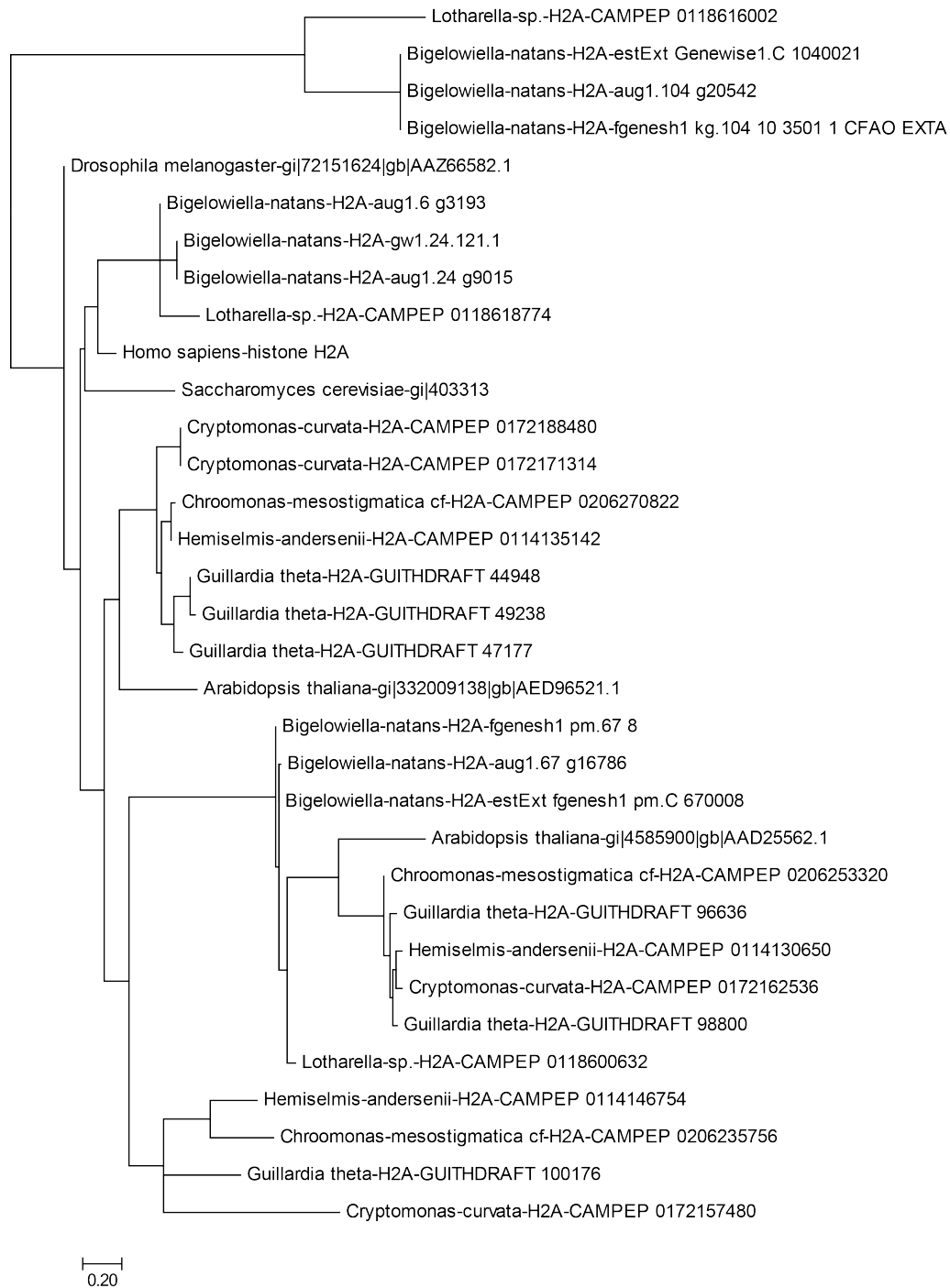
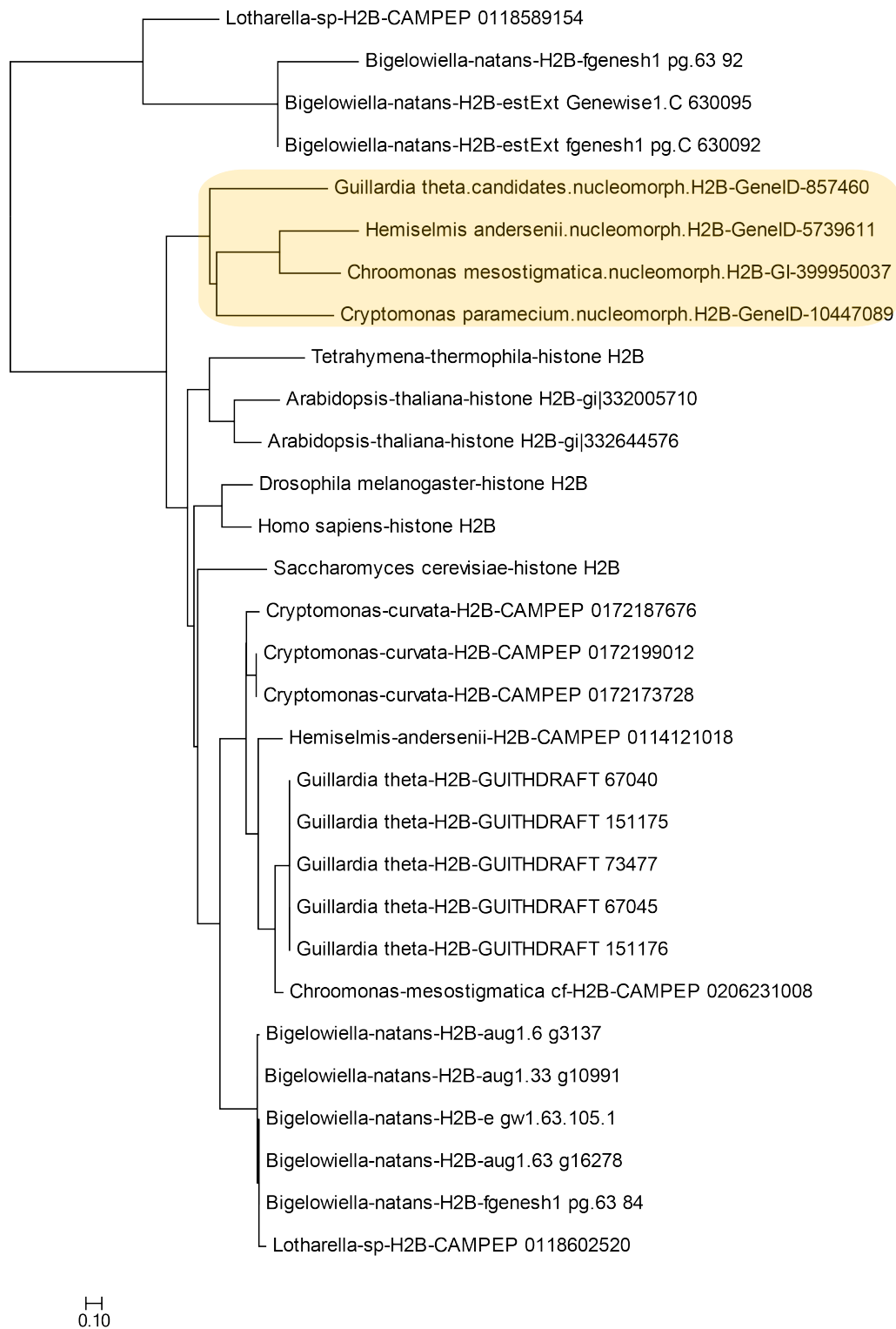


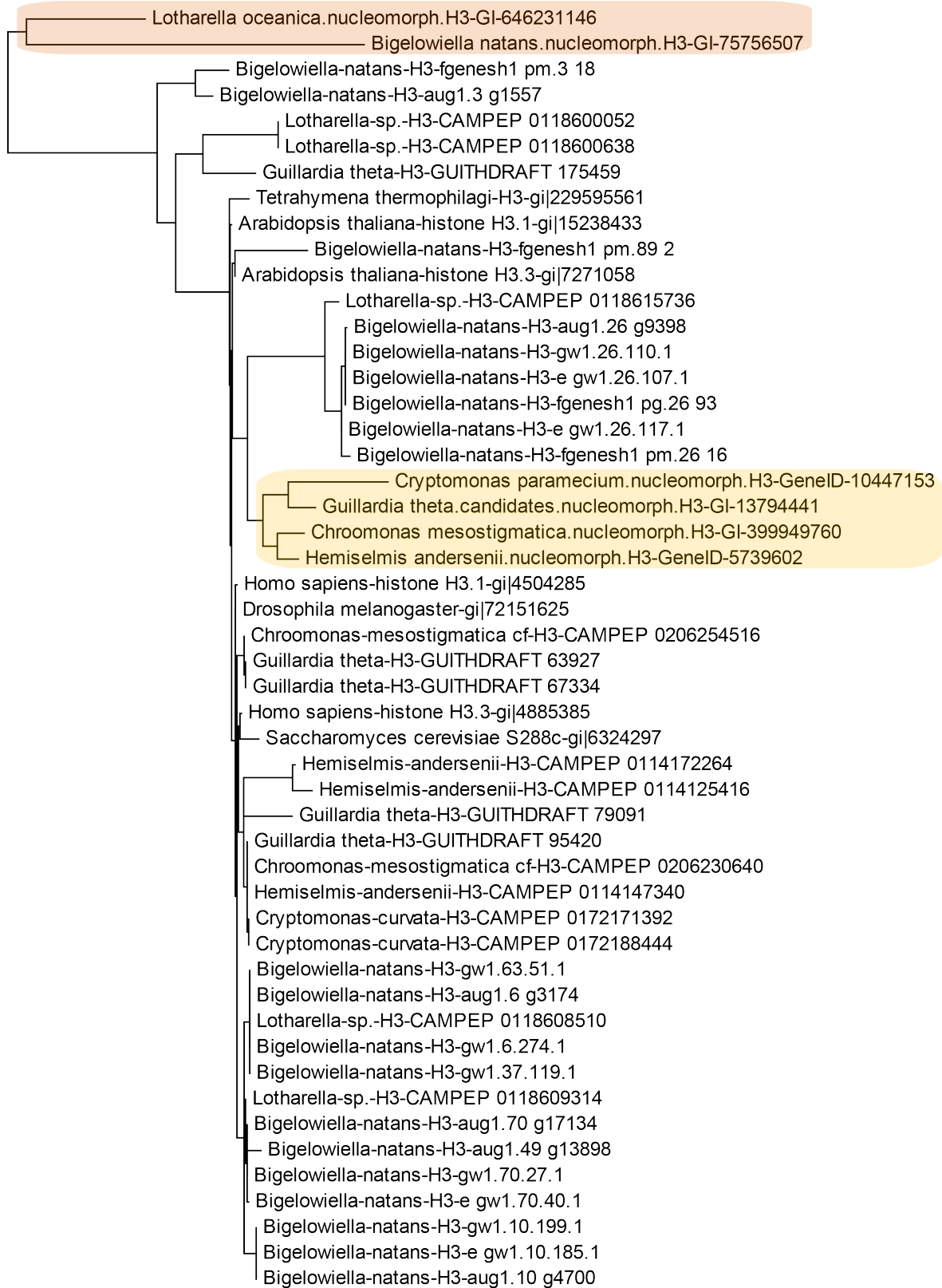
## Supplementary Figures



**Supplementary Figure 1: Maximum likelihood phylogenetic tree of all nucleus-coded H2A protein variants in cryptophytes and chlorarachniophytes.** The tree was generated using MUSCLE (Edgar 2004) multiple sequence alignments and RAxML (Stamatakis 2014; version 8.0.26) under the LG+G model and with 100 bootstrap replicates, and visualized using MEGA (Tamura et al. 2013). H2A sequences were obtained from available genome assemblies for *Bigelowiella natans* and *Guillardia theta* and from transcriptome assemblies for *Lotharella* sp., *Hemiselmis andersenii*, *Cryptomonas paramecium*, and *Chroomonas mesostigmatica* provided by the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) (Keeling et al. 2014) (after filtering out incompletely assembled proteins). Canonical H2A sequences from several other eukaryotes were also included.

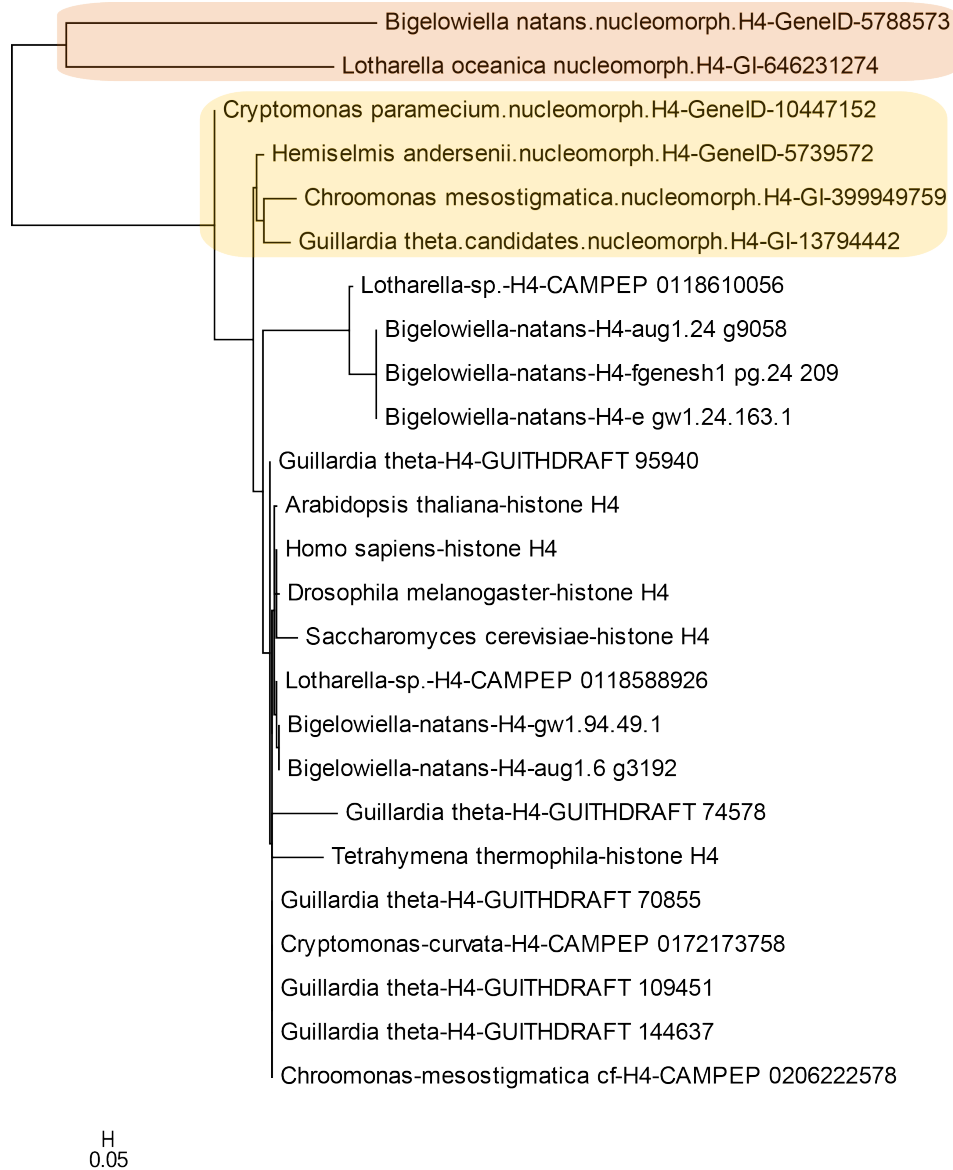


**Supplementary Figure 2: Maximum likelihood phylogenetic tree of all nucleus- and nucleomorph-coded H2B protein variants in cryptophytes and chlorarachniophytes.** The tree was generated using MUSCLE (Edgar 2004) multiple sequence alignments and RAxML (Stamatakis 2014; version 8.0.26) under the LG+G model and with 100 bootstrap replicates, and visualized using MEGA (Tamura et al. 2013). H2B sequences were obtained from nucleomorph genome assemblies, from available genome assemblies for *Bigelowiella natans* and *Guillardia theta* and from transcriptome assemblies for *Lotharella* sp., *Hemiselmis andersenii*, *Cryptomonas paramecium*, and *Chroomonas mesostigmatica* provided by the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) (Keeling et al. 2014) (after filtering out incompletely assembled proteins). Canonical H2A sequences from several other eukaryotes were also included.



H  
0.05

**Supplementary Figure 3 (preceding page): Maximum likelihood phylogenetic tree of all nucleus- and nucleomorph-coded H3 protein variants in cryptophytes and chlorarachniophytes.** The tree was generated using MUSCLE (Edgar 2004) multiple sequence alignments and RAxML (Stamatakis 2014; version 8.0.26) under the LG+G model and with 100 bootstrap replicates, and visualized using MEGA (Tamura et al. 2013). H3 sequences were obtained from nucleomorph genome assemblies, from available genome assemblies for *Bigelowiella natans* and *Guillardia theta* and from transcriptome assemblies for *Lotharella* sp., *Hemiselms andersenii*, *Cryptomonas paramecium*, and *Chroomonas mesostigmatica* provided by the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) (Keeling et al. 2014) (after filtering out incompletely assembled proteins). Canonical H2A sequences from several other eukaryotes were also included.



**Supplementary Figure 4: Maximum likelihood phylogenetic tree of all nucleus- and nucleomorph-coded H4 protein variants in cryptophytes and chlorarachniophytes.** The tree was generated using MUSCLE (Edgar 2004) multiple sequence alignments and RAxML (Stamatakis 2014; version 8.0.26) under the LG+G model and with 100 bootstrap replicates, and visualized using MEGA (Tamura et al. 2013). H4 sequences were obtained from nucleomorph genome assemblies, from available genome assemblies for *Bigelowiella natans* and *Guillardia theta* and from transcriptome assemblies for *Lotharella* sp., *Hemiselms andersenii*, *Cryptomonas paramecium*, and *Chroomonas mesostigmatica* provided by the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) (Keeling et al. 2014) (after filtering out incompletely assembled proteins). Canonical H2A sequences from several other eukaryotes were also included.



**Supplementary Figure 5: Alignment of cryptophyte nucleomorph H2B histones and the nuclear H2B histones from several representative eukaryotes. (A) *Chroomonas mesostigmatica* histone H2B; (B) *Cryptomonas paramecium* histone H2B; (C) *Guillardia theta* histone H2B; (D) *Hemiselmis andersenii* histone H2B.**

<b>H2B</b>	<i>Chroomonas mesostigmatica</i>			<i>Cryptomonas paramecium</i>			<i>Guillardia theta</i>			<i>Hemiselmis andersenii</i>			
	radius:	0	1	2	0	1	2	0	1	2	0	1	2
K 5	0	0	0	0	0	0	0	0	0	0	0	0	0
K 12	0	0	0	0	0	0	0	0	0	0	0	0	0
K 15	0	0	0	0	0	0	0	0	0	0	0	0	0
K 20	0	0	0	1	0	0	0	0	0	0	1	0	0
K 116	1	1	1	1	1	1	1	1	1	1	1	1	1
K 120	1	0	0	1	0	0	1	0	0	1	0	0	0

**Supplementary Figure 6: Conservation of key posttranscriptionally modified residues in the cryptophyte nucleomorph histone H2B.** The radius  $r$  refers to the size of the context considered when scoring conservation. When  $r = 1$ , only the residue itself is considered; when  $r = 1$ , a perfect match to the three-amino acid peptide also including the flanking residues on each side is required; when  $r = 2$ , the five-amino acid peptide also including the two flanking residues on each side is considered. A score of 1 means conservation according to these criteria, while a score of 0 means lack of conservation. Conservation was scored against the *Homo sapiens* histone H2B sequence.







## Supplementary References

- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**(5):1792–1797.
- Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, Armbrust EV, Archibald JM, Bharti AK, Bell CJ, Beszteri B, Bidle KD, Cameron CT, Campbell L, Caron DA, Cattolico RA, Collier JL, Coyne K, Davy SK, Deschamps P, Dyrman ST, Edvardsen B, Gates RD, Gobler CJ, Greenwood SJ, Guida SM, Jacobi JL, Jakobsen KS, James ER, Jenkins B, John U, Johnson MD, Juhl AR, Kamp A, Katz LA, Kiene R, Kudryavtsev A, Leander BS, Lin S, Lovejoy C, Lynn D, Marchetti A, McManus G, Nedelcu AM, Menden-Deuer S, Miceli C, Mock T, Montresor M, Moran MA, Murray S, Nadathur G, Nagai S, Ngam PB, Palenik B, Pawlowski J, Petroni G, Piganeau G, Posewitz MC, Rengefors K, Romano G, Rumpho ME, Ryneerson T, Schilling KB, Schroeder DC, Simpson AG, Slamovits CH, Smith DR, Smith GJ, Smith SR, Sosik HM, Stief P, Theriot E, Twary SN, Umale PE, Vaultot D, Wawrik B, Wheeler GL, Wilson WH, Xu Y, Zingone A, Worden AZ. 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol* **12**(6):e1001889.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**(9):1312–1313.
- Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol* **30**(12):2725–2729.