Phylogenomics of Marine Algae

Cheong Xin Chan*

The University of Queensland, Institute for Molecular Bioscience, and ARC Centre of Excellence in Bioinformatics, Brisbane, QLD 4072, Australia

* c.chan@imb.uq.edu.au (corresponding author)

ABSTRACT   *De novo* genome and transcriptome data from a number of marine algal species have recently become available, ranging from red, green and brown algae, as well as other photosynthetic eukaryotes, e.g. diatoms and dinoflagellates. Phylogenomic approaches are widely adopted to decipher the evolutionary relationships among diverse lineages. Novel algal genomes therefore provide an exciting analysis platform for understanding algal biology, ecophysiology and diversity, and at a broader scale, eukaryote evolution. In this brief communication, I highlight major findings from recent phylogenomic studies of marine algae and their impact to the research field. I then discuss the current trends and future directions of phylogenomics, and how we can apply this approach in studying biodiversity in the South China Sea.

(**Keywords:** algal genomics, biodiversity, endosymbiosis, lateral genetic transfer, marine algae, phylogenomics)

INTRODUCTION

Marine algae are essential primary producers in the oceans. Having a photoautotrophic (photosynthetic) lifestyle, these algae convert carbon dioxide into oxygen that is crucial for sustaining the entire marine ecosystems. The oceans, covering ca. 70% of the planet surface with an estimated area of 362 million km^2 and a total volume of 1.36 billion km^3 [1], consist of diverse marine flora and fauna. The diversity of algae, both in marine and freshwaters, has been conservatively estimated at about 300,000 species, with a rough estimate of over 1 million species [2]. Marine algae encompass microalgae, macroalgae (i.e. seaweeds) and phytoplanktons (e.g. the ubiquitous diatoms and dinoflagellates), living in a wide range of habitats ranging from the estuaries, ocean surface to coral reefs. Some marine algae produce high level of lipids [3] and hydrocolloid compounds [4], showing great potential in biotechnology, particularly in the industries of food and animal feed, pharmaceuticals, cosmetics, and perhaps more importantly, biofuels [5]. In addition, some algae represent the most ancient lineages of photosynthetic eukaryotes. Their genomes therefore contain clues that will help us understand the evolution of photosynthetic eukaryotes, and of all eukaryotes in general. A sound understanding of how eukaryotes first became photosynthetic is also important for deciphering the geological and atmospheric histories of planet Earth, e.g. the Great Oxygenation Event ca. 2.4 billion years ago [6].

Phylogenomics is the study of evolutionary histories among organismal lineages based on comparative analysis of genome-scale data. Extending from phylogenetic analysis at the gene level, phylogenomic inference is commonly observed based on gene-by-gene [7, 8], concatenated multi-genes [9-11] or whole-genome [12] comparisons. Recently genome sequences of marine algae have become available, providing an excellent analysis platform for phylogenomics. In this brief communication to the Second South China Sea conference, I highlight major findings from recent phylogenomic studies of marine algae, current trends and future directions, and how we can apply this approach in studying marine diversity in the South China Sea.

WHAT HAVE WE LEARNED?

As algal genome data are becoming available, many recent studies of algal genomes have been focusing on early evolution of eukaryotes, particularly of the origin of organelles, a key characteristic distinguishing these lineages from the prokaryotes (Bacteria and Archaea). The origin of plastid (chloroplast) in eukaryotes, especially, has gathered much interest within the research community. Two main types of plastid occur in eukaryotes: the simple, two-membrane-bound primary plastids, and the more structurally complex, three- or four-membrane-bound

secondary (or tertiary) plastids [13-15]. Primary plastids are found in red algae (Rhodophyta), green algae and plants (Viridiplantae), and glaucophyte algae (Glaucophyta); these three lineages are commonly grouped as the Archaeplastida or Plantae supergroup [16, 17]. The origin of these plastids is traced back to a cyanobacterial source, in which a cyanobacterium was engulfed by and retained within a heterotrophic host (i.e. primary endosymbiosis) [18], estimated to have occurred around 1 to 1.5 billion years ago [19, 20]. The engulfed endosymbiont gradually became the extant plastids. During endosymbiosis, genetic materials were transferred from the endosymbiont to the host nucleus. On the other hand, the more-complex plastids are found in all other algae, e.g. brown algae, diatoms and dinoflagellates. The origin of these plastids is complicated by multiple, serial events of endosymbiosis involving already plastid-bearing endosymbionts [21, 22], again, with genetic transfer instances during each event, complicating their evolutionary histories.

Ideally, a thorough assessment of the evolutionary histories of algae and other microbial eukaryotes using phylogenomics requires well-annotated, experimentally validated gene repertoires from all known species, which is currently impossible. However, availability of novel algal genome data allows for addressing interesting (and generating novel) biological hypotheses that would yield novel insights into eukaryote (and plastid) evolution. For instance, although the Plantae supergroup is expected to share a common origin, the support for this hypothesis had been based on studies of a small number of genes [11, 17]. This is partly due to lack of gene repertoires for red and glaucophyte algae, as data from these lineages are scarce in comparison to the genome data available for green algae and plants. Using novel red algal transcriptome of the unicellular *Porphyridium cruentum* and genome data from the multicellular, coralline red alga *Calliarthron tuberculosum*, a recent study [23] has demonstrated a strong support for Plantae supergroup (in ca. 50% of the analysed phylogenies), further reinforced by later study incorporating the first genome of glaucophyte algae, *Cyanophora paradoxa* [24]. These studies, together with the earlier works [11, 17, 25] demonstrate a single origin of primary plastids and a great extent of genetic transfer events among algal lineages, and from bacterial sources.

The evolution of algae that possess secondary (and tertiary) plastids, however, is more complicated and contentious. The positions of these lineages on the eukaryote tree of life are far from being resolved, as demonstrated in a number of studies [9, 26-28]. The

genomes of two diatom species have recently become available [29, 30]. A large-scale phylogenomics analysis of the diatom genes suggests a cryptic endosymbiosis involving green algae (specifically a prasinophyte-like endosymbiont) [31], additional to red algal endosymbiont that is commonly associated with secondary endosymbiosis. In an independent analysis that incorporated a larger number of red algal genes [23] and new genome of the brown seaweed *Ectocarpus siliculosus* [32] (Stramenopiles; as with diatoms), a proportion of genes encoding membrane transporters in diatoms (ca. 25% of examined phylogenies) are found to have a red and/or green algal prominence [33]. The extent of genetic transfer in prokaryotes is well known to be rampant [7, 8, 34] and to extend beyond gene boundaries [35, 36]. Interestingly, in a recent transcriptome analysis of the dinoflagellate *Alexandrium tamarense* (that causes harmful algal blooms, or "red tide") [37], the extent of genetic transfer in microbial eukaryotes is shown to be comparable to that in prokaryotes, despite more-complex coding capacity in eukaryotes. The dinoflagellates can be considered as the worst-case scenario in terms of the complexity of algal evolution, because tertiary and quarternary endosymbiosis events involving other eukaryotic (e.g. haptophyte-like) cells have been postulated [22, 38, 39], and genes of bacterial origin have been reported in other studies [40, 41]. Some have argued that the cryptic green algal endosymbiosis hypothesis could be an artefact due to the lack of red algal genes [42, 43], and that the chlamydial origin of some of the Plantae genes as reported in [24] are due to technical biases [44]. Nevertheless, all these studies demonstrate algal and bacterial genetic transfer as key contributing factors to the adaption and survival of the ubiquitous microbial species in the fluctuating marine environments.

Many of the red algal gene repertoires available are from unicellular species, with the complete, highly reduced genome of the hyperthermophilic *Cyanidioschyzon merolae* [45]. Recently, extensive transcriptome data (ca. 4.7 million expressed sequence tags) from two multicellular, macroalgal species of *Porphyra* has become available [46]. The seaweed *Porphyra* is well known for its application in food (e.g. "nori" used in sushi wraps) due to its high nutritional value, with its mariculture valued at USD1.3 billion a year [47]. Phylogenomic analysis of the *Porphyra* membrane transporters supports the important role of genetic transfer in environmental adaption of microbial eukaryotes [48]. These studies generated novel insights into red algal biology [49], e.g. a fatty acid biosynthesis pathway that is distinct from that in plants [48]. In

addition, other studies have demonstrated that green algal derived genes in microbial eukaryotes are important for the function of light-harvesting complex superfamily [50], as well as for protection from oxidative damage of genes involved in carotenoid biosynthesis [51]. Genetic transfer has also recently been demonstrated in other algal lineages, such as the cryptophyte *Guillardia theta* and chlorarachniophyte *Bigelowiella natans*, with their respective relict endosymbiont nucleus within the cell (i.e. nucleomorphs) [52]. Many of these findings remain to be experimentally validated. However, genetic transfer is undoubtedly a key driving force in the evolution of algae (and microbial eukaryotes), for instance, in adapting to the fluctuating concentrations of dissolved oxygen and/or redox-sensitive transition metals, over evolutionary time-scales.

Recent development of single-cell genomics [53, 54] provides the opportunity for capturing snapshots of genomes (and other omic data). This approach is useful for studying genomics and its variation within a population, and the genetic mechanisms underlying responses of the organisms to their environments. This exciting capability of observing evolving genome *in situ* (i.e., experimental evolution) would greatly enhance our understanding of ecology and evolution of specific organisms, as well as their interactions with one another and with the environments. Phylogenomic analysis of three single-cell genomes of an unculturable marine "algal" species of picobiliphytes [55] has demonstrated that these cells are in fact heterotrophic instead of photoautotrophic as originally thought [56]. This study showcased the typical case of keptoplasty, in which the observed plastid within the picobiliphyte cells when they were first described [56] was "stolen" from another source, i.e. the plastid was within an algal cell that was engulfed by the picobiliphyte. Interestingly, the variation of genome data among these three cells suggests different physiological conditions of the cell, e.g. one that is infected by a virus, and one that is actively feeding bacterial cells. Therefore, phylogenomics based on single-cell genome data, in this case, has uncovered hidden biodiversity in the marine environment at the cell-by-cell level, which was previously not possible using the conventional genomic approaches based on cultured cells.

CURRENT TRENDS AND FUTURE PERSPECTIVES

Owing to its sequence-centric nature, phylogenomic approaches are heavily dependent on the availability and quality of sequence data. Thanks to the rapid development of next-generation sequencing technologies, biological research is now data-rich, in which omic data of immense quantity are generated at great breadth and huge depth, ranging from genomes, transcriptomes, epigenomes and exomes, to meta-genomes and meta-transcriptomes. As phylogenomic analyses become more common in the literature, a question remains: is the approach adopted in these studies still the state-of-the-art, or should we spend more time in developing one that is better? In other words, where is the balance between extracting as much information as we can from the rapidly growing data using our current know-how, *versus* exploring approaches that would take us perhaps closer to the truth? This question has no easy answer.

With generation of sequence data becoming a routine practice in most laboratories, the limitations of computational and human power in data management, interpretation and analysis cannot be overstated. There is a need to decipher as much biological information from these data as possible. The quality of sequence data *vis-à-vis* stochastic sequence variation, convergence, long-branch attraction, incomplete sequence data (e.g. transcriptome, gene fragments) in addition to contaminations, represents the biggest hurdle in phylogenomics (and any sequence analysis) [57-59]. Due to this, one could argue that current approaches would yield biased inferences, thus resulting observations that would be of little use, or at the extreme, useless. A common ground between the two schools of thoughts is crucial for the field to move forward. One could always improve the phylogenetic framework, e.g. in perfecting multiple sequence alignment, identification of homologous sequence groups or phylogenetic algorithms to reduce inaccurate biological interpretations. On the other hand, providing more-efficient scalability, higher computing capacity, better implementations and sampling strategies among existing data, phylogenomic studies could yield valuable insights into algal biology and evolution. Although phylogeny itself is a working hypothesis, not *bona fide* truth, these studies, particularly of *de novo* genomes in which little (if any) prior information is known, provide an excellent test bed for novel, hypothesis-driven research into algal biology and evolution, e.g. genome innovation relative to environmental adaptation.

Next, do we have the right data to justify our conclusions or to test our hypotheses? Most sequenced genomes are economically and medically important species, with little data available from marine algae, or the more-ancient,

abundant lineages of microbial eukaryotes. The latter would provide invaluable insights into early eukaryote evolution. Where genome data is unavailable, the use of transcriptome data in phylogenomic analysis has been reported [37, 60]. However, assembled transcriptome data (e.g. based on expressed sequence tags) contain partial gene transcripts and could be biased by environmental conditions during which genetic materials were harvested from the organisms or culture. Multiple sequence alignment of these sequences alongside with other (putatively homologous) full-length sequences inevitably creates undesirable "gappy" aligned positions (i.e. phylogenetically non-informative sites) that would affect subsequent inference of phylogeny. In these cases, an attractive strategy is to use alignment-free methods in calculating sequence distances (e.g. using *k*-mers) [61-64], which does not require contiguity of homologous sequences to be conserved. Alternative phylogenetic representations independent from the tree-like structure, e.g. the use of networks [65, 66], would also provide a fresh perspective into genome evolution. Although these approaches unlikely compensate our limitation in inadequate taxon sampling, they certainly help in reducing data biases.

CONCLUDING REMARKS

The South China Sea contains abundant marine biodiversity due to the warm, tropical climate in the region. Phylogenomic approaches will be useful to assess the abundance of species in the reigon, and perhaps more importantly, genome data from tropical algal species (currently lacking in the public domain) will allow for large-scale comparative studies with the other temperate and extremophilic species, and therefore help enhancing our understanding of algal adaptation and evolution with respect to climate change. In addition, combining all types of omic data (e.g., including metabolome and non-coding elements) in a systems biology approach, we can better examine the interactions among these organisms and their responses to abiotic stresses, and learn how they adapt to the environments. Given the multi-disciplinary nature of algal genomic studies, regional cooperation and collaborative network, either within the East or Southeast Asian region, or externally with other world experts (e.g., as promoted by the South China Sea Conference initiative) should be encouraged and forged.

REFERENCES

1. Amante C. and Eakins B. W. (2009). ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. In *NOAA Technical Memorandum NESDIS NGDC-24*. NOAA National Geophysical Data Center. Boulder, Colarado. pp. 1-19.

2. Guiry M. D. (2012). How many species of algae are there? *J. Phycol.* **48**: 1057-1063.

3. Mata T. M., Martins A. A. and Caetano N. S. (2010). Microalgae for biodiesel production and other applications: a review. *Renew. Sust. Energ. Rev.* **14**: 217-232.

4. De Ruiter G. A. and Rudolph B. (1997). Carrageenan biotechnology. *Trends Food Sci. Technol.* **8**: 389-395.

5. Dismukes G. C., Carrieri D., Bennette N., Ananyev G. M. and Posewitz M. C. (2008). Aquatic phototrophs: efficient alternatives to land-based crops for biofuels. *Curr. Opin. Biotechnol.* **19**: 235-240.

6. Scott C., Lyons T. W., Bekker A., Shen Y., Poulton S. W., Chu X. and Anbar A. D. (2008). Tracing the stepwise oxygenation of the Proterozoic ocean. *Nature* **452**: 456-459.

7. Beiko R. G., Harlow T. J. and Ragan M. A. (2005). Highways of gene sharing in prokaryotes. *Proc. Natl. Acad. Sci. U. S. A.* **102**: 14332-14337.

8. Puigbò P., Wolf Y. I. and Koonin E. V. (2010). The tree and net components of prokaryote evolution. *Genome Biol. Evol.* **2**: 745-56.

9. Baurain D., Brinkmann H., Petersen J., Rodríguez-Ezpeleta N., Stechmann A., Demoulin V., Roger A. J., Burger G., Lang B. F. and Philippe H. (2010). Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. *Mol. Biol. Evol.* **27**: 1698-1709.

10. Burki F., Inagaki Y., Brate J., Archibald J. M., Keeling P. J., Cavalier-Smith T., Sakaguchi M., Hashimoto T., Horak A., Kumar S., Klaveness D., Jakobsen K. S., Pawlowski J. and Shalchian-Tabrizi K. (2009). Large-scale phylogenomic analyses reveal that two enigmatic protist lineages, telonemia and centroheliozoa, are related to photosynthetic chromalveolates. *Genome Biol. Evol.* **1**: 231-8.

11. Nozaki H., Iseki M., Hasegawa M., Misawa K., Nakada T., Sasaki N. and Watanabe M. (2007). Phylogeny of primary photosynthetic eukaryotes as deduced from slowly evolving nuclear genes. *Mol. Biol. Evol.* **24**: 1592-1595.

12. Rannala B. and Yang Z. H. (2008). Phylogenetic inference using whole genomes. *Annu. Rev. Genomics Hum. Genet.* **9**: 217-231.

13. McFadden G. I. (2001). Primary and secondary endosymbiosis and the origin of plastids. *J. Phycol.* **37**: 951-959.

14. Chan C. X., Gross J., Yoon H. S. and Bhattacharya D. (2011). Plastid origin and evolution: new models provide insights into old problems. *Plant Physiol.* **155**: 1552-1560.

15. Howe C. J., Barbrook A. C., Nisbet R. E. R., Lockhart P. J. and Larkum A. W. D. (2008). The origin of plastids. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **363**: 2675-2685.

16. Cavalier-Smith T. (1981). Eukaryote kingdoms: seven or nine? *Biosystems* **14**: 461-481.

17. Rodríguez-Ezpeleta N., Brinkmann H., Burey S. C., Roure B., Burger G., Löffelhardt W., Bohnert H. J., Philippe H. and Lang B. F. (2005). Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr. Biol.* **15**: 1325-1330.

18. Margulis L. (1970). *Origin of Eukaryotic Cells.* Yale University Press, New Haven.

19. Douzery E. J. P., Snell E. A., Bapteste E., Delsuc F. and Philippe H. (2004). The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? *Proc. Natl. Acad. Sci. U. S. A.* **101**: 15386-15391.

20. Yoon H. S., Hackett J. D., Ciniglia C., Pinto G. and Bhattacharya D. (2004). A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.* **21**: 809-818.

21. Reyes-Prieto A., Weber A. P. and Bhattacharya D. (2007). The origin and establishment of the plastid in algae and plants. *Annu. Rev. Genet.* **41**: 147-68.

22. Yoon H. S., Hackett J. D., Van Dolah F. M., Nosenko T., Lidie L. and Bhattacharya D. (2005). Tertiary endosymbiosis driven genome evolution in dinoflagellate algae.*Mol. Biol. Evol.* 22: 1299-1308.

23 Chan C. X., Yang E. C., Banerjee T., Yoon H. S., Martone P. T., Estevez J. M. and Bhattacharya D. (2011). Red and green algal monophyly and extensive gene sharing found in a rich repertoire of red algal genes. *Curr. Biol.* **21**: 328-333.

24. Price D. C., Chan C. X., Yoon H. S., Yang E. C., Qiu H., Weber A. P., Schwacke R., Gross J., Blouin N. A., Lane C., Reyes-Prieto A., Durnford D. G., Neilson J. A. D., Lang B. F., Burger G., Steiner J. M., Löffelhardt W., Meuser J. E., Posewitz M. C., Ball S., Arias M. C., Henrissat B., Coutinho P. M., Rensing S. A., Symeonidi A., Doddapaneni H., Green B. R., Rajah V. D., Boore J. and Bhattacharya D. (2012). *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. *Science* **335**: 843-847.

25. Nozaki H., Maruyama S., Matsuzaki M., Nakada T., Kato S. and Misawa K. (2009). Phylogenetic positions of Glaucophyta, green plants (Archaeplastida) and Haptophyta (Chromalveolata) as deduced from slowly evolving nuclear genes. *Mol. Phylogenet. Evol.* **53**: 872-880.

26. Burki F., Okamoto N., Pombert J. F. and Keeling P. J. (2012). The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. *Proc. R. Soc. B* **279**: 2246-2254.

27. Burki F., Shalchian-Tabrizi K., Minge M., Skjæveland Å., Nikolaev S. I., Jakobsen K. S. and Pawlowski J. (2007). Phylogenomics reshuffles the eukaryotic supergroups. *PLoS ONE* **2**: e790.

28. Parfrey L. W., Grant J., Tekle Y. I., Lasek-Nesselquist E., Morrison H. G., Sogin M. L., Patterson D. J. and Katz L. A. (2010). Broadly sampled multigene analyses yield a well-resolved eukaryotic tree of life. *Syst. Biol.* **59**: 518-33.

29. Armbrust E. V., Berges J. A., Bowler C., Green B. R., Martinez D., Putnam N. H., Zhou S., Allen A. E., Apt K. E., Bechner M., Brzezinski M. A., Chaal B. K., Chiovitti A., Davis A. K., Demarest M. S., Detter J. C., Glavina T., Goodstein D., Hadi M. Z., Hellsten U., Hildebrand M., Jenkins B. D., Jurka J., Kapitonov V. V., Kroger N., Lau W. W., Lane T. W., Larimer F. W., Lippmeier J. C., Lucas S., Medina M., Montsant A., Obornik M., Parker M. S., Palenik B., Pazour G. J., Richardson P. M., Rynearson T. A., Saito M. A., Schwartz D. C., Thamatrakoln K., Valentin K., Vardi A., Wilkerson F. P. and Rokhsar D. S. (2004). The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**: 79-86.

30. Bowler C., Allen A. E., Badger J. H., Grimwood J., Jabbari K., Kuo A., Maheswari U., Martens C., Maumus F., Otillar R. P., Rayko E., Salamov A., Vandepoele K., Beszteri B., Gruber A., Heijde M., Katinka M., Mock T., Valentin K., Verret F., Berges J. A., Brownlee C., Cadoret J. P., Chiovitti A., Choi C. J., Coesel S., De Martino A., Detter J. C., Durkin C., Falciatore A., Fournet J., Haruta M., Huysman M. J., Jenkins B. D., Jiroutova K., Jorgensen R.

E., Joubert Y., Kaplan A., Kroger N., Kroth P. G., La Roche J., Lindquist E., Lommer M., Martin-Jezequel V., Lopez P. J., Lucas S., Mangogna M., McGinnis K., Medlin L. K., Montsant A., Oudot-Le Secq M. P., Napoli C., Obornik M., Parker M. S., Petit J. L., Porcel B. M., Poulsen N., Robison M., Rychlewski L., Rynearson T. A., Schmutz J., Shapiro H., Siaut M., Stanley M., Sussman M. R., Taylor A. R., Vardi A., von Dassow P., Vyverman W., Willis A., Wyrwicz L. S., Rokhsar D. S., Weissenbach J., Armbrust E. V., Green B. R., Van de Peer Y. and Grigoriev I. V. (2008). The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**: 239-44.

31. Moustafa A., Beszteri B., Maier U. G., Bowler C., Valentin K. and Bhattacharya D. (2009). Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* **324**: 1724-1726.

32. Cock J. M., Sterck L., Rouze P., Scornet D., Allen A. E., Amoutzias G., Anthouard V., Artiguenave F., Aury J. M., Badger J. H., Beszteri B., Billiau K., Bonnet E., Bothwell J. H., Bowler C., Boyen C., Brownlee C., Carrano C. J., Charrier B., Cho G. Y., Coelho S. M., Collen J., Corre E., Da Silva C., Delage L., Delaroque N., Dittami S. M., Doulbeau S., Elias M., Farnham G., Gachon C. M., Gschloessl B., Heesch S., Jabbari K., Jubin C., Kawai H., Kimura K., Kloareg B., Kupper F. C., Lang D., Le Bail A., Leblanc C., Lerouge P., Lohr M., Lopez P. J., Martens C., Maumus F., Michel G., Miranda-Saavedra D., Morales J., Moreau H., Motomura T., Nagasato C., Napoli C. A., Nelson D. R., Nyvall-Collen P., Peters A. F., Pommier C., Potin P., Poulain J., Quesneville H., Read B., Rensing S. A., Ritter A., Rousvoal S., Samanta M., Samson G., Schroeder D. C., Segurens B., Strittmatter M., Tonon T., Tregear J. W., Valentin K., von Dassow P., Yamagishi T., Van de Peer Y. and Wincker P. (2010). The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature* **465**: 617-21.

33. Chan C. X., Reyes-Prieto A. and Bhattacharya D. (2011). Red and green algal origin of diatom membrane transporters: insights into environmental adaptation and cell evolution. *PLoS ONE* **6**: e29138.

34. Zhaxybayeva O., Gogarten J. P., Charlebois R. L., Doolittle W. F. and Papke R. T. (2006). Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome Res.* **16**: 1099-1108.

35. Chan C. X., Beiko R. G., Darling A. E. and Ragan M. A. (2009). Lateral transfer of genes and gene fragments in prokaryotes. *Genome Biol. Evol.* **1**: 429-438.

36. Chan C. X., Darling A. E., Beiko R. G. and Ragan M. A. (2009). Are protein domains modules of lateral genetic transfer? *PLoS ONE* **4**: e4524.

37. Chan C. X., Soares M. B., Bonaldo M. F., Wisecaver J. H., Hackett J. D., Anderson D. M., Erdner D. L. and Bhattacharya D. (2012). Analysis of *Alexandrium tamarense* (Dinophyceae) genes reveals the complex evolutionary history of a microbial eukaryote. *J. Phycol.* **48**: 1130-1142.

38. Hackett J. D., Anderson D. M., Erdner D. L. and Bhattacharya D. (2004). Dinoflagellates: a remarkable evolutionary experiment. *Am. J. Bot.* **91**: 1523-1534.

39. Wisecaver J. H. and Hackett J. D. (2011). Dinoflagellate genome evolution. *Annu. Rev. Microbiol.* **65**: 369-387.

40. Nosenko T. and Bhattacharya D. (2007). Horizontal gene transfer in chromalveolates. *BMC Evol. Biol.* **7**: 173.

41. Slamovits C. H., Okamoto N., Burri L., James E. R. and Keeling P. J. (2011). A bacterial proteorhodopsin proton pump in marine eukaryotes. *Nat. Commun.* **2**: 183.

42. Burki F., Flegontov P., Obornik M., Cihlář J., Pain A., Lukeš J. and Keeling P. J. (2012). Re-evaluating the green versus red signal in eukaryotes with secondary plastid of red algal origin. *Genome Biol. Evol.* **4**: 626-635.

43. Deschamps P. and Moreira D. (2012). Reevaluating the green contribution to diatom genomes. *Genome Biol. Evol.* **4**: 683-8.

44. Dagan T., Roettger M., Stucken K., Landan G., Koch R., Major P., Gould S. B., Goremykin V. V., Rippka R., de Marsac N. T., Gugger M., Lockhart P. J., Allen J. F., Brune I., Maus I., Pühler A. and Martin W. F. (2012). Genomes of Stigonematalean cyanobacteria (Subsection V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids. *Genome Biol. Evol. 5:31-44.*

45. Matsuzaki M., Misumi O., Shin-i T., Maruyama S., Takahara M., Miyagishima S. Y., Mori T., Nishida K., Yagisawa F., Nishida K., Yoshida Y., Nishimura Y., Nakao S., Kobayashi T., Momoyama Y., Higashiyama T., Minoda A., Sano M., Nomoto H., Oishi K., Hayashi H., Ohta F., Nishizaka S., Haga S., Miura S., Morishita T., Kabeya Y., Terasawa K., Suzuki Y., Ishii Y., Asakawa S., Takano H., Ohta N., Kuroiwa H., Tanaka K., Shimizu N., Sugano S., Sato N., Nozaki H., Ogasawara N., Kohara Y. and Kuroiwa T. (2004). Genome sequence of the

ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* **428**: 653-657.

46. Chan C. X., Blouin N. A., Zhuang Y., Zäuner S., Prochnik S. E., Lindquist E., Lin S., Benning C., Lohr M., Yarish C., Gantt E., Grossman A. R., Lu S., Müller K., W. Stiller J., Brawley S. H. and Bhattacharya D. (2012). *Porphyra* (Bangiophyceae) transcriptomes provide insights into red algal development and metabolism. *J. Phycol.* **48**: 1328-1342.

47. Blouin N. A., Brodie J. A., Grossman A. C., Xu P. and Brawley S. H. (2011). Porphyra: a marine crop shaped by stress. *Trends Plant Sci.* **16**: 29-37.

48. Chan C. X., Zauner S., Wheeler G., Grossman A. R., Prochnik S. E., Blouin N. A., Zhuang Y. Y., Benning C., Berg G. M., Yarish C., Eriksen R. L., Klein A. S., Lin S. J., Levine I., Brawley S. H. and Bhattacharya D. (2012). Analysis of *Porphyra* membrane transporters demonstrates gene transfer among photosynthetic eukaryotes and numerous sodium-coupled transport systems. *Plant Physiol.* **158**: 2001-2012.

49. Stiller J. W., Perry J., Rymarquis L. A., Accerbi M., Green P. J., Prochnik S., Lindquist E., Chan C. X., Yarish C., Lin S. J., Zhuang Y. Y., Blouin N. A. and Brawley S. H. (2012). Major developmental regulators and their expression in two closely related species of *Porphyra* (Rhodophyta). *J. Phycol.* **48**: 883-896.

50. Peers G., Truong T. B., Ostendorf E., Busch A., Elrad D., Grossman A. R., Hippler M. and Niyogi K. K. (2009). An ancient light-harvesting protein is critical for the regulation of algal photosynthesis. *Nature* **462**: 518-521.

51. Frommolt R., Werner S., Paulsen H., Goss R., Wilhelm C., Zauner S., Maier U. G., Grossman A. R., Bhattacharya D. and Lohr M. (2008). Ancient recruitment by chromists of green algal genes encoding enzymes for carotenoid biosynthesis. *Mol. Biol. Evol.* **25**: 2653-2667.

52. Curtis B. A., Tanifuji G., Burki F., Gruber A., Irimia M., Maruyama S., Arias M. C., Ball S. G., Gile G. H., Hirakawa Y., Hopkins J. F., Kuo A., Rensing S. A., Schmutz J., Symeonidi A., Elias M., Eveleigh R. J., Herman E. K., Klute M. J., Nakayama T., Obornik M., Reyes-Prieto A., Armbrust E. V., Aves S. J., Beiko R. G., Coutinho P., Dacks J. B., Durnford D. G., Fast N. M., Green B. R., Grisdale C. J., Hempel F., Henrissat B., Hoppner M. P., Ishida K., Kim E., Koreny L., Kroth P. G., Liu Y., Malik S. B., Maier U. G., McRose D., Mock T., Neilson J. A., Onodera N. T., Poole A. M., Pritham

E. J., Richards T. A., Rocap G., Roy S. W., Sarai C., Schaack S., Shirato S., Slamovits C. H., Spencer D. F., Suzuki S., Worden A. Z., Zauner S., Barry K., Bell C., Bharti A. K., Crow J. A., Grimwood J., Kramer R., Lindquist E., Lucas S., Salamov A., McFadden G. I., Lane C. E., Keeling P. J., Gray M. W., Grigoriev I. V. and Archibald J. M. (2012). Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* **492**: 59-65.

53. Lasken R. S. (2007). Single-cell genomic sequencing using Multiple Displacement Amplification. *Curr. Opin. Microbiol.* **10**: 510-516.

54. Woyke T., Xie G., Copeland A., Gonzalez J. M., Han C., Kiss H., Saw J. H., Senin P., Yang C., Chatterji S., Cheng J. F., Eisen J. A., Sieracki M. E. and Stepanauskas R. (2009). Assembling the marine metagenome, one cell at a time. *PLoS ONE* **4**: e5299.

55. Yoon H. S., Price D. C., Stepanauskas R., Rajah V. D., Sieracki M. E., Wilson W. H., Yang E. C., Duffy S. and Bhattacharya D. (2011). Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science* **332**: 714-717.

56. Not F., Valentin K., Romari K., Lovejoy C., Massana R., Tobe K., Vaulot D. and Medlin L. K. (2007). Picobiliphytes: a marine picoplanktonic algal group with unknown affinities to other eukaryotes. *Science* **315**: 253-255.

57. Stiller J. W. (2011). Experimental design and statistical rigor in phylogenomics of horizontal and endosymbiotic gene transfer. *BMC Evol. Biol.* **11**: 259.

58. Philippe H., Brinkmann H., Lavrov D. V., Littlewood D. T. J., Manuel M., Wörheide G. and Baurain D. (2011). Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol.* **9**: e1000602.

59. Philippe H., Snell E. A., Bapteste E., Lopez P., Holland P. W. H. and Casane D. (2004). Phylogenomics of eukaryotes: impact of missing data on large alignments. *Mol. Biol. Evol.* **21**: 1740-1752.

60. Struck T. H., Paul C., Hill N., Hartmann S., Hösel C., Kube M., Lieb B., Meyer A., Tiedemann R., Purschke G. and Bleidorn C. (2011). Phylogenomic analyses unravel annelid evolution. *Nature* **471**: 95-98.

61. Höhl M. and Ragan M. A. (2007). Is multiple-sequence alignment required for accurate inference of phylogeny? *Syst. Biol.* **56**: 206-221.

62. Vinga S. and Almeida J. (2003). Alignment-free sequence comparison - a review. *Bioinformatics* **19**: 513-523.

63. Reinert G., Chew D., Sun F. and Waterman M. S. (2009). Alignment-free sequence comparison (I): statistics and power. *J. Comput. Biol.* **16**: 1615-34.

64. Jing J., Burden C. J., Forêt S. and Wilson S. R. (2010). Statistical considerations underpinning an alignment-free sequence comparison method. *J. Korean Stat. Soc.* **39**: 325-335.

65. Huson D. H. and Scornavacca C. (2011). A survey of combinatorial methods for phylogenetic networks. *Genome Biol. Evol.* **3**: 23-35.

66. Dagan T. (2011). Phylogenomic networks. *Trends Microbiol.* **19**: 483-491.