The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae

J. Mark Cock, Lieven Sterck, Pierre Rouzé, Delphine Scornet, Andrew E. Allen, Grigoris Amoutzias, Veronique Anthouard, François Artiguenave, Jean-Marc Aury, Jonathan H Badger, et al.

To cite this version:

J. Mark Cock, Lieven Sterck, Pierre Rouzé, Delphine Scornet, Andrew E. Allen, et al.. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. Nature, Nature Publishing Group, 2010, 465 (7298), pp.617-621. 10.1038/nature09016 . cea-00906990

**HAL Id: cea-00906990**

https://hal-cea.archives-ouvertes.fr/cea-00906990

Submitted on 15 Oct 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike| 4.0 International License
Cock JM, Sterck L, Rouzé P, et al.  
**The Ectocarpus genome and the independent evolution of multicellularity in brown algae.**  
*Nature* 2010, 465(7298), 617-621.

Copyright:  
This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit [http://creativecommons.org/licenses/by-nc-sa/3.0/](http://creativecommons.org/licenses/by-nc-sa/3.0/)

Link for published article:  
[http://dx.doi.org/10.1038/nature09016](http://dx.doi.org/10.1038/nature09016)

**Date deposited:** 25th November 2013
The Ectocarpus genome and the independent evolution of multicellularity in brown algae

J. Mark Cock,1,2 Lieven Sterck,1,4 Pierre Rouze,3,4 Delphine Scornet,1,2 Andrew E. Allen,5 Grigoris Amoutzias,1,4 Veronique Anthouard,6 Francois Artiguenave,6 Jean-Marc Aury,9 Jonathan H. Badger5 Bank Beszteri7,21 Kenny Billiu1,4, Eric Bonnet,1,5 John H. Bothwell9,10 Chris Bowler11,12 Catherine Boyen1,2 Colin Brownlee10 Carl J. Carrano13, Bénédicte Charrier1,2 Ga Youn Cho1,2 Susana M. Coelho1,2, Jonas Collén1,2, Erwan Corre14, Corinne Da Silva,16 Ludovic Delage1,2 Nicolas Delaroque15, Simon M. Dittami1,2 Sylvie Doulbeau16 Marek Elias17 Garry Farnham18, Claire M. M. Gachon19 Bernhard Gschlossi1,2, Svenja Heesch18, Kamel Jabbari11,14 Claire Jubit14, Hiroshi Kawai19, Kei Kimura20, Bernard Kloareg1,2 Frithjof C. Küpper18, Daniel Lang13, Aude Le Bail11 Catherine Leblanc1,2 Patrice Lefort22 Martin Lohr23 Pascal J. Lopez11 Cindy Martens14, Florian Maumus11, Gurvan Michel1,2 Diego Miranda-Saavedra24, Julia Morales25,26 Hervé Moreau27 Taizo Motomura20 Chikako Nagasato20, Carolyn A. Napoli28, David R. Nelson29, Pi Nyvall-Colén1,2, Akira F. Peters1,2,4, Cyril Pommier30 Philippe Potin1,2 Julie Pouilain3, Hadi Quesneville3, Betsy Read1,2 Stefan A. Rensing2, Andrés Ritter1,2,32 Sylvie Rousseau1,2 Manoj Samanta5, Gaelle Samson6, Declan C. Schroeder10, Béatrice Ségurens6 Martina Strittmatter18 Thierry Tonon1,2, James W. Tregear16, Klaus Valentijn7, Peter von Dassow34 Takahiro Yamagishi19, Yves Van de Peer3,4 Patrick Wincker6

Brown algae (Phaeophyceae) are complex photosynthetic organisms with a very different evolutionary history to green plants, to which they are only distantly related1. These seaweeds are the dominant species in rocky coastal ecosystems and they exhibit many interesting adaptations to these, often harsh, environments. Brown algae are also one of only a small number of eukaryotic lineages that have evolved complex multicellularity (Fig. 1). We report the 214 million base pair (Mb) genome sequence of the filamentous seaweed Ectocarpus siliculosus (Dillwyn) Lyngbye, a model organism for brown algae1,4, closely related to the kelps6 (Fig. 1). Genome features such as the presence of an extended set of light-harvesting and pigment biosynthesis genes and new metabolic processes such as halide metabolism help explain the ability of this organism to cope with the highly variable tidal environment. The evolution of multicellularity in this lineage is correlated with the presence of a rich array of signal transduction gene families. Of particular interest is the presence of a family of receptor kinases, as the independent evolution of related molecules has been linked with the emergence of multicellularity in both the animal and green plant lineages. The Ectocarpus genome sequence represents an important step towards developing this organism as a model species, providing the possibility to combine genomic and genetic2 approaches to explore these and other aspects of brown algal biology further.

The 16,256 protein coding genes present in the 214 Mb haploid male genome of E. siliculosus are rich in introns (seven per gene on average), have long 3’ untranslated regions (average size: 845 bp) and are often located very close to each other on the chromosome (29% of the intergenic regions between divergently transcribed genes are less than 400 bp long; Table 1 and Supplementary Information 2.1). Repeated sequences, including DNA transposons, retrotransposons and helitrons, make up 22.7% of the Ectocarpus genome. Small RNAs mapped preferentially to transposons, indicating that they have a role...
in silencing these elements despite the absence of detectable levels of cytosine methylation in the genome (Supplementary Information 2.1). Sequencing also revealed the presence of an integrated copy of a large DNA virus, closely related to the *Ectocarpus* phaeovirus EsV-1 (ref. 8; Fig. 2a). Approximately 50% of individuals in natural *Ectocarpus* populations show symptoms of viral infection9,10 but the sequenced genome is also predicted to encode 21 putative dehalogenases and two haloalkane dehalogenases. These enzymes may serve to protect *Ectocarpus*

![Diagram](image)

**Figure 1 | Simplified representation of the evolutionary tree of the eukaryotes showing the five major groups that have evolved complex multicellularity (indicated in colour).** Here we define groups showing complex multicellularity as those that include macroscopic organisms with defined, recognizable morphologies and composed of multiple cell types. Coloured bars indicate the approximate, relative times of emergence of complex multicellularity in each lineage. The inset tree to the right indicates the relationship of *Ectocarpus* to selected brown algal genera. Kelps are represented in the tree by the genus *Laminaria*.

in it to adapt to an environment with highly variable light conditions. The high levels of phenolic compounds in brown algae are thought to protect against ultraviolet radiation, in a manner analogous to flavonoids in terrestrial plants11. Homologues of most of the terrestrial plant flavonoid pathway genes were found in *Ectocarpus* but these are completely absent from diatom or green algal genomes (Supplementary Information 2.2.9). The diverse complement of enzymes involved in the metabolism of reactive oxygen species (Supplementary Information 2.2.11) is also likely to represent an important adaptation to osmotic and light stresses.

In the Laminarias, the high concentration of apoplastic iodide is thought to be used in a new anti-oxidant system that, through the emission of iodine, has an impact on atmospheric chemistry12. *Ectocarpus* also accumulates halides, although to a significantly lower level than in kelps (Supplementary Information 2.2.10). This difference was reflected in the genome; only one vanadium-dependent bromoperoxidase was found in contrast to the large families of haloperoxidases in *Laminaria digitata*13. The *Ectocarpus* genome does, however, encode 21 putative dehalogenases and two haloalkane dehalogenases. These enzymes may serve to protect *Ectocarpus*

![Graph](image)

**Figure 2 | An integrated viral sequence in the *Ectocarpus* genome.** a, Representation of the linear and circular forms of the EsV-1 genome compared to the inserted viral genome. Genes on the upper and lower strands are above and below the line, respectively. A short region of the viral genome are indicated. Dashed line, algal DNA. ITRA and ITRA', inverted terminal repeats. b, Mean expression levels of the inserted viral genes: the graph shows the mean of the normalized expression values (4 replicates) ± s.d. of the genes that were included in the microarray experiments carried out in ref. 29. Expression data are shown for the control condition, but gene expression profiles were highly similar under stress conditions (not shown). Each bar represents the expression value for one coding sequence, the bars are in the same order as the corresponding genes along the supercontigs. Red bars correspond to virus genes, blue bars to host genes. Supercontigs 0062, 0052 and 0028 are adjacent on the genetic map, supercontig 0371 is part of another linkage group. The hybridization signals for 95% of the negative controls (median of four random probes on the same array) were between 19 and 59 (indicated by the two dotted lines).

| Table 1 | *Ectocarpus* genome statistics |
|---|---|
| Size of the sequenced genome (Mbp) | 195.8 |
| Number of supercontigs (scaffolds) over 2kbp | 1,561 |
| Supercontig (scaffold) N50 (bp) | 504,428 |
| Number of contigs | 14,043 |
| Contig N50 (bp) | 32,862 |
| Percentage of the 91,041 cDNA sequences that match 97.4% |
| the genome |
| Genomic G+C content (%) | 53.6% |
| Percentage of repeated sequences | 22.7% |
| Number of genes | 16,256 |
| Average gene length (bp) | 6,859 |
| Average coding sequence length (bp) | 1,563 |
| Number of introns | 113,619 |
| Average intron length (bp) | 703.8 |
| Average number of introns per gene | 6.98 |
| Number of exons | 129,875 |
| Average exon length (bp) | 242.2 |
| Number of single exon genes | 856 |
| Number of genes with protein similarity support | 10,278 (62.2%) |
| (Blast e-value cutoff < e-10) |
| Number of genes with expressed sequence tag support | 9,601 (59%) |
| Number of genes with tiling array support | 6,474 (40%) |
against halogenated compounds produced by kelps as defence molecules\textsuperscript{12}, allowing it to grow epiphytically on these organisms\textsuperscript{14,15}.

The cell walls of brown algae contain unusual polysaccharides such as alginates and fucans\textsuperscript{16}, with properties that are important both in terms of resistance to mechanical stresses and as protection from predators. Analysis of the \textit{Ectocarpus} genome failed to detect homologues of many of the enzymes that are known, from other organisms, to have roles in alginate biosynthesis and in the remodelling of alginates, fucans and cellulose, indicating that brown algae have independently evolved enzymes to carry out many of these processes. However, a number of polysaccharide modifying enzymes, such as mannuronan C5 epimerases, sulphotransferases and sulphatases, were identified. These enzymes are likely to modulate physicochemical properties of the cell wall, influencing rigidity, ion exchange\textsuperscript{17} and resistance to abiotic stress.

Comparison of genomes from a broad range of organisms (Fig. 3) indicated that the major eukaryotic groups have retained distinct but overlapping sets of genes since their evolution from a common ancestor, with new gene families evolving independently in each lineage. On average, lineages that have given rise to multicellular organisms have lost fewer gene families and evolved more new gene families than unicellular lineages. However, we were not able to detect any significant, common trends, such as a tendency for the multicellular lineages to gain families belonging to particular functional (gene ontology) groups.

Analysis of the gene families that are predicted to have been gained by the \textit{Ectocarpus} genome since divergence from the unicellular diatoms indicated a significant gain in ontology terms associated with protein kinase activities, and these genes include a particularly interesting family of membrane-spanning receptor kinases. Receptor kinases have been shown to have key roles in developmental processes such as differentiation and cellular patterning in both the animal and green plant lineages\textsuperscript{18}. Animal tyrosine and green plant serine/threonine receptor kinases form two separate monophyletic clades, indicating that these families than unicellular lineages. However, we were not able to detect any significant, common trends, such as a tendency for the multicellular lineages to gain families belonging to particular functional (gene ontology) groups.

The ion channels in the \textit{Ectocarpus} genome illustrate how the evolutionary fates of eukaryotic lineages have probably depended not only on the evolution of new gene functions but also on the retention of genes already present in ancestral genomes. Along similar lines, there is evidence that, compared to unicellular organisms, multicellular organisms have tended to retain a more complete \textit{Rad51} family, which encodes DNA repair proteins including members with important roles during meiosis\textsuperscript{20}. This is also the case in the stramenopiles, where \textit{Ectocarpus} has a markedly more complete \textit{Rad51} gene family than the other sequenced members of the group (Supplementary Information 2.3, 12). \textit{Ectocarpus} also possesses a more extensive set of GTPase genes.
The emergence of the brown algal lineage. The former include the diverse origins of the genes that make up the genome, many of which were acquired via endosymbiotic events (Supplementary Information 2.3.15), whereas the latter include the recent emergence of new gene families. Other kinases than the unicellular diatoms (Supplementary Information 2.3.7) present from an early stage of eukaryotic evolution. Sixty-seven candidate target sites were identified for 12 of the 26 microRNAs.

Analysis of a large set of small RNA sequences allowed the identification of 26 microRNAs in Ectocarpus (Supplementary Table 17). This observation, together with the identification of microRNAs in three other eukaryotic groups, the archaebplastid, opisthokont and amoebozoan lineages, indicates that these regulatory molecules were present from an early stage of eukaryotic evolution. Sixty-seven candidate target sites were identified for 12 of the 26 microRNAs. Interestingly, 75% of these target sequences occur in genes with leucine-rich repeat (LRR) domains (Supplementary Information 2.3.14). The LRR genes include many members of the ROCO (Roc GTPase plus COR (C-terminal of Roc) domain) family that are predicted to have evolved since the split from the diatoms. Taken together, these observations indicate that a significant proportion of the microRNAs identified may regulate recently evolved processes. This is interesting in the light of suggestions that microRNAs may have had a key role in the evolution of complex multicellularity in the animal lineage.

Analysis of the Ectocarpus genome has revealed traces both of its ancient evolutionary past and of more recent events associated with the emergence of the brown algal lineage. The former include the diverse origins of the genes that make up the genome, many of which were acquired via endosymbiotic events (Supplementary Information 2.3.15), whereas the latter include the recent emergence of new gene families and the evolution of an unusual genome architecture, in terms both of gene structure and organization (Supplementary Information 2.1). It is likely that the evolution of complex multicellularity within brown algae depended on events spanning both timescales. The conservation of completeness and diversity within key gene families over the long term seems to have been as important as the more recent evolution of novel proteins, such as the brown algal receptor kinase family.

**METHODS SUMMARY**

Genome and cDNA sequencing were carried out using the Ectocarpus siliculosus strain Ec 32, which is a meiotic offspring of a field sporophyte collected in 1988 in San Juan de Marcona, Peru. The genome sequence was assembled using 2,233,253 and 903,939 paired, end-sequences from plasmid libraries with 3 and 10 kbp inserts respectively, plus 58,155 paired, end-sequence reads from a small-insert bacterial artificial chromosome library. Annotation was carried out using the EuGene program and optimized by manual correction of gene models and functional assignments. Sequencing of 91,041 cDNA reads, corresponding to six different cDNA libraries, and a whole genome tiling array analysis provided experimental confirmation of a large proportion of the transcribed part of the genome (Table 1). Small RNAs were characterized by generating 7,114,682 experimental confirmation of a large proportion of the transcribed part of the genome (Table 1). Small RNAs were characterized by generating 7,114,682 small RNAs were characterized by generating 7,114,682 small RNAs were characterized by generating 7,114,682 small RNAs were characterized by generating 7,114,682
LETTERS

3. Peters, A. F., Marie, D., Scornet, D., Kloareg, B. & Cock, J. M. Proposal of Ectocarpus siliculosus (Ectocarpales, Phaeophyceae) as a model organism for brown algal genetics and genomics. J. Phycol. 40, 1079–1088 (2004).

4. Charrier, B. et al. Development and physiology of the brown alga Ectocarpus siliculosus: two centuries of research. New Phytol. 177, 319–332 (2008).

5. Coelho, S. M. et al. Complex life cycles of multicellular eukaryotes: new approaches based on the use of model organisms. Gene 406, 152–170 (2007).

6. Kawai, H., Hanyuda, T., Draisma, S. G. A. & Müller, D. G. Molecular phylogeny of Discorisorum mesarthrocarpum (Phaeophyceae) with a reinstatement of the order Discorisorombiales. J. Phycol. 43, 186–194 (2007).

7. Phillips, N., Burrowes, R., Rousseau, F., de Reviers, B. & Saunders, G. W. Resolving evolutionary relationships among the brown algae using chloroplast and nuclear genes. J. Phycol. 44, 394–405 (2008).

8. Delaroque, N. et al. The complete DNA sequence of the Ectocarpus siliculosus virus EsV-1 genome. Virology 287, 112–132 (2001).

9. Dixon, N. M., Leadbeater, B. S. C. & Wood, K. R. Frequency of viral infection in a field population of Ectocarpus fasciculatus (Ectocarpales, Phaeophyceae). Physiologia 39, 258–263 (2000).

10. Müller, D. G. et al. Massive prevalence of viral DNA in Ectocarpus (Phaeophyceae, Ectocarpales) from two habitats in the North Atlantic and South Pacific. Biol. Mar. 43, 157–159 (2000).

11. Rozema, J. et al. The role of UV-B radiation in aquatic and terrestrial ecosystems—an experimental and functional analysis of the evolution of UV-absorbing compounds. J. Photochem. Photobiol. B 66, 2–12 (2002).

12. Köpper, F. C. et al. Iodide accumulation provides kelp with an inorganic antioxidant impacting atmospheric chemistry. Proc. Natl Acad. Sci. USA 105, 6954–6958 (2008).

13. Colin, C. et al. The brown algal kelp Laminaria digitata features distinct bromoperoxidase and iodoperoxidase activities. J. Biol. Chem. 278, 23545–23552 (2003).

14. Russell, G. Formation of an ectocarpoid epiflora on blades of Laminaria digitata. Mar. Ecol. Prog. Ser. 11, 181–187 (1983).

15. Russell, G. Parallel growth-patterns in algal epiphytes and Laminaria blades. Mar. Ecol. Prog. Ser. 13, 303–304 (1983).

16. Kloareg, B. & Quatrano, R. S. Structure of the cell walls of marine algae and the chloroplasts of Ectocarpus siliculosus (Ectocarpales, Phaeophyceae) as a model organism for brown algal genetics and genomics. Proc. Natl Acad. Sci. USA 103, 10328–10333 (2006).

17. De Smet, I., Voss, U., Jürgens, G. & Beeckman, T. Receptor-like kinases shape the plant. Nature Cell Biol. 11, 1166–1173 (2009).

18. Shiu, S. H. & Bleecker, A. B. Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. Proc. Natl Acad. Sci. USA 98, 10763–10768 (2001).

19. Cock, J. M., Vanooijhuyse, V. & Gaude, T. Receptor kinase signalling in plants and animals: distinct molecular systems with mechanistic similarities. Curr. Opin. Cell Biol. 14, 230–236 (2002).

20. Arnaout, M. A., Goodman, S. L. & Xiong, J.-P. Structure and mechanics of integrin-based cell adhesion. Curr. Opin. Cell Biol. 19, 495–507 (2007).

21. Nakayama, Y., Fujii, K., Sokabe, M. & Yoshimura, K. Molecular and electrophysiological characterization of a mechanosensitive channel expressed in the chloroplasts of Chlamydomonas. Proc. Natl Acad. Sci. USA 104, 5883–5888 (2007).

22. Wheeler, G. L. & Brownelee, C. Ca²⁺ signalling in plants and green algae – changing channels. Trends Plant Sci. 13, 506–514 (2008).

23. Goddard, H., Manison, N., Tomos, D. & Brownelee, C. Elemental propagation of calcium signals in response-specific patterns determined by environmental stimulus strength. Proc. Natl Acad. Sci. USA 97, 1932–1937 (2000).

24. Coelho, S. M. et al. Spatiotemporal patterning of reactive oxygen production and Ca²⁺ wave propagation in Fucus rhizoid cells. Plant Cell 14, 2369–2381 (2002).

25. Lin, Z., Gong, H., Nii, M. & Ma, H. Origins and evolution of the recA/RAD51 gene family: evidence for ancient gene duplication and endosymbiotic gene transfer. Proc. Natl Acad. Sci. USA 103, 10328–10333 (2006).

26. Griffiths-Jones, S., Saini, H., van Dongen, S. & Enright, A. miRBase: tools for microRNA genomics. Nucleic Acids Res. 36, D154–D158 (2008).

27. Marin, I., van Egmond, W. N. & van Haastert, P. J. M. The Roco protein family: a functional perspective. FASEB J. 22, 3103–3110 (2008).

28. Peterson, K. J., Dietrich, M. R. & McPeek, M. A. MicroRNAs and metazoan macroevolution: insights into canalization, complexity, and the Cambrian explosion. Bioessays 31, 736–747 (2009).

29. Dittami, S. M. et al. Global expression analysis of the brown alga Ectocarpus siliculosus (Phaeophyceae) reveals large-scale reprogramming of the transcriptome in response to abiotic stress. Genome Biol. 10, R66 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank Dieter G. Müller for his help and advice. The project was supported by the French GIS ‘Institut de la Génomique Marine’, the Centre National de Recherche Scientifique, the European Union network of excellence Marine Genomics Europe, the GIS Europôle Mer, the Inter-University Network for Fundamental Research (P6/25, BioMagNet), the ‘Conseil Général’ of the Finistère department and the University Pierre and Marie Curie.

Author Contributions J.M.C. coordinated genome analysis and manuscript preparation. P.W. and Y.V.d.P. coordinated genome assembly and centralized and enabled the annotation process, respectively. P.W. and Y.V.d.P. should be considered joint last authors. L.S. and P.R. implemented the automated annotation of the genome and made substantial contributions to genome annotation and analysis. D.S. developed and implemented protocols for library construction. L.S., P.R. and D.S. should be considered joint second authors. All other authors are members of the genome sequencing consortium and contributed annotation, analyses or data to the genome project.

Author Information The annotated Ectocarpus genome sequence can be obtained through the EMBL Nucleotide Sequence Database (accession numbers CABU01000001–CABU01013533, FN647682–FN649242, FN649726–FN649760) and can be browsed at the Bogas website (http://bioinformatics.psb.ugent.be/webtools/bogas/). cDNA sequence data are available through accession numbers FP245546–FP31261 and small RNA sequences and tiling array data have been submitted to the GEO database (accession numbers ERA000209 and GSE19912, respectively). The Ectocarpus microRNAs have been submitted to miRBase (accession numbers esi-MIR3450–esi-MIR3469). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.M.C. (cock@sb-roscoff.fr).

©2010 Macmillan Publishers Limited. All rights reserved