Le modèle algue brune pour l'analyse fonctionnelle et évolutive du déterminisme sexuel

Alexandre Cormier

Université Pierre et Marie Curie

Ecole doctorale Complexité du vivant (ED 515)

Laboratoire de Biologie Intégrative des Modèles Marins UMR 8227

Equipe de Génétique des algues, Station Biologique de Roscoff

Le modèle algue brune pour l'analyse fonctionnelle et évolutive du déterminisme sexuel

Par Alexandre Cormier

Thèse de doctorat en Bio-informatique

Dirigée par Susana Coelho et Mark Cock

Présentée et soutenue publiquement le 16 novembre 2015

Devant le jury composé de :

Dr. Leroy Philipe (INRA, Clermont-Ferrand)	Rapporteur
Dr. Renou Jean-Pierre (INRA, Angers) :	Rapporteur
Pr. Carbone Alessandra (UPMC, Paris) :	Examinatrice
Dr. Brunaud Véronique (INRA, Orsay) :	Examinatrice
Dr. Le Roux Frédérique (Ifremer, Roscoff) :	Représentante ED 515
Dr. Coelho Susana (CNRS-UPMC, Roscoff):	Directrice de thèse
Dr. Cock Mark (CNRS-UPMC, Roscoff) :	Directeur de thèse
Dr. Corre Erwan (CNRS-UPMC, Roscoff) :	Encadrant de thèse

Remerciements

Je tiens en premier lieu tout particulièrement à remercier mes encadrants de thèse, Susana Coelho, Mark Cock et Erwan Corre qui m'ont permis de réaliser ma thèse, pour leur encadrement et leur soutien pour mener à bien ce projet.

Je remercie également les membres de mon comité de thèse, Angela Falciatore, Armel Salmon et Lieven Sterck pour m'avoir suivie durant ces trois années, pour leurs conseils et leur intérêt à l'égard de mon sujet de thèse.

Je tiens à remercier Philipe Leroy, Jean-Pierre Renou, Alessandra Carbone et Véronique Brunaud d'avoir accepté de faire partie de mon jury de thèse et pour le temps consacré à la lecture de mon manuscrit.

Je remercie aussi les personnes de mon équipe, présentes et passées, Nicolas Macaisne, Olivier Godfroy, Sophia Ahmed, Komlan Avia, Rémy Luthringer, Fiona Lerck, Martina Strittmatter, Alok Arun, Delphine Scornet, Agnieszka Lipinska, Marie-Mathilde Perrineau, Simon Bourdareau et Laure Mignerot avec que j'ai eu plaisir à travailler et échanger, mais aussi les personnes de l'UMR 8227, Simon Dittami, Ludovic Delage, Catherine Leblanc, Thierry Thonon, Catherine Boyen, Lionel Cladière et bien d'autres.

Egalement, un grand merci aux membres de la plateforme ABiMS, Gildas Le Corguillé, Misharl Monsoor, Xi Liu, Loraine Guéguen, Camille Vacquié, Julien Wollbret, Wilfrid Carré, Gwendoline Andrès, Eric Duvignac, Jean-Michel Aroumougom, Joseph Kervellec, Antoine Bisch, Nolwenn Dantec, Mark Hoebeke, Olivier Quenez et Ludovic Legrand pour m'avoir accueillis et soutenu pendant ces trois ans. Je tiens tout particulièrement à remercier le responsable de la plateforme, Christophe Caron, pour son soutien et son aide tout au long de ma thèse, mais aussi lors de mes différents séjours à la Station.

Ik wil mijn dank uitspreken aan het hele Bioinformatics & Evolutionary Genomics team van de Gentse VIB. Vooral aan Lieven Sterck, Yves van de Peer, Pierre Rouzé, Li Zhen en Stéphane Rombouts. Bedankt voor de twee maanden die ik in het team heb kunnen doorbrengen om een deel van mijn thesis projekt te voltooien, en voor de voortreffelijke momenten tijdens de Brainstorming.

Een grote dank aan Lieven Sterck voor de hulp, de steun, het advies die hij me gedurende deze drie jaar heeft gegeven.

Merci à tous mes amis, Océane, Ronflex, Manide, Obivalion, Maxime, Paulo, Thibaut, Lucas et bien d'autres pour m'avoir soutenu et encouragé.

Un grand merci à Laurence Cladière pour son soutien, son intérêt pour ma thèse et ses conseils.

Je remercie spécialement Léa Cabioch, pour m'avoir soutenu, supportée et qui a adoucit les derniers jours de ma thèse.

Enfin merci à ma famille pour m'avoir permis d'arriver jusque-là et qui m'a toujours soutenu dans ce que j'entreprenais et pour m'avoir donné les moyens d'y arriver.

Table des matières

Préambule

Les algues brunes représentent un large groupe d'organismes qui comprend environ 1800 espèces réparties en 265 genres et 14 ordres. Elles sont majoritairement présentes dans le milieu marin et plus rarement dans le milieu dulçaquicole (seulement 5 genres étant trouvés dans ce milieu). Les espèces marines se trouvent principalement dans les eaux tempérées à froides au niveau de la zone intertidale et dans la partie infralittorale, pouvant aller jusqu'à des profondeurs de 220 m si le niveau de turbidité de l'eau est suffisamment faible. Elles vivent attachées sur des substrats rocheux, mais peuvent, en fonction de leur taille, se fixer sur d'autres types de substrats comme les digues, les quais, des mollusques, des zostères ou sur d'autres algues. Elles présentent une morphologie très variable allant de filaments microscopiques à des individus pouvant atteindre jusqu'à 60 mètres de long (*Macrocystis*) et possédant une structure complexe avec plusieurs types cellulaires. Elles jouent un rôle important dans la biodiversité et l'écologie des océans (Dayton 1985; Steneck et al. 2002; Bartsch et al. 2008) comme composante essentielle des écosystèmes côtiers et peuvent former des couvertures denses et étendues telles que les « forêts » de laminaires. La coloration brunâtre de ces algues est liée à la présence de pigments accessoires dans les chloroplastes, les xanthophylles et principalement la fucoxanthine (Paillard 2011).

L'utilisation par l'homme des algues brunes est courante dans des secteurs économiques variés, allant de la pharmacologie, la cosmétologie, en passant par l'agriculture, dans l'agroalimentaire - de manière brute ou transformée - et dans l'industrie textile (Kijjoa and Sawangwong 2004; Smit 2004; A.D. Hughes et al. 2012). En utilisation directe, pour l'alimentation humaine, elle se fait principalement dans les pays asiatiques comme la Chine et le Japon. De manière transformée, on connait surtout leur utilisation via l'alginate, un polysaccharide obtenu majoritairement à partir des laminaires. Les alginates sont utilisés par les industriels comme texturants polyvalents, épaississants, agents gélifiants, stabilisants, cryoprotecteurs (pour les aliments surgelés) et films comestibles.

Les algues brunes font partie du groupe de Phaeophyceae appartenant au clade des Hétérocontes, autrement appelé Stramenopiles (Baldauf 2003). Elles sont l'un des 5 groupes d'eucaryotes qui ont évolué vers la complexité multicellulaire de manière indépendante, avec les animaux, les plantes vertes, les champignons et les algues rouges (Cock et al. 2010). Cette histoire

évolutive indépendante a fait que les algues brunes ont développé un large éventail de nouveaux processus au niveau métabolique, cellulaire et écologique, qui sont rares ou absents dans les autres clades. Par exemple, la modification des pigments photosynthétiques, avec la présence majoritaire de la fucoxanthine, afin de tenir compte de l'absorption d'une partie du spectre lumineux par l'eau lors de l'immersion des algues (Charrier et al. 2008). Ou bien encore une voie métabolique des halogènes spécifiques (La Barre et al. 2010) et un stockage du carbone, accumulé lors de la photosynthèse, sous forme soluble, dans la laminarine (Stewart 1974) ou bien le mannitol (Davis et al. 2003).

L'importante distance phylogénétique et cette histoire évolutive indépendante par rapport aux autres espèces eucaryotes ont mis en évidence la nécessité d'identifier et de développer un modèle biologique pour répondre aux questions spécifiques de la biologie des algues brunes et pour fournir un support pour les applications de génétique et de génomique. Dans ce cadre, *Ectocarpus* a été proposé comme modèle pour le groupe des algues brunes en 2004 (Peters et al. 2004) et un projet de séquençage du génome d'*Ectocarpus siliculosus* a été initié. Le choix d'*Ectocarpus* comme modèle a été motivé par le fait qu'elle soit étudiée depuis de nombreuses années (Dillwyn 1809; Müller 1967; Müller 1976; Bolton 1983; Schmid and Dring 1993; Maier 1995; Silva et al. 1996; Maier 1997a; StacheCrain et al. 1997; Maier 1997b; Busch and Schmid 2001; Peters et al. 2004). Elle est facilement cultivable en laboratoire et son cycle de vie peut être complété dans une période comprise entre 2 et 3 mois en boîte de Pétri (Peters et al. 2004; Charrier et al. 2008). Enfin, *Ectocarpus* possède un génome relativement petit (215 Mpb) par rapport à d'autres algues brunes.

Le projet génome s'est achevé en 2010 avec la mise à disposition du génome et de l'annotation, représentant une première étape donnant accès à une importante source d'information (Cock et al. 2010). La création d'une carte génétique a permis l'organisation des super-contigs en 34 groupes de liaison (Heesch et al. 2010), valeur qui se rapproche d'une estimation empirique d'approximativement 25 chromosomes (Müller 1967). De nombreux autres outils et ressources ont été développés autour de ce modèle depuis la disponibilité du génome, incluant par exemple des puces micro-array (Coelho et al. 2011a; Dittami et al. 2011), des EST (Expressed Sequence Tag), un tiling array, des techniques de protéomique et des outils bio-informatiques comme la mise en place d'un réseau métabolique (Prigent et al. 2014) ou d'un logiciel de prédiction de la localisation subcellulaire des protéines (Gschloessl et al. 2008).

L'adaptation de ces techniques au génome d'*Ectocarpus* a permis d'explorer diverses questions biologiques parmi lesquelles, l'étude de la détermination et de l'évolution du sexe. L'un des premiers

défis afin de répondre à cette question a été d'identifier le chromosome sexuel mâle dans le génome de référence d'*Ectocarpus*. Cette thèse a été initiée sur ce premier objectif et s'est poursuivie avec l'identification du chromosome femelle et l'analyse comparative de la structure des deux chromosomes. La disponibilité du génome complet a permis de réaliser des analyses d'expression différentielle des gènes entre individus mâles et femelles, d'identifier les gènes différentiellement exprimés et d'analyser leur évolution moléculaire. La disponibilité de nombreuses données de séquençage afin de répondre à cette question a en outre permis de proposer une nouvelle version de l'annotation structurale et fonctionnelle des gènes d'*Ectocarpus*.

Objectifs

L'objectif de cette thèse était d'étudier les mécanismes à l'origine de la détermination du sexe chez l'algue brune *Ectocarpus*. Cette thèse comprend deux parties distinctes. Une première partie centrée sur l'identification et l'analyse de la structure des chromosomes sexuels ainsi que l'identification et la caractérisation des gènes différentiellement exprimés entre mâles et femelles. La seconde partie est focalisée sur l'amélioration de l'annotation structurale du génome d'*Ectocarpus*.

Plus spécifiquement, les objectifs étaient les suivants :

- Compléter l'identification des régions spécifiques au mâle et à la femelle dans le génome d'*Ectocarpus* et réaliser une analyse structurale des chromosomes sexuels mâle et femelle par l'étude de la structure génétique (i.e. la présence et le type des éléments transposables, la densité de gènes, la structure des gènes, la présence de pseudogènes, etc.) en comparaison avec la structure de la région pseudo-autosomique et les autosomes (Chapitre 1 – article 1 et 2).
- Réaliser une étude comparative des transcriptomes mâles et femelles afin d'identifier les gènes exprimés spécifiquement ou préférentiellement dans l'un des deux sexes à plusieurs stades du cycle de vie d'*Ectocarpus* (Chapitre 1 – article 3).
- Réaliser une nouvelle annotation structurale des gènes d'*Ectocarpus* afin d'améliorer et compléter l'annotation du génome par l'apport d'informations liées aux nouvelles technologies de séquençage (Chapitre 2 – Article 1).
- Valider la présence de micro-ARN chez *Ectocarpus* et analyser les caractéristiques au niveau de l'expression différentielle entre mâles et femelles et des cibles de micro-ARN identifiés (Chapitre 2 – Article 2).

Chaque chapitre du manuscrit commencera par une introduction, suivie de plusieurs sous parties correspondant à chaque article publié ou en cours de publication. Enfin, une discussion et conclusion générale sur l'ensemble des chapitres sera proposée en fin du manuscrit.

Chapitre 1 : Evolution et analyse fonctionnelle de la détermination du sexe chez les algues brunes

Introduction

Dans le contexte de l'analyse de l'apparition et de l'évolution du sexe chez les Eucaryotes, il faut distinguer la reproduction sexuée du déterminisme sexuel. La reproduction sexuée, appelée aussi sexe méiotique, est un phénomène extrêmement répandu qui assure la production de nouvelles combinaisons génétiques dans presque toutes les lignées eucaryotes, même ancestrales. C'est un processus qui se déroule en deux étapes, la première, la syngamie (i.e. fusion de deux cellules haploïdes), va engendrer la formation d'un zygote diploïde qui lors de la seconde étape, la méiose, va générer la formation de nouvelles cellules haploïdes et ainsi compléter le cycle. L'apparition de la reproduction sexuée serait très ancienne et remonterait au début de l'évolution des eucaryotes (Cavalier-smith 2002), mais nos connaissances sont très limitées et son origine reste l'une des grandes énigmes de la biologie (Speijer et al. 2015).

Les mécanismes du déterminisme du sexe, quant à eux, i.e. le développement d'individu vers l'un ou l'autre des deux types de sexe (mâle ou femelle), ont émergé de manière indépendante et répétée au sein de plusieurs lignées d'Eucaryotes et les voies qui déterminent les spécificités sexuelles sont très variées.

La détermination des sexes chez les Eucaryotes

L'apparition à de multiples reprises de systèmes de détermination du sexe a permis de voir émerger une grande diversité de mécanismes. Cependant, ils peuvent être regroupés en deux catégories de détermination du sexe, épigénétique ou bien génétique. Même si certaines espèces, comme le zebrafish, présentent un état intermédiaire entre ces deux grandes catégories, avec un système de détermination du sexe génétique mais qui peut être influencé par des facteurs abiotiques (Liew and Orbán 2014).

Détermination épigénétique du sexe

Le premier type de détermination sexuelle est lié à des facteurs épigénétiques (Détermination Epigénétique du Sexe - DES) de type abiotique ou bien abiotique. Ce type de déterminisme du sexe est phylogénétiquement dispersé et retrouvé dans divers taxons tels que les plantes, les nématodes, les amphipodes, les mollusques, les poissons ou encore les vertébrés amniotes (**Figure 1**) (Janzen and Phillips 2006). Dans le cas d'une influence abiotique, le sexe n'est pas déterminé au moment de la fécondation, mais durant le développement de l'embryon (Bull 1983). Cette influence abiotique peut être liée à la température, la disponibilité des ressources, le pH ou encore la photopériode. Par exemple, le cas le plus connu est celui de la détermination du sexe par la température lors du développement de l'embryon chez l'ensemble des crocodiliens et des rhynchocéphales (Janzen and Phillips 2006), pour certaines tortues (Pieau et al. 1994) et poissons (Godwin et al. 2003). Chez le Copépode *Pachypygus gibber*, la faible ressource en nourriture entraîne une augmentation du nombre d'individus mâles (Becheikh et al. 1998). Des influences de type biotique peuvent aussi être à l'origine d'une modification du sexe au cours de la vie d'un individu. Chez la crépidule, le sexe est déterminé par la position de l'individu dans la population. Un individu sur un substrat sans congénères à proximité se développera en femelle. A l'inverse, si l'individu s'installe sur un substrat avec une colonie déjà établie ou directement sur ses congénères, il se développera en mâle (Wright 1988; Proestou et al. 2008). Chez certains poissons, la proportion d'un sexe par rapport à l'autre favorisera le développement de l'un des deux sexes (Godwin et al. 2003).

L'avantage théorique de ce système est la capacité donnée à la population de pouvoir s'adapter dans des environnements « inégaux » en modifiant la structure de la population afin d'optimiser le fitness (Bull 1985). En d'autres termes, un individu adoptera le sexe qui possède la plus grande capacité à survivre et à transmettre son patrimoine génétique dans l'environnement où il se situe. Cependant, ce type de déterminisme du sexe a un coût, comme le développement de l'intersexualité et un biais au niveau du sexe ratio. De plus, dans le cas des systèmes dépendant de la température, le sexe ratio est tributaire des conditions environnementales et de la sélection du site de nidification par le parent.

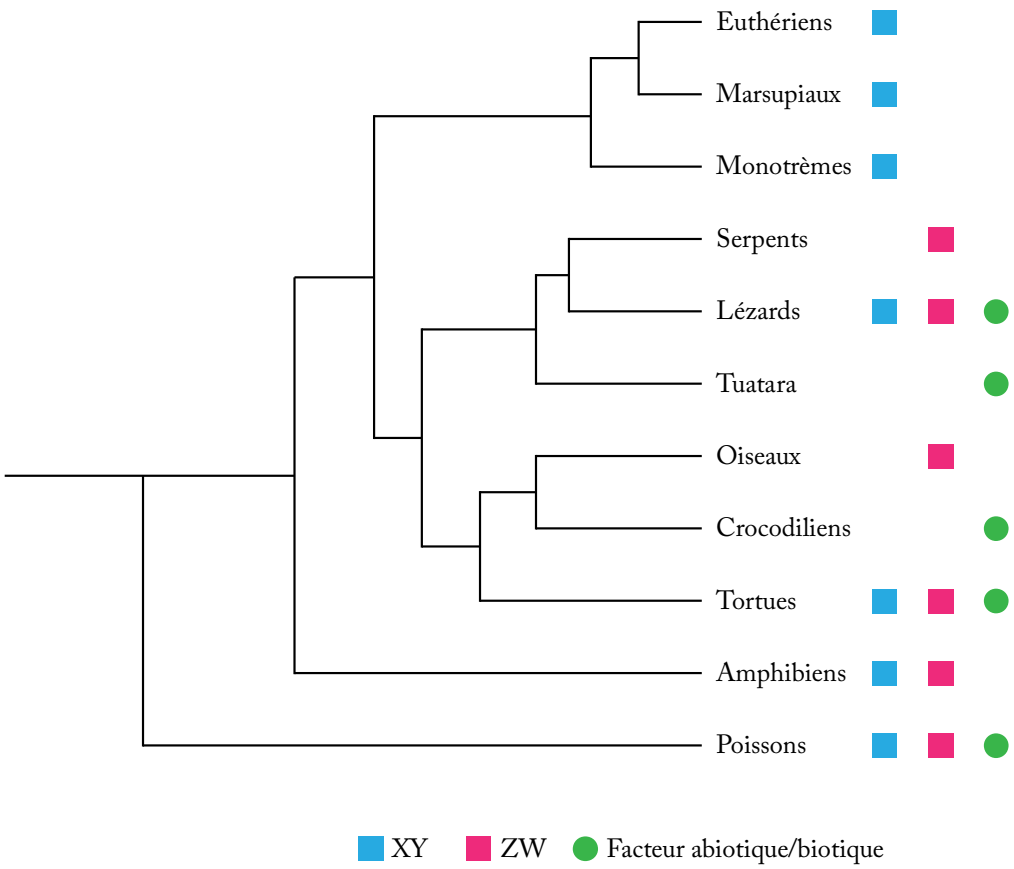


Figure 1 : Répartition des différents systèmes de détermination du sexe chez les vertébrés (Ezaz et al. 2006).

Détermination génétique du sexe

La détermination du sexe peut aussi être liée à des facteurs génétiques (Détermination Génétique du Sexe - DGS). Plusieurs systèmes existent, le modèle classique étant le contrôle de la détermination du sexe de manière monogénique avec des systèmes chromosomiques tels que XY chez les mammifères et ZW chez les oiseaux ou bien UV chez certaines algues ou bien l'hépatique *Marchantia polymorpha,* pour lesquelles un unique locus contrôle la détermination (**Figure 1**). Chez le papillon bombyx (ZW), le chromosome sexuel femelle totalement dégénéré ne possède pas de gènes codant pour des protéines, mais une protéine non codante, un piRNA (Piwi-interacting RNA) qui va déterminer le sexe par l'activation de l'isoforme femelle du gène Bmdsx (Kiuchi et al. 2014).

D'un autre côté, le système de polygénétique fait intervenir plusieurs loci indépendants ou bien des combinaisons d'allèles afin de déterminer le sexe (**Figure 2**) (Kosswig 1964). Ce type de détermination est trouvé dans une grande diversité de groupes phylogénétiques comme les insectes, les mammifères, les poissons ou encore chez les plantes (Moore and Roberts 2013). Par exemple chez la souris naine d'Afrique (*Mus minutoides*), le système de détermination est de type XYW. L'apparition

de la copie W provient de l'évolution d'un chromosome X et est caractérisée par l'acquisition d'une mutation contre la masculinisation (Veyrunes et al. 2010). Des systèmes plus complexes, faisant intervenir différents allèles, sont observés chez des espèces de poissons de la famille des Cichlidae (Ser et al. 2010). Cependant, le fait de posséder plusieurs chromosomes sexuels ne donne pas forcément un système de détermination du sexe polygénétique. L'ornithorynque possède 5 X et 5 Y, mais la détermination sexuelle est identique au système XY monogénétique grâce à une ségrégation commune des X et des Y lors de la méiose (Grützner et al. 2004). Les travaux théoriques sur les deux systèmes tendent à montrer que le système polygénique multiloci serait une première étape de transition du passage de l'hermaphrodisme au système monogénétique (Rice 1987; Bachtrog et al. 2014).

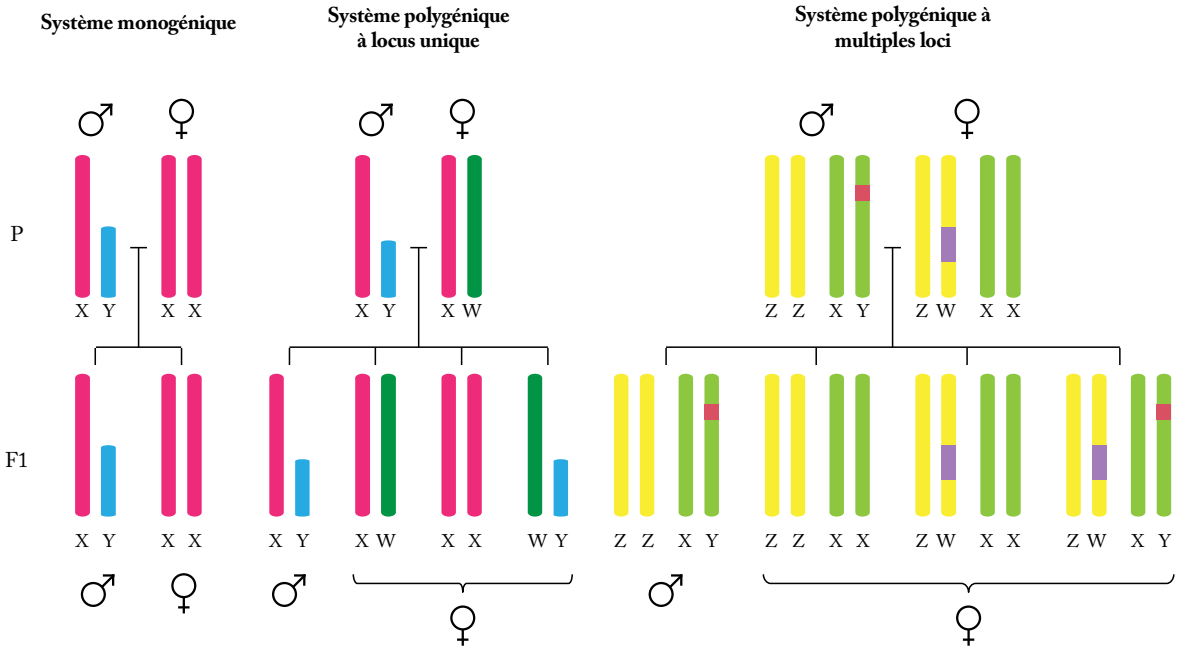


Figure 2 : Modèle de transmission du sexe dans le système monogénétique XY et les systèmes polygénétiques. Adapté de (Moore and Roberts 2013). Dans le système XX/XY, le sexe dans la progéniture est déterminé par le parent mâle qui porte le chromosome Y. Dans les systèmes polygénétiques XYW, le sexe de la progéniture est aussi porté par le chromosome Y, mais son effet est annulé par le second chromosome sexuel femelle W. Dans un système multilocus, plusieurs allèles (en rouge et violet) sont présents et situés sur différents chromosomes dont la ségrégation est indépendante. L'allèle présent sur le chromosome W annule l'effet de l'allèle sur le chromosome Y, déterminant pour le mâle. De fait, les individus ZW / XY sont des femelles. De même que dans le système polygénétique à simple locus, il en résulte la formation d'un seul type génotypique mâle et de trois types génotypiques femelles, l'absence des allèles donnant un individu femelle.

Les différents cycles de vie sexués chez les Eucaryotes

Les types de détermination du sexe, présentés précédemment, sont étroitement corrélés au cycle de vie de l'espèce. Chez les Eucaryotes, le cycle de vie est caractérisé par l'alternance de deux phases, diploïde et haploïde. La transition de la phase diploïde à la phase haploïde est réalisée par la méiose tandis que la transition de la phase haploïde vers la phase diploïde est accomplie grâce à la syngamie ou autrement appelée fusion des gamètes. Cependant, il existe des variations entre les espèces, principalement liées à la durée de chaque phase et au niveau de l'activité mitotique. Ces différences peuvent être catégorisées en trois variantes du cycle de vie : le cycle de vie à majorité haploïde, le cycle de vie à majorité diploïde et le cycle de vie à alternance des générations ou cycle haploïde-diploïde.

Le cycle de vie haploïde est caractérisé par la dominance de la phase haploïde sur la phase diploïde (**Figure 3a**). La phase diploïde est réduite au minimum sans développement d'un individu multicellulaire, le zygote faisant rapidement la méiose, produisant ainsi des méiospores. Les méiospores se développent en organismes multicellulaires qui produisent des gamètes haploïdes, fusionnant pour redonner un zygote. Ce type de cycle de vie est observé chez les organismes tels que les champignons ou encore les algues vertes.

Le cycle de vie diploïde est caractérisé par la dominance de la phase diploïde sur la phase haploïde (**Figure 3b**). La syngamie engendre le développement d'un organisme multicellulaire diploïde, qui, selon les espèces, peut produire des individus avec des sexes séparés ou bien des individus hermaphrodites. Comme expliqué dans la section précédente, le sexe peut être déterminé de manière environnementale ou de manière génétique. La phase haploïde est réduite à l'état de gamètes. Ce type de cycle de vie est observé principalement chez les Opisthocontes, certaines algues vertes et brunes ou encore chez les Alvéolés et les Excavés.

Le cycle de vie haploïde-diploïde est caractérisé par une alternance du développement d'individus haploïdes, appelé gamétophytes, et d'individus diploïdes, les sporophytes (**Figure 3c**). Après la méiose chez le sporophyte, les méiospores engendrent le développement de deux individus haploïdes, un mâle et une femelle. La fusion de leurs gamètes permet le développement d'un nouveau sporophyte. Ce type de cycle de vie est observé chez les algues rouges, les plantes terrestres, certaines algues brunes ou chez les champignons.

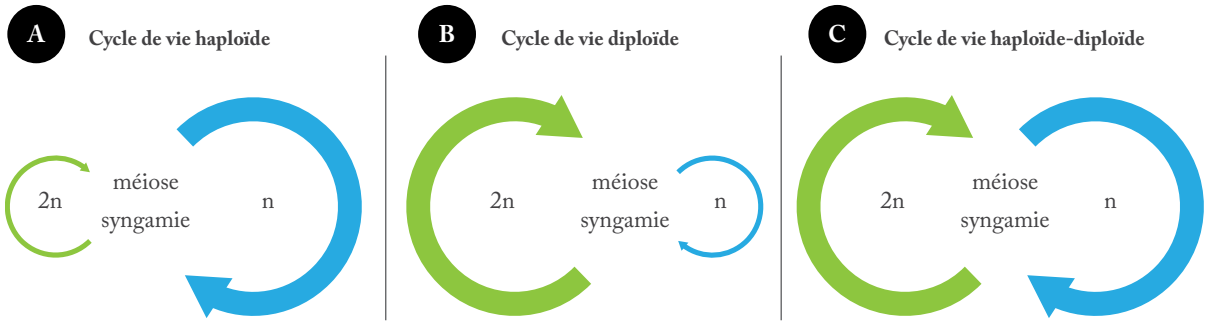


Figure 3 : Les principaux cycles de vie. Selon la phase dominante, entre haploïde et diploïde, on identifie trois catégories de cycle de vie. A) cycle de vie haploïde B) cycle de vie diploïde C) cycle de vie haploïde-diploïde. La flèche en gras représente la phase dominante.

Les chromosomes sexuels : origine et évolution

Les différents types de chromosomes sexuels

La détermination génétique du sexe monogénétique chez les Eucaryotes inclut trois grands types de systèmes chromosomiques sexuels : les systèmes de détermination sexuelle diploïde XY et ZW, et le système de détermination sexuelle haploïde UV (**Figure 4**).

Dans les systèmes diploïdes XY et ZW, le sexe est déterminé au moment de la fusion des gamètes. Le système XY est principalement observé chez les mammifères tandis que le système ZW est principalement observé chez les oiseaux et les reptiles. Cependant, les deux systèmes peuvent cohabiter au sein d'un même clade, auquel peut s'ajouter la présence d'un déterminisme sexuel biotique, par exemple chez les poissons (Ezaz et al. 2006). Le système XY est défini comme étant hétérogamétique mâle, ce dernier portant les deux chromosomes sexuels X et Y, le Y étant spécifique au mâle et le X est hérité de la mère, tandis que la femelle porte deux copies du chromosome femelle X (**Figure 4a**). Dans le système ZW, contrairement au système XY, l'hétérogamétie est femelle. Cette dernière porte les deux chromosomes sexuels Z et W, le chromosome W est spécifique à la femelle et la progéniture femelle hérite toujours du chromosome Z de leur père. Le mâle, lui, hérite du chromosome Z de chacun des deux parents (**Figure 4b**). Pour ces deux types de détermination génétique du sexe, ce dernier est exprimé durant la phase diploïde, phase qui est très largement majoritaire au niveau de la durée par rapport à la phase haploïde, réduite à la vie des gamètes. Il s'agit de systèmes de base, où de nombreuses variations peuvent être observées en termes de nombre de chromosomes sexuels. Par exemple chez l'ornithorynque, la femelle possède cinq paires de chromosomes X tandis que le mâle possède cinq chromosomes X et cinq chromosomes Y (Ferguson-Smith and Rens 2010).

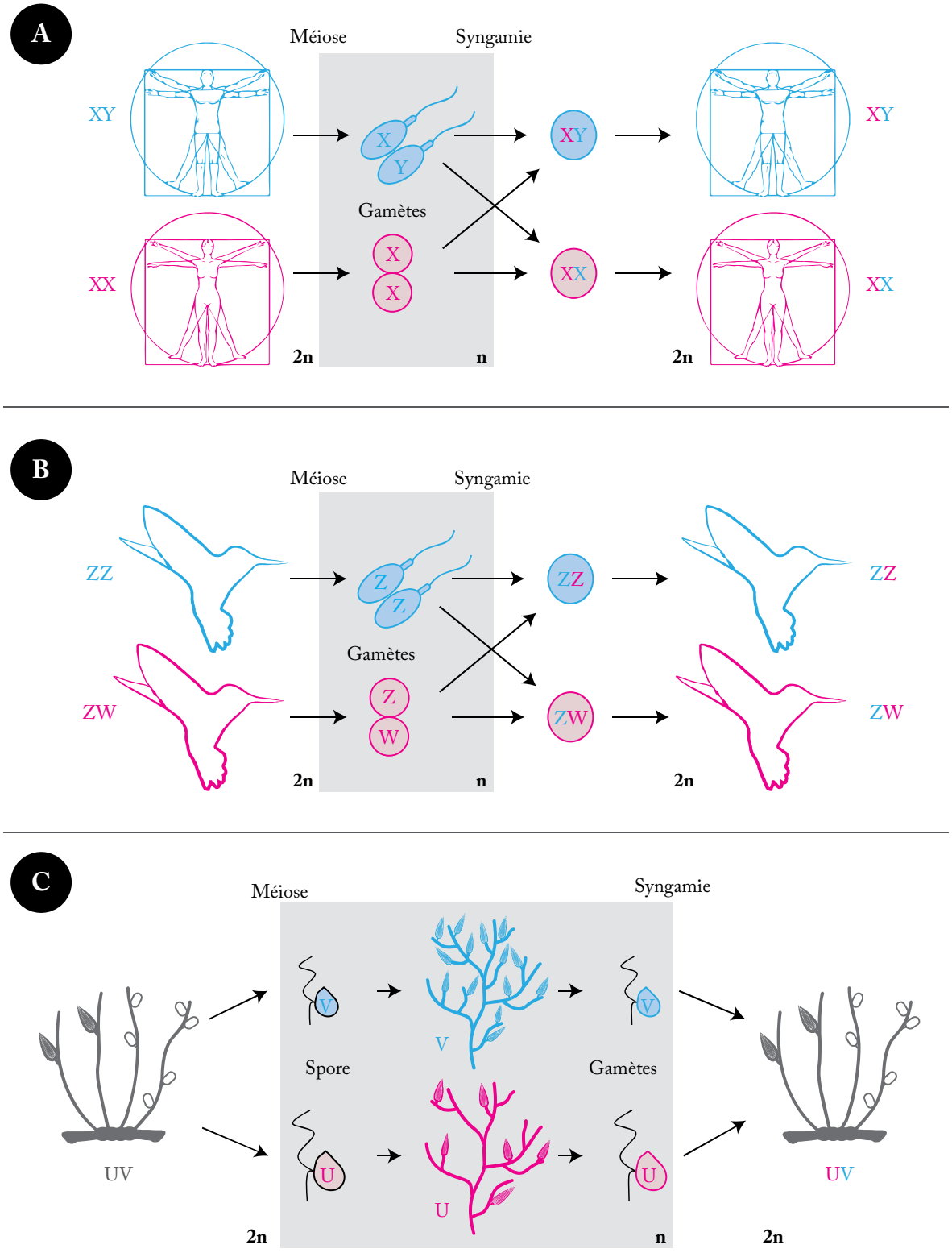


Figure 4 : Les trois principaux types de chromosomes sexuels chez les Eucaryotes. Adapté de (Bachtrog et al. 2014). A) Le système XY dans lequel le chromosome Y est spécifique au mâle. B) Le système ZW dans lequel le chromosome W est spécifique à la femelle. Pour ces deux systèmes, le sexe est exprimé durant la phase diploïde. C) Le système UV dans lequel le chromosome V est spécifique au mâle et le chromosome U est spécifique à la femelle. Dans ce système, le sexe est exprimé durant la phase haploïde.

Le système de détermination sexuelle chromosomique haploïde UV est, quant à, lui présent chez les algues et les bryophytes (Bachtrog et al. 2011). Dans le système UV, le sexe n'est pas déterminé au moment dans la fertilisation, contrairement aux systèmes diploïdes XY et ZW, mais le sexe de la descendance méiotique est déterminé par celui qui porte le chromosome femelle (U) ou le chromosome mâle (V) après la méiose, en fonction du chromosome sexuel reçu par la spore (**Figure 4c**). Il n'y a donc pas de sexe homogamétique, et les chromosomes U et V sont toujours hémizygotes durant la phase diploïde (UV).

Contrairement aux deux autres systèmes de chromosomes sexuels, les connaissances et les données disponibles pour le système UV sont relativement limitées. Bien que les espèces Eucaryotes qui possèdent un système UV soient probablement aussi communes que celles du type XY et ZW, très peu d'espèces du système UV ont été jusqu'à présent caractérisées (Immler and Otto 2015).

Formation et évolution des chromosomes sexuels

Les théories actuelles sur l'origine des chromosomes sexuels ont été émises dans les années 1980 et sont régulièrement revues afin de tenir compte des derniers résultats (Charlesworth and Charlesworth 1980; Bull 1983; Charlesworth et al. 2005). Pour en faciliter la compréhension, l'explication sera, dans un premier temps, basée sur le modèle XY, pour les systèmes de détermination sexuelle diploïde.

Les théories prédisent que l'évolution des chromosomes sexuels se fait à partir d'une paire d'autosomes par l'apparition d'un locus de la détermination du sexe, via par exemple l'acquisition d'un gène déterminant pour le mâle (**Figure 5**). L'apparition d'un gène de stérilité pour les femelles sur le proto-Y et d'un gène promoteur pour les mâles sur le proto-Y permet la transition d'une population jusque-là hermaphrodite vers une population sexuée. La présence de ces gènes dans les chromosomes induit la mise en place de mécanismes de réduction de la recombinaison, tels que des réarrangements chromosomiques comme les inversions, les transpositions et l'accumulation d'éléments transposables (Bergero and Charlesworth 2009). La suppression de la recombinaison engendre l'apparition d'une région spécifique au sexe (Sex Determining Region - SDR) afin d'éviter la recombinaison entre les loci déterminants pour le sexe, ce qui pourrait causer la stérilité ou bien une réversion vers l'hermaphrodisme. A court terme, ces modifications peuvent entraîner une augmentation de la taille du Y, pouvant devenir plus grand que le X.

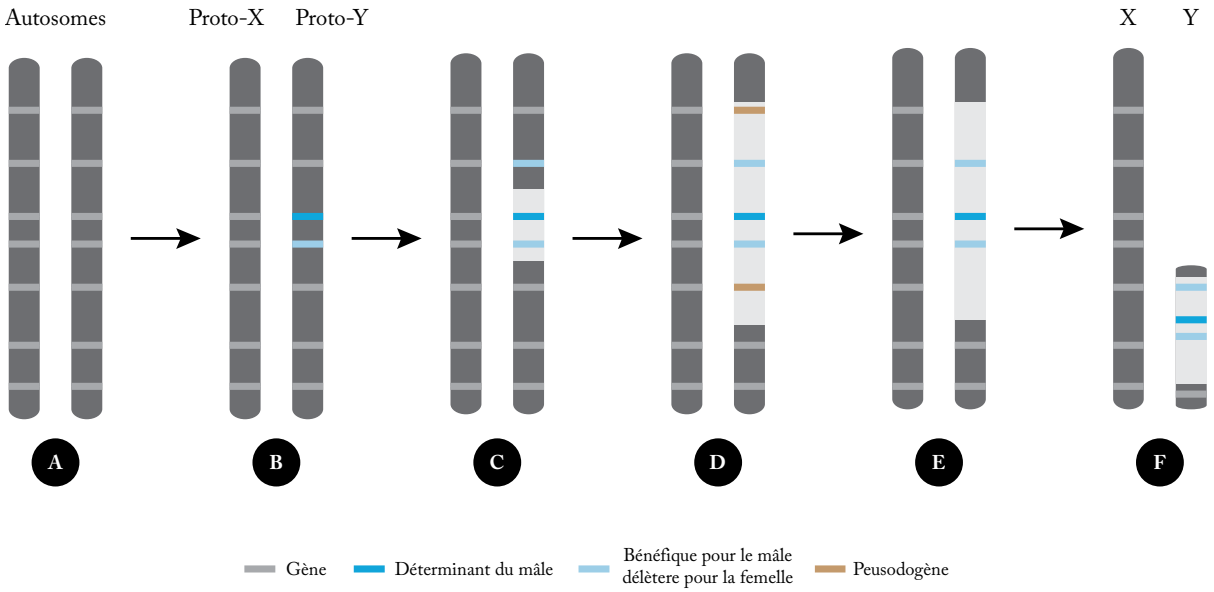


Figure 5 : Modèle d'évolution des chromosomes sexuels dans le système XY (Bachtrog 2013). a) Les chromosomes sexuels sont formés à partir d'une paire d'autosomes. b) Dans une population hermaphrodite, l'acquisition d'un locus déterminant pour l'un des sexes, mâle par exemple, associé à une mutation de stérilité pour la femelle, permet l'émergence de sexes séparés, via la formation d'un proto-X et d'un proto-Y. c) L'accumulation de mutations sexuellement antagonistes au niveau de la région déterminante pour le sexe favorise la cessation de la recombinaison et son extension. d) La région déterminante pour le sexe peut s'étendre par l'accumulation de mutations, l'intégration de gènes déterminant pour le mâle et principalement les évènements d'inversions dans le chromosome Y. L'accumulation de mutations par le manque de recombinaison peut entraîner un phénomène de pseudogénisation dans le chromosome Y. e) L'accumulation d'éléments répétés liée à l'arrêt de la recombinaison fait qu'il est possible que la taille globale du chromosome Y dépasse celle du chromosome X. f) La dégénérescence génétique due à l'absence de recombinaison peut engendrer la perte de portions d'ADN non fonctionnel chez le chromosome Y et entraîner une diminution de la taille de ce dernier.

La SDR peut intégrer de manière progressive des loci bénéfiques pour le développement du mâle, dans le cas du système XY. Ces loci présentant un antagonisme sexuel (Rice 1996) – i.e. qu'ils sont bénéfiques pour la mâle, mais désavantageux, voire néfastes, pour la femelle - favorisent l'insertion de ces régions dans la SDR mâle. Cette accumulation progressive peut laisser des traces au cours de l'évolution, des strates, qui représentent les régions qui ont été incorporées dans la partie non recombinante du chromosome sexuel à différentes périodes. Étant donné que ces régions ne sont pas intégrées aux mêmes périodes évolutives, et les contraintes n'étant pas les mêmes au sein de la région recombinante, il est possible de distinguer les différentes régions intégrées par l'analyse de leurs niveaux de divergence évolutive (**Figure 6**). L'intégration progressive de ces régions bénéfiques pour le mâle dans le chromosome Y peut mener à une extension de la région spécifique du sexe (SDR) à occuper une majeure partie du chromosome. Les régions pseudo-autosomales (PAR) ont alors une

taille limitée, qui permet de maintenir l'appariement des deux chromosomes durant la méiose (Charlesworth et al. 2005). Des strates évolutives ont été identifiées et étudiées dans un certain nombre d'organismes tels que les animaux (Lahn and Page 1999; Vicoso et al. 2013; Wright et al. 2014) et les végétaux (Bergero et al. 2007; Wang et al. 2012).

La suppression de la recombinaison engendre l'accumulation de mutations délétères dans la séquence nucléotidique, la perte de fonction de certains gènes engendrant la formation de pseudogènes. A long terme, l'accumulation de mutations peut entraîner une diminution de la taille du chromosome sexuel hétérogamétique Y lié à la dégénérescence génétique. A l'inverse, la maintenance de la recombinaison chez les chromosomes homomorphiques X permet d'éviter la dégénérescence de ces derniers et explique leur maintenance par l'élimination des mutations délétères et des différents éléments répétés.

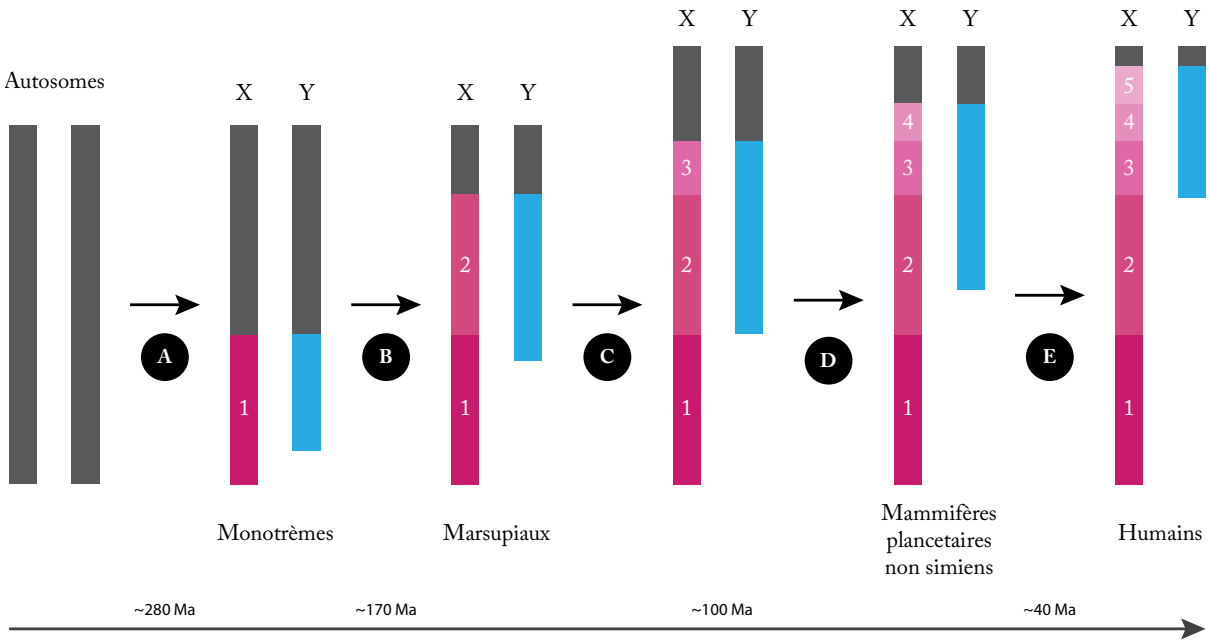


Figure 6 : Modèle d'évolution ayant amené à la présence de cinq strates évolutives dans le chromosome X chez l'humain. Adapté de (Lahn and Page 1999; J.F. Hughes et al. 2012). Chaque inversion réduit la taille de la partie pseudo-autosomale par la suppression de la recombinaison entre le X et le Y A) Emergence des chromosomes sexuels XY chez les mammifères, première inversion du chromosome Y (240 – 320 Ma). B) Deuxième inversion dans le chromosome Y (130 – 170 Ma). C) Expansion de la région pseudo-autosomale et troisième inversion dans le chromosome Y (80 – 130 Ma). D) Quatrième inversion dans le chromosome Y (80 – 130 Ma). E) Cinquième inversion dans le chromosome Y (30 – 50 Ma). En noir, les parties recombinantes. En rose et bleu, les parties non recombinantes chez les individus, respectivement femelles et mâles.

Dans le système UV, le sexe est exprimé au stade haploïde, les femelles portant le chromosome U et les mâles le chromosome V. Les travaux théoriques prédisent que les chromosomes

U et V devraient présenter des caractéristiques similaires, à savoir une taille similaire, la même vitesse et le même degré de dégénérescence (Bull 1978; Immler and Otto 2015). De même que dans le système XY, la suppression de la recombinaison durant la phase diploïde entraîne l'accumulation de mutation délétère. Mais contrairement au système XY, les théories suggèrent que le temps relativement important passé par les chromosomes U et V dans la phase haploïde engendre une dégénérescence plus lente des SDR par l'exposition des chromosomes à une sélection purificatrice (purifying selection). Quant à l'évolution de la taille des chromosomes U et V, elle serait liée à une intégration et non une perte de régions génomiques, contenant des gènes bénéfiques pour le stade haploïde (Bull 1978).

Bien que le système UV soit décrit depuis longtemps (Allen 1917), ces hypothèses restent non testées du fait que peu de données empiriques soient disponibles pour les organismes avec un système UV (McDaniel et al. 2007; Yamato et al. 2007; McDaniel et al. 2013), comparé aux autres systèmes.

Structure des chromosomes sexuels

A travers les clades et les espèces, il existe une grande variation de la structure interne des chromosomes sexuels. Comme montré précédemment, dans beaucoup d'espèces, les chromosomes sexuels incluent une partie non recombinante, appelée région déterminant le sexe (Sex Determining Region - SDR) qui est bordée par une ou deux parties appelées régions pseudo-autosomiques (Pseudo Autosomal Region - PAR) qui sont encore capables de recombiner.

La PAR est particulièrement importante pour maintenir un appariement et une ségrégation correctes des chromosomes sexuels durant la méiose, ce qui explique sa persistante dans une grande majorité des espèces (Rouyer et al. 1986). Cette présence nécessaire peut entraîner un niveau de recombinaison par base supérieur par rapport à celui observé chez les autosomes, et en conséquence, les régions distales de la PAR pourraient approcher le modèle de transmission des autosomes, avec une chance égale d'être transmise à chaque sexe. La comparaison réalisée entre les plantes et les animaux sur le nombre de PAR montre que, le plus souvent, ces derniers ne possèdent qu'une seule PAR. Etant donné que les chromosomes sexuels des animaux sont généralement plus âgés que ceux des plantes, il semblerait que la présence de deux PAR soit plus importante chez les chromosomes sexuels récents (Otto et al. 2011). Cependant, elle reste facultative et pour des espèces comme les marsupiaux, elle est absente et l'appariement des chromosomes sexuels est achiasmatique (i.e. dont la méiose n'implique pas la recombinaison entre la paire de chromosomes) (Patel et al. 2010). Pour

d'autres espèces comme les Diptères et les Lépidoptères pour lesquelles la recombinaison est absente chez les mâles (hétérogamétique), la PAR n'est pas présente (Gethmann 1988).

La taille de la PAR est variable en fonction des espèces. Les chromosomes sexuels dont l'histoire évolutive est plus récente ont tendance à posséder des PAR d'une taille plus importante que les chromosomes sexuels plus anciens, mais la corrélation entre âge et taille de la PAR reste cependant à vérifier (Charlesworth et al. 2005; Otto et al. 2011). Ainsi, chez les espèces ayant récemment acquis un chromosome sexuel, une grande variation de taille de la PAR peut être observée, allant d'une taille plutôt limitée comme chez l'épinoche (*Gasterosteus aculeatus*) où la PAR représente 15,8 % (Ross and Peichel 2008) à une taille plus importante chez la papaye avec une PAR qui représente 83 % du chromosome sexuel (Yu et al. 2009). Chez les espèces ayant une histoire évolutive plus ancienne au niveau des chromosomes sexuels, comme chez la souris, la PAR ne représente que 1 % de la taille totale du chromosome Y et chez l'homme, 4,6 %. La disponibilité de nombreuses données pour le groupe des oiseaux a permis de montrer qu'il y a une forte variation de la taille de la PAR au sein de ce groupe et entre les différentes lignées de ce groupe et entre les différentes espèces (Zhou et al. 2014). Par exemple, dans le groupe des paléognathes, pour l'emu et l'autruche, 65 % du chromosome sexuel Z est toujours recombinant, tandis que seulement 1 % l'est chez le tinamou à gorge blanche.

Cette évolution qui tend vers la diminution de la PAR est consistante avec l'hypothèse de l'antagonisme sexuel pour l'évolution de la SDR (i.e. que plus les gènes sexuellement antagonistes vont migrer vers la SDR au cours du temps, plus la taille de la SDR va prendre le pas sur la taille de la PAR). La pression de sélection spécifique au sexe agissant dans la sélection différentielle des gènes chez le mâle et la femelle est le mécanisme couramment accepté pour expliquer la suppression progressive de la recombinaison dans la PAR. La « contraction » de la PAR et « l'expansion » de la SDR se déroule par étape dans le temps et crée des « strates évolutives ». Ces strates sont bien caractérisées pour plusieurs espèces de différents groupes, telles que l'homme (Lahn and Page 1999), la plante *Silene* (Bergero et al. 2007), les serpents (Vicoso et al. 2013) ou encore les oiseaux (Zhou et al. 2014). Cette contraction par étape est due à l'expansion progressive de la SDR par la suppression de la recombinaison.

Un premier mécanisme pour expliquer la maintenance de la PAR est simplement, comme présenté précédemment, est son rôle pour le bon déroulement de la ségrégation des chromosomes durant la méiose. Un second mécanisme pour expliquer la persistance de la PAR est la translocation de régions autosomiques contenant des gènes présentant un antagonisme au niveau du sexe. Ces

translocations sont particulièrement favorables pour les régions génomiques sous l'effet d'une sélection par le sexe. Elles facilitent la divergence dans la fréquence des allèles entre les mâles et les femelles (Charlesworth and Charlesworth 1980). Un dernier mécanisme pour expliquer la maintenance de la PAR est la résolution de l'antagonisme sexuel par l'expression différentielle de ces gènes entre les sexes. Ce mécanisme a été observé chez l'Emeu (Vicoso et al. 2013). Dans cette espèce, la présence d'un gène dont l'expression est favorable pour le mâle et délétère pour la femelle entraîne une diminution de l'expression du gène chez les femelles. Cela engendre une résolution de l'antagonisme sexuel qui supprime l'effet de pression de sélection qui aurait normalement tendu à supprimer la recombinaison au niveau de ce gène entre les chromosomes Z et W et provoquer le passage du gène de la PAR vers la SDR.

La PAR présente des caractéristiques particulières comparées aux autosomes et la SDR. Elle possède par exemple un niveau de recombinaison supérieur au niveau de celui des autosomes dans certaines régions (Lien et al. 2000; Kondo et al. 2001) et tend à avoir une accumulation d'éléments répétés (Smeds et al. 2014). Le fait que les gènes de la PAR soient partiellement liés au sexe devrait influencer la dynamique de leur évolution. Des travaux théoriques prédisent que la PAR devrait être enrichie par des gènes antagonistes au niveau du sexe (Charlesworth and Charlesworth 1980; Clark 1988).

L'arrêt de la recombinaison engendre diverses modifications au niveau de la structure de la SDR, comme l'accumulation de mutations délétères, d'éléments répétés et d'amplicons (White 1973; Skaletsky et al. 2003; Bachtrog 2013), mais aussi des évènements de pseudogénisation des gènes codants (Bachtrog 2005; Zhou and Bachtrog 2012). La caractéristique unique de la transmission limitée du Y de père en fils favorise la conservation et le recrutement de gènes bénéfiques et spécifiques pour le mâle au sein de ce chromosome (Brosseau 1960; Rice 1996b; Lahn and Page 1997; Carvalho et al. 2000; Carvalho et al. 2001). Cela explique que le chromosome Y, au cours de son évolution, a perdu une majorité des gènes le composant à l'origine (Bull 1983; Charlesworth 1991; Rice 1996b; Bachtrog et al. 2011). Les gènes encore présents dans le chromosome Y possèdent, le plus souvent, des fonctions associées à la régulation de l'expression. Certains gènes, comme le SRY chez les Thériens, ont évolué pour acquérir des fonctions spécifiques, comme la spermatogénèse ou le développement (Cortez et al. 2014).

Expression des gènes biaisés par le sexe

Les mâles et les femelles d'une même espèce partagent un génome commun, mais expriment deux phénotypes différents, avec des variations plus ou moins prononcées aux niveaux comportemental, morphologique et physiologique. Cependant, seul un nombre limité de gènes sont spécifiques aux chromosomes sexuels, comme expliqué précédemment, et un nombre aussi limité de gènes ne peut expliquer à lui seul de telles différences entre les individus des deux sexes. La différence entre les deux phénotypes peut cependant être expliquée par une régulation différente de l'expression des gènes à l'échelle du génome (« expression des gènes biaisés par le sexe ») et non uniquement aux gènes localisés dans les chromosomes sexuels.

Le développement des technologies NGS a permis de réaliser de nombreuses expériences afin d'identifier et d'étudier les gènes différentiellement exprimés entre les sexes pour de nombreuses espèces comme la drosophile (Perry et al. 2014), les oiseaux (Pointer et al. 2013; Uebbing et al. 2013), les poissons (Böhne et al. 2014; Sharma et al. 2014), les nématodes (Albritton et al. 2014) ou encore les algues brunes (Martins et al. 2013; Lipinska et al. 2015). Les gènes biaisés par le sexe (GBS) montrent deux profils d'expression. Ils peuvent être spécifiques à un sexe, le gène est alors exprimé uniquement dans l'un des deux sexes. Ou bien, ils peuvent être différentiellement régulés. Dans ce cas de figure, un gène est soit exprimé de manière plus importante dans l'un des sexes (up-regulated) ou à l'inverse, il est exprimé de manière plus faible (down-regulated) (Ellegren and Parsch 2007; Parsch and Ellegren 2013). Les gènes qui ne présentent pas de différences significatives de niveau d'expression entre mâles et femelles sont qualifiés de « non biaisés ». Il est important de noter que dans l'étude de l'expression des gènes biaisés par le sexe, on assume que le niveau de transcrits (mARN) est hautement corrélé avec le niveau de protéines, mais une régulation traductionnelle des mARN ne peut être exclue.

Origines des GBS

L'origine des gènes biaisés par le sexe est multiple et peut se regrouper dans trois catégories présentées dans les paragraphes suivants.

Antagonisme sexuel

L'une des possibilités expliquant l'apparition des GBS est la présence d'un antagonisme au niveau des sexes entre les gènes. Un gène ancestral neutre, exprimé de manière équivalente dans les

deux sexes peut acquérir une mutation augmentant ou diminuant le niveau d'expression du gène et lui conférer un avantage pour l'un des deux sexes et être préjudiciable à l'autre. On voit alors l'apparition d'un antagonisme de l'expression entre les sexes. Sous l'effet de l'évolution et sans régulation de l'expression spécifique au sexe, un équilibre du niveau d'expression sera atteint et représentera un compromis entre les optimums d'expression pour les deux sexes (Ellegren and Parsch 2007). Etant donné que l'équilibre ne peut être optimal pour chaque sexe, une modification du niveau d'expression du gène peut apparaître afin d'optimiser la valeur sélective (fitness), et engendrer la formation d'un gène biaisé par le sexe via des mécanismes de régulation de type *cis* et *trans* (**Figure 7a**) (Williams and Carroll 2009).

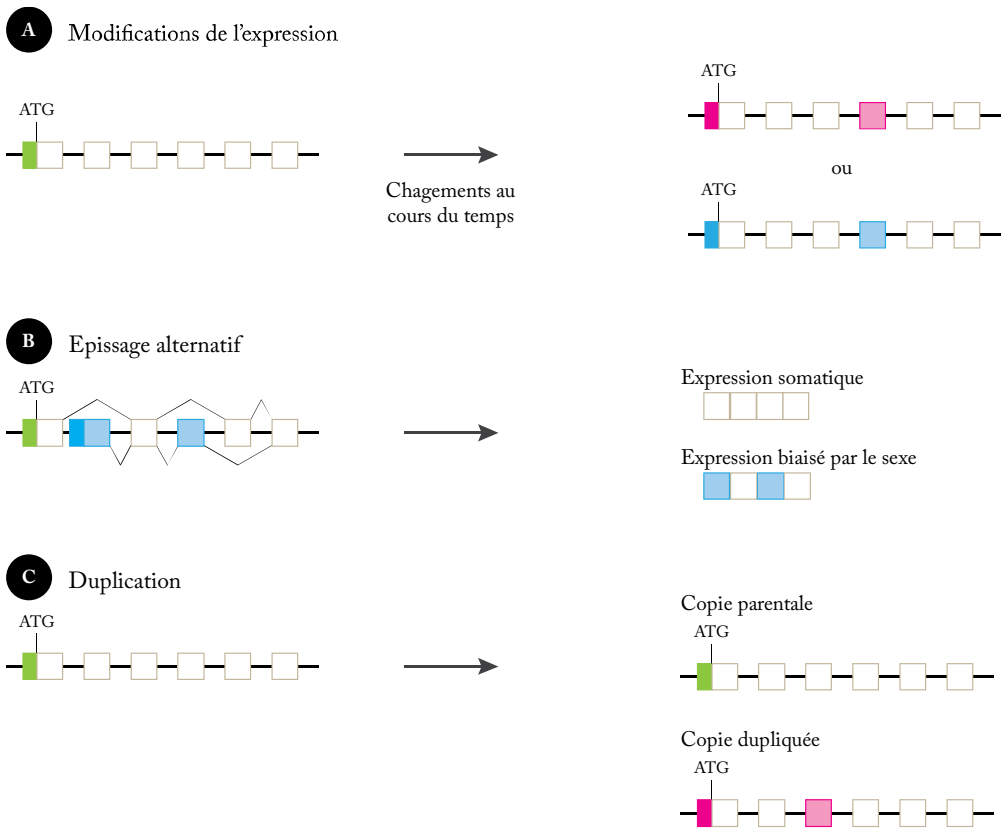


Figure 7 : Les différentes voies d'acquisition de gènes biaisés par le sexe (Gallach et al. 2011). A) Un changement dans la région *cis* régulatrice peut amener à une modification du niveau d'expression d'un gène soit chez la femelle (rose) ou bien chez le mâle (bleu). B) La présence d'isoformes pour un gène peut amener à l'utilisation préférentielle de l'une des isoformes par l'un des sexes. C) La duplication d'un gène peut amener l'une des copies à créer un gène biaisé par l'un des deux sexes. Dans les trois exemples présentés, le type d'expression du gène (neutre, mâle et femelle) est représenté par la boîte pleine, tandis que les boîtes vides représentent les différents exons. Le vert et le marron sont utilisés pour les gènes/exons exprimés de la même manière dans les deux sexes. Le bleu fait référence aux gènes/exons dont l'expression est biaisée chez les mâles. Enfin, le rose fait référence aux gènes/exons biaisés dont l'expression est biaisée chez les femelles.

Epissage alternatif

Le conflit sexuel pour un gène peut être résolu par un partitionnement de la séquence codante du gène pour créer des évènements d'épissage alternatif et ainsi permet la création d'isoformes à un gène (Parsch and Ellegren 2013). Dans ce cas, un type d'isoforme est privilégié par un sexe et une autre forme par l'autre sexe (**Figure 7b**). Les études disponibles, principalement chez la drosophile, confirment la présence de différence d'expression entre isoformes d'un même gène chez les mâles et les femelles (Gan et al. 2010; Chang et al. 2011). Ainsi, chez la drosophile, 36 % des gènes possèdent au moins deux isoformes et pour environ 23 % de ces gènes, l'expression des isoformes est biaisée par le sexe (Daines et al. 2011). Cependant, l'étude de l'épissage alternatif dans le cas de l'expression différentielle des gènes entre mâles et femelles reste encore une source assez peu explorée.

Duplication de gènes

Un autre mécanisme qui peut être une source de GBS est la duplication de gènes. Dans ce cas de figure, un gène va se dupliquer dans le génome et ainsi entraîner la formation d'une copie fonctionnelle. L'expression du gène parental reste inchangée tandis que la copie peut évoluer pour avoir un profil d'expression spécifique à l'un des sexes (**Figure 7c**). La duplication de gènes est une méthode particulièrement efficace pour fournir des réseaux de gènes qui sont spécifiques au sexe avec une régulation différente de ces réseaux pour les copies de gènes (Gallach and Betrán 2011). Il a cependant été observé, chez la drosophile, que la duplication favorise le développement de gènes biaisés au niveau de l'expression, majoritairement chez le mâle (Wyman et al. 2012).

Expression et évolution des gènes biaisés par le sexe

Au niveau de l'expression, il est difficile d'identifier une tendance générale quant au nombre de gènes biaisés par le sexe. En effet, ce nombre est variable et dépend de nombreux paramètres expérimentaux comme le design expérimental des expériences, les tissus analysés (un tissu spécifique, plusieurs types de tissus, l'organisme complet), l'organisme lui-même, le stade du cycle de vie, les outils d'analyses bio-informatique et statistique. Ces paramètres peuvent changer le nombre de gènes identifiés comme biaisés par le sexe de manière importante (Ellegren and Parsch 2007). Néanmoins, les différentes études semblent montrer un nombre important de gènes différentiellement exprimés entre les sexes. Par exemple chez la drosophile, ce nombre de gènes varie entre 50 et 75 % de la totalité des gènes exprimés (Assis et al. 2012). Il est important de constater que ces valeurs ne sont pas stables en fonction du cycle de vie, et que la quantité et les gènes impliqués varient au cours de la vie d'un individu en fonction des tissus analysés durant son développement (Parsch and Ellegren 2013).

Toujours dans le cas de la drosophile, l'analyse des gènes biaisés par le sexe au cours du cycle de vie a montré que la proportion de ces gènes était stable durant le développement des individus jusqu'à leur maturité (Perry et al. 2014). Au niveau des gènes eux-mêmes, certains sont biaisés tout au long de la vie de l'individu tandis que d'autres le sont uniquement à certains stades du cycle de vie.

Les différentes études réalisées au niveau des gènes biaisés par le sexe l'ont été sur des espèces dont le niveau de dimorphisme sexuel est relativement important et montrent qu'il existe une corrélation entre les niveaux d'expression des gènes biaisés par le sexe et le dimorphisme phénotypique observé chez les individus (Pointer et al. 2013). Pour les espèces avec un faible dimorphisme sexuel, les premiers résultats commencent à apparaître (Lipinska et al. 2015).

Au niveau de leur évolution moléculaire, les gènes biaisés par le sexe, plus particulièrement ceux biaisés chez le mâle dans le système XY, ont tendance à évoluer de manière plus rapide que les gènes non biaisés par le sexe (e.g. Zhang et al. 2004). Ce phénomène est observé chez la drosophile où les gènes biaisés par le sexe chez les mâles sont beaucoup plus divergents entre les différentes espèces du groupe que les gènes biaisés par le sexe chez femelles ou les gènes non biaisés par le sexe. La divergence est d'autant plus forte quand les gènes biaisés par le sexe chez le mâle sont uniquement exprimés chez ce dernier (Richards et al. 2005). La même observation a été faite chez *C. elegans* et certains mammifères (Cutter and Ward 2005; Khaitovich et al. 2005). Il a été observé chez les gènes biaisés par le sexe chez les mâles, une modification du biais d'usage des codons. Le biais d'usage des codons est un phénomène qui résulte de l'utilisation préférentielle, par les espèces, de codons dont la traduction est plus efficace pour la formation des ARN (Duret 2000; Duret and Mouchiroud 2000; Carbone et al. 2003). Chez la drosophile (Hambuch and Parsch 2005), le maïs et le blé (Whittle et al. 2007), une réduction du biais des codons est observée chez les gènes sexuellement biaisés chez les mâles. Cette évolution plus rapide des gènes biaisés par le sexe chez les mâles peut s'expliquer par la possibilité qu'ont ces gènes sous une pression de sélection plus faible à accumuler des mutations sans avoir d'impacts sur la valeur sélective (fitness). L'autre possibilité, supportée par des données expérimentales chez la drosophile (Zhang and Parsch 2005; Pröschel et al. 2006; Sawyer et al. 2007) et certains mammifères (Nielsen et al. 2005), est une augmentation de la pression de sélection sur ces gènes qui tend à un remplacement rapide de nucléotides. Cette augmentation de la sélection positive des gènes biaisés par le sexe chez les mâles peut être le résultat de la sélection naturelle, de la sélection sexuelle ou d'un antagonisme sexuel.

Les algues brunes et l'étude de l'évolution des sexes

La plupart des connaissances acquises sur le déterminisme sexuel portent sur les animaux, les plantes et les champignons. Très peu d'informations sont disponibles sur le fonctionnement du déterminisme sexuel dans les autres groupes Eucaryotes et il reste encore beaucoup à explorer.

L'importante distance phylogénétique séparant les algues brunes des autres groupes d'Eucaryotes (Charrier et al. 2008) fait que ce groupe présente un intérêt particulier pour l'étude de la détermination sexuelle et l'évolution des chromosomes sexuels (**Figure 8**). Dans le cadre de l'étude de la détermination du sexe, les connaissances et les données disponibles pour le système UV sont relativement limitées comparé aux deux autres systèmes XY et ZW. Plus précisément, des données génomiques sont disponibles pour l'algue verte Volvox (Ferris et al. 2010), pour l'algue brune *Ectocarpus sp.* (Ahmed et al. 2014) et l'hépatique Marchantia, dont seul le chromosome V est disponible (Yamato et al. 2007).

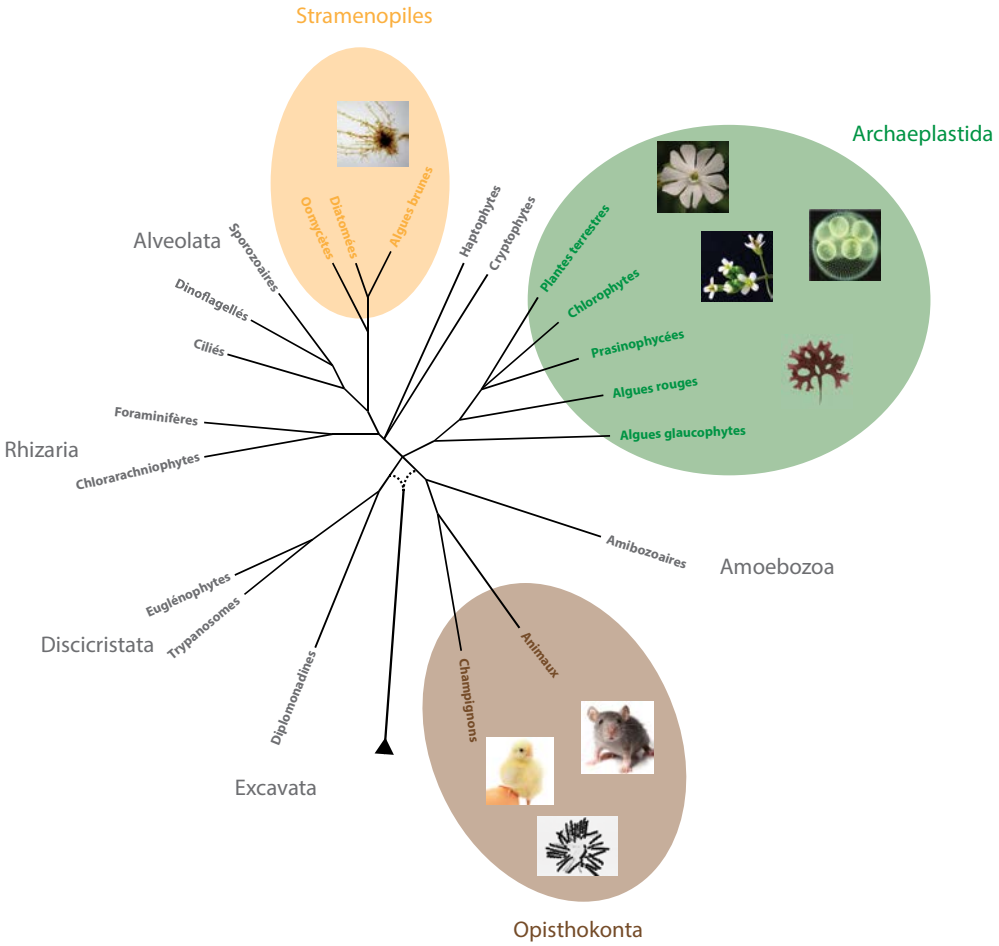


Figure 8 : Arbre phylogénétique des Eucaryotes (Baldauf 2008). En jaune, vert et marron, les principaux groupes étudiés dans le cadre de la détermination du sexe.

Ectocarpus est le modèle génétique et génomique retenu pour le groupe des algues brunes depuis la publication du génome mâle (Cock et al. 2010). *Ectocarpus* est une petite algue brune filamenteuse qui peut atteindre 30 cm dans la nature, mais qui peut devenir fertile en laboratoire à partir de 1-3cm (Charrier et al. 2008).

Ectocarpus a un cycle de vie haploïde-diploïde qui implique une alternance entre deux générations multicellulaires hétéromorphiques indépendantes, la phase gamétophytique et la phase sporophytique (**Figure 9**). La phase diploïde voit le développement d'un sporophyte asexué, qui une fois mature, libère des méiospores haploïdes. Les méiospores vont se développer en gamétophytes haploïdes, mâles lorsqu'ils possèdent le chromosome V, ou bien femelles lorsqu'ils possèdent le chromosome U. Les gamétophytes adultes produisent des gamètes mâles et femelles qui peuvent fusionner et donner le développement d'un sporophyte, complétant le cycle de vie d'*Ectocarpus*. Les gamètes non fécondés peuvent se développer de manière parthénogénétique et former un parthéno-sporophyte fonctionnel. Il présente la particularité de ne pas être morphologiquement différenciable du sporophyte diploïde (Peters et al. 2008). Une fois adulte, le parthéno-sporophyte va produire des gamètes qui vont se développer en gamétophytes. Le sexe est exprimé durant la phase haploïde menant à la différenciation des individus mâles et femelles. Les individus mâles et femelles de la phase gamétophytique ne montrent qu'un niveau limité de différenciation sexuelle au niveau morphologique, la différence étant sur le nombre de structures reproductrices, plus abondantes chez le mâle.

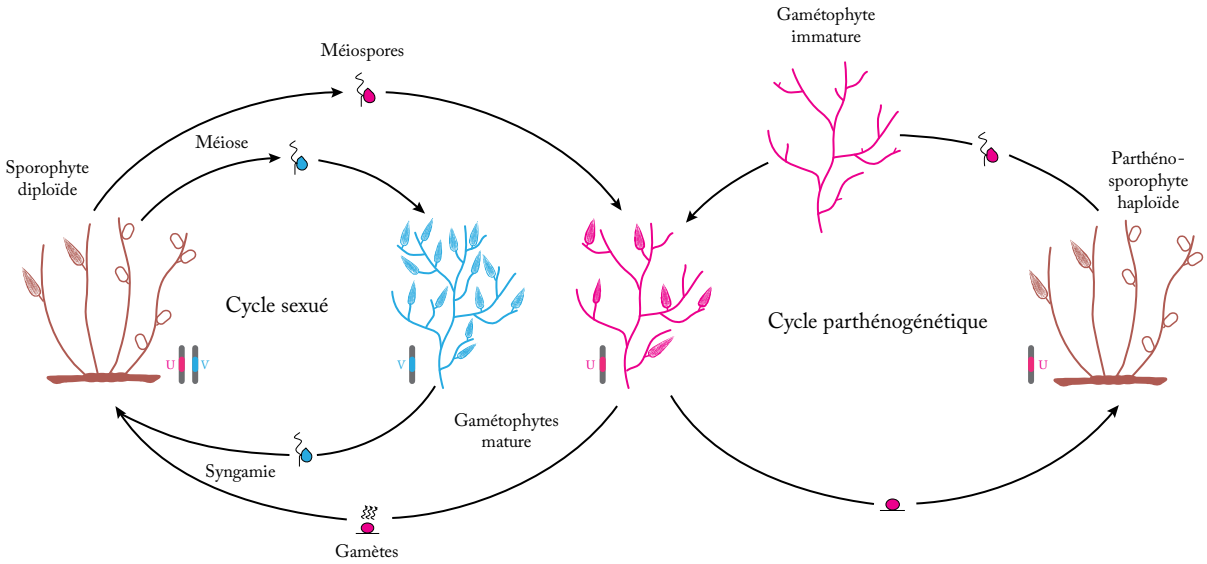


Figure 9 : Le cycle de vie d'*Ectocarpus sp*.

De nombreux outils et ressources ont été développés autour de ce modèle depuis la disponibilité du génome (**Figure 10**), incluant la PCR quantitative (Le Bail et al. 2008), des techniques de mutagénèse (Bail et al. 2012) et le développement de différents mutants (Peters et al. 2008; Coelho et al. 2011b), la cryopréservation (Heesch et al. 2012), une carte génétique (Heesch et al. 2010), des puces micro-array (Coelho et al. 2011a; Dittami et al. 2011), des EST (Cock et al. 2010), des techniques de protéomique et métabolomique (Ritter et al. 2010; Ritter et al. 2014) et des outils bio-informatique comme la mise en place d'un réseau métabolique (Prigent et al. 2014), la recherche et l'identification de miRNA (microRNA) (Billoud et al. 2014; Tarver et al. 2015) ou encore la prédiction de la localisation subcellulaire des protéines (Gschloessl et al. 2008). D'autres méthodes sont en cours de développement chez *Ectocarpus*, telles que des méthodes de tilling, pour la création de librairies de mutants, et de RNAi.

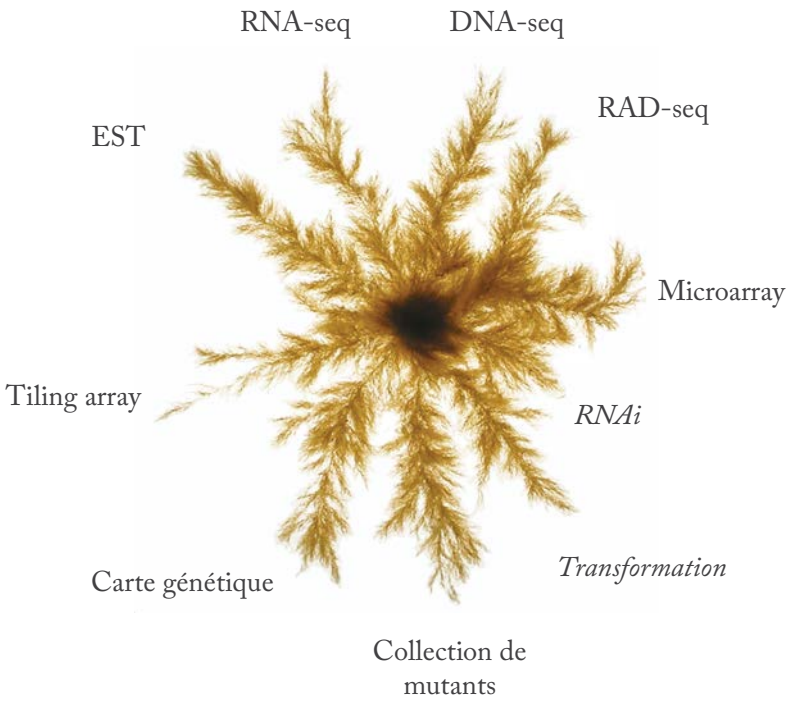


Figure 10 : Aperçu des développements réalisés autour d'*Ectocarpus*. En italique, les techniques toujours en cours de développement pour *Ectocarpus*.

Apport de la bio-informatique pour l'étude du déterminisme du sexe

Dans le cadre de l'étude du déterminisme du sexe, plusieurs types d'informations sont nécessaires afin de pouvoir en comprendre les mécanismes mis en jeu : l'accès à l'information génomique, permettant l'étude des chromosomes sexuels, et l'information transcriptomique pour étudier, par exemple, l'expression des gènes entre mâles et femelles.

Le développement des techniques NGS a permis un énorme bond en avant dans l'étude des données génomiques et transcriptomiques, en donnant un accès à moindre coût et de manière massive à ces sources d'informations, permettant une très grande variété d'analyses comme la détection et l'annotation d'isoformes, les analyses d'expression différentielle, l'accès à des transcriptomes sans le génome de référence, la détection de SNP ou SNV, etc. De nombreuses méthodologies et logiciels ont été développés autour de ces données afin de les traiter et ainsi répondre aux différentes questions biologiques posées.

DNA-seq

L'accès à la séquence génomique est indispensable afin de caractériser le génome et surtout les chromosomes sexuels. Pour cela, le traitement des données de séquençage demande une méthodologie de préparation et d'assemblage des données, mais aussi d'évaluation de la qualité de l'assemblage afin d'obtenir une séquence génomique optimale (Ekblom and Wolf 2014; Wajid and Serpedin 2014).

L'assemblage de données génomique pour d'obtenir la séquence finale du génome, représente un défi à plusieurs niveaux. Au niveau biologique, il existe une grande hétérogénéité entre les espèces, certaines pouvant présenter un très haut niveau de duplication, par exemple chez le blé (Salse et al. 2008) ou le riz (Yu et al. 2005), augmentant considérablement la difficulté de l'assemblage. Comme présenté précédemment, les chromosomes sexuels sont caractérisés entre autres par une grande richesse au niveau des éléments répétés. Cette caractéristique fait qu'il est d'autant plus difficile de réaliser un assemblage correct de ces régions. Mais les dernières évolutions dans les techniques de séquençage permettent aujourd'hui de passer outre cette limitation, par exemple dans le cadre de l'assemblage du chromosome sexuel chez le champignon *Microbotryum lychnidis-dioicae*, réalisé à l'aide de reads longs PacBio (Badouin et al. 2015).

Une fois le génome disponible, il reste à identifier clairement les séquences correspondantes aux chromosomes sexuels. Avec la disponibilité massive des données de séquences, de nouvelles méthodologies plus ou moins complexes ont été développées, par exemple la méthode YGS (Y chromosome Genome Scan) qui compare la fréquence des kmers entre données de séquençage provenant d'individus mâles et femelles (Carvalho and Clark 2013). D'autres approches ont montré leur efficacité, par exemple la détection via la recherche de différences de taux de couverture lors du mapping des reads de librairies mâles et femelles (Chen et al. 2012), le séquençage de BAC provenant des chromosomes sexuels après leur identification avec des sondes de gènes identifiés comme étant liés aux chromosomes sexuels (Wang et al. 2012; Blavet et al. 2015), des approches de RAD-seq (Qiu et al. 2015), ou encore des méthodologies d'assemblage *de novo* combinées à l'utilisation d'un génome assemblé et annoté d'une espèce proche afin d'identifier et reconstruire les chromosomes sexuels (Vicoso et al. 2013).

RNA-seq

L'accès aux informations du transcriptome permet de disposer de la séquence des gènes, soit via l'annotation du génome de référence ou par l'assemblage *de novo* de données RNA-seq. Le développement des techniques de RNA-seq a permis d'accéder de manière facilitée à cette information, et de lever les limitations liées aux techniques de micro-array et EST (Malone and Oliver 2011). Elle a permis, en outre, d'apporter tout une série de nouvelles possibilités, par exemple l'accès à l'information sur les évènements d'épissage alternatif, la détection de fusions de gènes, l'accès au niveau d'expression aussi bien des gènes faibles que ceux fortement exprimés (Ozsolak and Milos 2011). Plusieurs publications recensent les principaux types de méthodologies avec les outils associés (Oshlack et al. 2010; Chen et al. 2011), mais ne donnent qu'une vue limitée des outils disponibles. Différentes initiatives ont été lancées afin de recenser de la manière la plus exhaustive l'ensemble des outils disponibles et de les catégoriser selon leur fonction (Henry et al. 2014). Dans le cas des outils pour le traitement des données issues de séquençage RNA-seq, le site OMICtools recense, en septembre 2015, cent quarante-trois outils dédiés à ce type d'analyse. Cependant, certaines suites logicielles sont plus connues que d'autres, telles que le pipeline d'analyse de données RNA-seq Tuxedo (Trapnell et al. 2012) pour l'analyse de RNA-seq avec génome de référence. Dans le cadre de l'étude du déterminisme du sexe, la principale utilisation des données RNA-seq est liée à l'identification des gènes biaisés par le sexe et l'étude de leurs caractéristiques et leur évolution, mais aussi l'assemblage et l'identification des gènes spécifiques aux mâles et femelles (Bergero and

Charlesworth 2011; Chang et al. 2011; Chibalina and Filatov 2011; Assis et al. 2012; Muyle et al. 2012; Martins et al. 2013; Sun et al. 2013; Lipinska et al. 2015).

Design expérimental et préparation des données

La première étape consiste à réaliser un plan d'expérience, qui doit prendre en considération bien évidemment la question biologique qui est posée, le type de matériel biologique à utiliser, la plateforme de séquençage la plus adaptée, mais aussi les contraintes qui peuvent intervenir pour la post-analyse (Strickler et al. 2012). Par exemple, dans le cas d'analyse de l'expression différentielle entre plusieurs conditions, il est indispensable de disposer de réplicats biologiques, au minimum deux, afin de prendre en compte la variation biologique entre les individus ainsi que la profondeur de séquençage pour identifier correctement les gènes différentiellement exprimés (Liu et al. 2014; Sims et al. 2014). D'autres contraintes peuvent influencer aussi la prise de décision sur le type de séquençage à réaliser, par exemple l'utilisation de données single-end ou paired-end, la taille des reads ou bien la profondeur de séquençage désirée.

Une fois le séquençage réalisé, une phase d'exploration des données est nécessaire (**Figure 11a**) afin de s'assurer de la qualité des données avec des outils tels que FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), HTSeq-qa (Anders et al. 2015) ou bien Kraken (Davis et al. 2013). En fonction du résultat et du type d'analyse à réaliser, il est parfois nécessaire de préparer les données. Ces étapes comprennent la plupart du temps une phase de « nettoyage » et de sélection des reads selon un score de qualité et une taille minimale. Elles peuvent inclure une étape de décontamination et de suppression des adaptateurs utilisés lors du séquençage. De nombreux outils sont disponibles pour réaliser ces différentes tâches de préparation des données. Certains sont génériques et peuvent exécuter plusieurs types de tâches, telles que Trimmomatic (Bolger et al. 2014), PRINSEQ (Schmieder and Edwards 2011a) ou FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) et d'autres, dédiés à la réalisation de traitements spécifiques, tels que DeconSeq sont utilisés pour supprimer les amorces de séquençage (Schmieder and Edwards 2011b) ou bien RiboPicker pour supprimer les reads correspondant à des contaminations d'ARNr (Schmieder et al. 2012).

Une fois la qualité des données validés, il est possible de passer à leur exploitation au travers des différents pipelines d'analyse RNA-seq.

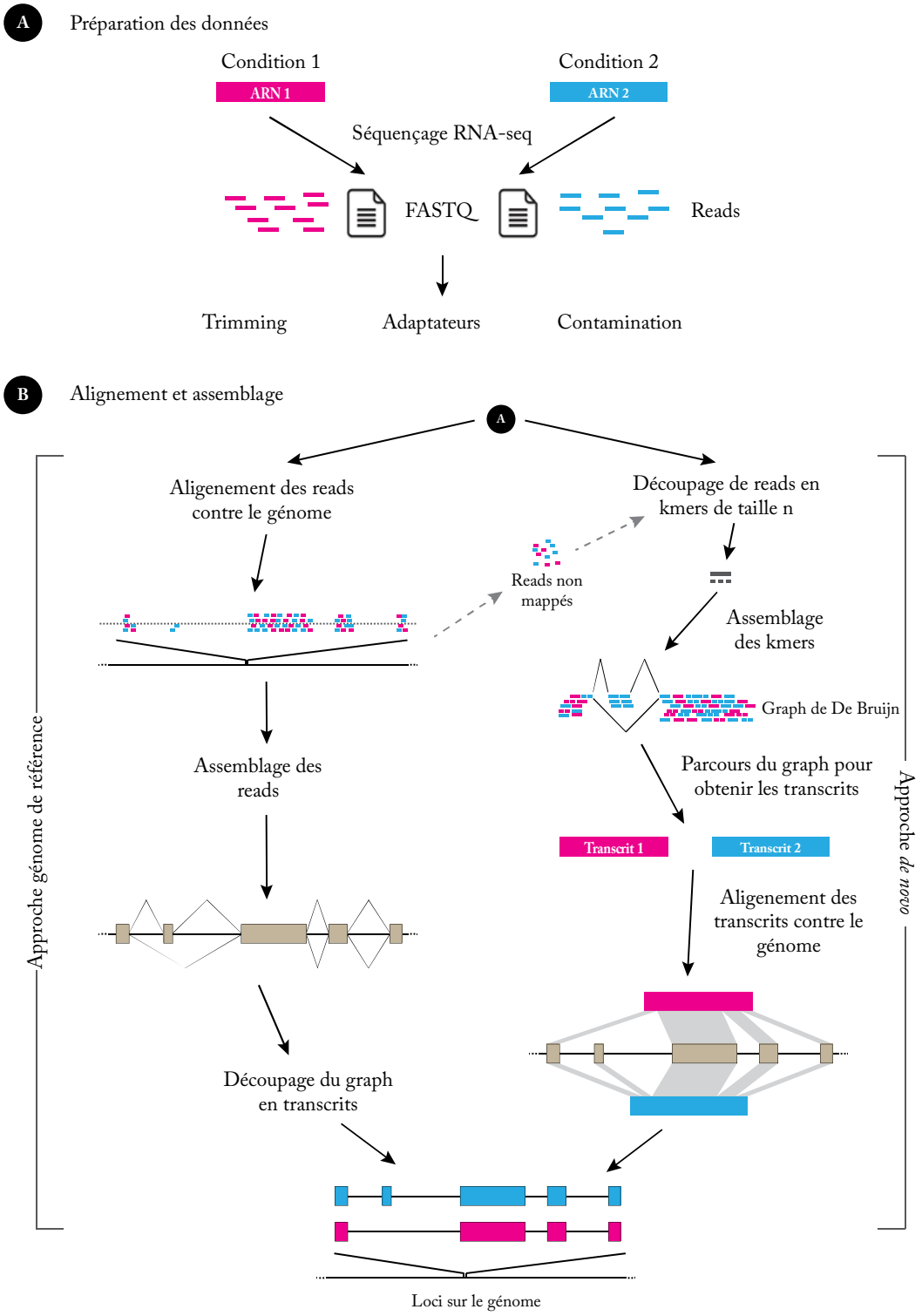


Figure 11 : Exemple de pipeline de préparation et d'assemblage des données issues de séquençage RNA-seq. Adapté de (Oshlack et al. 2010; Garber et al. 2011). A) Une fois le séquençage réalisé, les données de sortie sont préparées pour la suite des analyses. Cette étape de préparation consiste principalement à vérifier la qualité des données, retirer les adaptateurs de séquençage, supprimer les bases de mauvaise qualité, supprimer les contaminants. B) Les données traitées sont ensuite alignées et assemblées, avec deux voies possibles selon la disponibilité ou non d'un génome de référence.

Alignement et assemblage de lectures

A partir des données obtenues après séquençage, deux méthodologies d'assemblage des données sont disponibles (**Figure 11b**). La première, basée sur l'utilisation du génome de référence, consiste à aligner les lectures (reads) des différentes librairies contre le génome de référence, avec ou sans l'utilisation des annotations structurales des gènes. La seconde possibilité, communément appelée approche *de novo*, obligatoire en l'absence de génome de référence, est de partir de l'information contenue dans les reads pour les assembler entre eux afin d'obtenir la séquence des transcrits. Ces deux approches sont décrites de manière plus spécifique dans les paragraphes suivants.

Approche avec génome de référence

La disponibilité d'un génome de référence permet de faciliter la tâche de mapping et d'assemblage des reads. La première étape consiste à réaliser un mapping des reads contre le génome de référence, avec des mappeurs dédiés. Il faut en effet tenir compte de la structure des gènes, les données RNA-seq étant obtenues à partir de mRNA et le mapping réalisé sur un génome. Il existe donc des reads qui sont situés sur les jonctions entre exons, côte à côte sur le mRNA mais pouvant être distants sur le génome à cause de la présence d'introns. Des outils ont été développés afin de tenir compte de cette structure particulière, les « splicing-aware aligner », tels que TopHat (Trapnell et al. 2009), SpliceMap (Au et al. 2010) ou bien GSNAP (Wu and Nacu 2010). Ces différents outils utilisent des méthodologies différentes pour réaliser le mapping des reads contre un génome (**Figure 12**). Par exemple, TopHat et SpliceMap, utilisant la méthode dite « exon-first », vont d'abord chercher à mapper les reads sur la totalité de leur longueur, sans autoriser de gaps. Ensuite, les reads restants sont découpés en sous-parties et mappés contre le génome. Une extension est ensuite réalisée afin de retrouver la structure du read, en autorisant l'insertion de gap (**Figure 12a**). L'autre méthode, dite de « seed-extend », utilisée par des outils tels que GSNAP, va d'abord créer les différentes graines (seeds) pour chaque read et va ensuite mapper ces seeds contre le génome. Une fois la seed mappée, un processus d'extension est réalisé afin de retrouver le mapping complet du read, en autorisant l'insertion de gaps (**Figure 12b**).

En plus du génome, il est possible de fournir aux mappeurs les annotations structurales des gènes afin de les guider dans leur tâche. La disponibilité d'annotations lors de cette étape peut grandement influencer le taux de mapping des reads, surtout dans le cas des reads situés au niveau de la jonction entre exons. Par exemple, dans le cas d'une analyse comparative du taux de mapping sans et avec l'annotation de référence chez l'humain, plus d'un tiers des reads situés au niveau des jonctions

n'étaient pas mappés en l'absence de l'annotation (Zhao 2014). Une fois l'étape de mapping réalisée, il est possible de passer directement à la partie expression différentielle lorsque l'on dispose d'une annotation structurale des gènes pour le génome (**Figure 13a**). Un comptage du nombre de reads par gène et par condition est alors réalisé avec des outils tels que HTSeq (Anders et al. 2015) ou featureCounts (Liao et al. 2014).

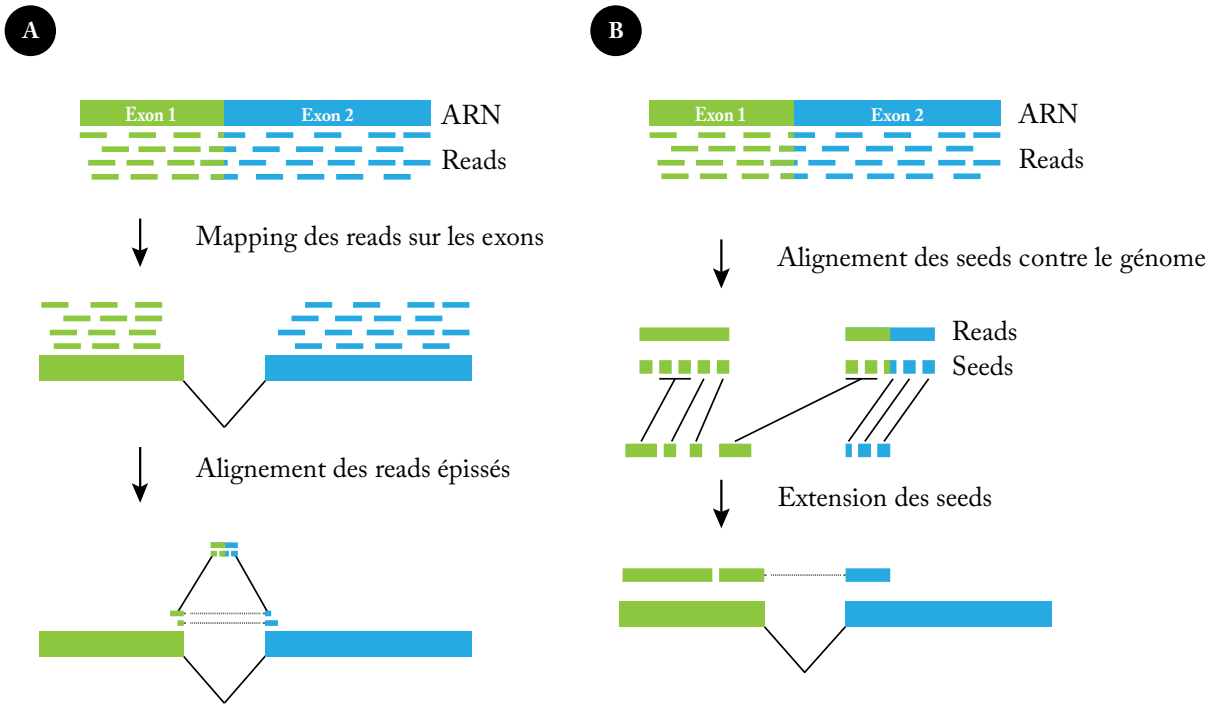


Figure 12 : Les différentes stratégies pour le mapping des reads RNA-seq contre un génome (Garber et al. 2011). A) Méthode « exon-first » | TopHat, SpliceMap. Dans un premier temps, l'ensemble des reads non épissés sont mappés contre le génome. Les reads non-mappés sont divisés en sous-parties et mappés sur le génome. Un processus d'extension des parties mappées est réalisé afin de vérifier qu'il s'agit bien site d'épissage alternatif B) Méthode « seed-extend » | GSNAP. Les reads sont découpés en sous-parties afin d'obtenir les différentes graines (seeds) qui sont alors alignées contre le génome. Un processus d'extension de la seed est ensuite réalisé pour obtenir des alignements de plus grande taille, pouvant inclure les intervalles provenant de la présence de sites d'épissage.

En l'absence d'annotations structurales des gènes, il est nécessaire de passer par une étape d'assemblage des reads afin de reconstruire les transcrits et isoformes (**Figure 11b**). Plusieurs outils d'assemblage des transcrits sont disponibles, tels que Cufflinks (Trapnell et al. 2010), StringTie (Pertea et al. 2015) ou Scripture (Guttman et al. 2010). Les outils comme Cufflinks et StringTie utilisent les informations de mapping des reads afin de créer un graph et le parcours de ce dernier permet d'identifier les différentes isoformes. Cependant, même en présence d'une annotation de référence, l'étape d'assemblage des transcrits peut être réalisée afin, par exemple, de compléter

l'annotation structurale déjà existante, soit par la prédiction de nouveaux gènes ou par la prédiction des isoformes des gènes déjà annotés. L'annotation est fournie, en plus des mapping des reads, à l'assembleur afin de le « guider » et va permettre, en plus de l'assemblage des annotations de référence, la découverte de nouveaux gènes et de nouvelles isoformes. De plus, la mise à disposition de l'annotation permet d'assurer un transfert des annotations de référence dans le nouvel assemblage afin d'en évaluer plus facilement la qualité, mais permet aussi d'assembler plus facilement des gènes avec une faible couverture RNA-seq, avec des méthodologies telles que le RABT (reference annotation based transcript assembly) (Roberts et al. 2011). Ce genre d'approche d'assemblage de transcriptome avec annotation de référence a été utilisé avec succès chez l'homme (Trapnell et al. 2010) et chez la drosophile (Daines et al. 2011).

L'approche avec génome de référence est un domaine où les développements logiciels sont très fréquents, avec des mises à jour régulières des outils, de nouvelles versions d'un outil ou encore de nouveaux outils dans le but d'améliorer les performances autant au niveau de la qualité des prédictions que des performances computationnelles (Langmead et al. 2009; Trapnell et al. 2009; Trapnell et al. 2010; Langmead and Salzberg 2012; Dobin et al. 2013; Kim et al. 2013; Pertea et al. 2015).

Approche sans génome de référence ou *de novo*

Dans le cas de l'absence d'un génome de référence, l'approche d'assemblage des reads *de novo* permet, à partir de l'information contenue dans les reads, d'assembler directement les transcrits et leurs isoformes. Plusieurs outils sont disponibles pour répondre à cette problématique, tels que Trinity (Grabherr et al. 2011), SOAPdenovo (Luo et al. 2012), Oases (Schulz et al. 2012) ou encore Trans-ABySS (Robertson et al. 2010). Ces outils se basent sur la création de graphs de De Brujin à partir de reads découpés en kmers de taille n et le parcours de ces graphs, afin de retrouver les différents transcrits et les isoformes du transcriptome. L'assemblage de transcriptome *de novo* pose plusieurs problèmes. Par exemple, la création des graphs de De Brujin est particulièrement sensible aux erreurs de séquençage, provoquant la création de transcrits partiels ou de transcrits chimériques (McGettigan 2013). Il est donc indispensable d'avoir un contrôle et une préparation des données en amont particulièrement rigoureux. Les assembleurs *de novo*, du fait des algorithmes utilisés, sont très consommateurs de ressources, autant au niveau matériel avec une forte utilisation en RAM et CPU, mais aussi en terme de temps de calcul (Lu et al. 2013). Cependant, l'évolution des programmes fait que certains points sont améliorés entre les différentes versions, par exemple pour le logiciel Trinity,

dont le temps d'exécution a nettement diminué, de l'ordre de 60%, entre la première et la dernière version du logiciel (http://trinityrnaseq.github.io/performance/).

Malgré les contraintes associées à cette méthodologie, au niveau de la préparation des données et de la consommation en ressources, elle présente l'avantage de se passer de génome, ce qui permet d'augmenter considérablement le nombre d'espèces pouvant être étudiées au niveau transcriptomique, tout en fournissant une qualité suffisante au niveau de la reconstruction des transcrits (Grabherr et al. 2011; Strickler et al. 2012).

Expression différentielle

Une fois les données de comptage de reads obtenues pour chaque gène dans les différentes conditions et pour l'ensemble des réplicats, il est possible de réaliser des analyses d'expression différentielle entre les gènes. Cependant, il est nécessaire, dans un premier temps, de corriger les données de comptage par une étape de normalisation. En effet, les données de séquençage présentent des biais inhérents à la technique, tels qu'une variation de la couverture en reads liée à la taille des gènes (Oshlack and Wakefield 2009) ou encore la variation de taille entre les librairies (Mortazavi et al. 2008). Plusieurs méthodes sont disponibles pour normaliser les données, comme la méthode TMM (Trimmed Mean of M values) (Robinson and Oshlack 2010), la méthode RPKM (Reads Per Kilobase per Million mapped read) (Mortazavi et al. 2008) ou la méthode implémentée dans le package d'analyses d'expression différentielle DESeq (Anders and Huber 2010). Elles ne sont pas équivalentes, certaines étant plus efficaces que d'autres (Dillies et al. 2012), et le choix d'une méthode adéquate est d'autant plus important qu'elle a une grande influence sur les analyses statistiques qui en découlent (Hoffmann et al. 2002).

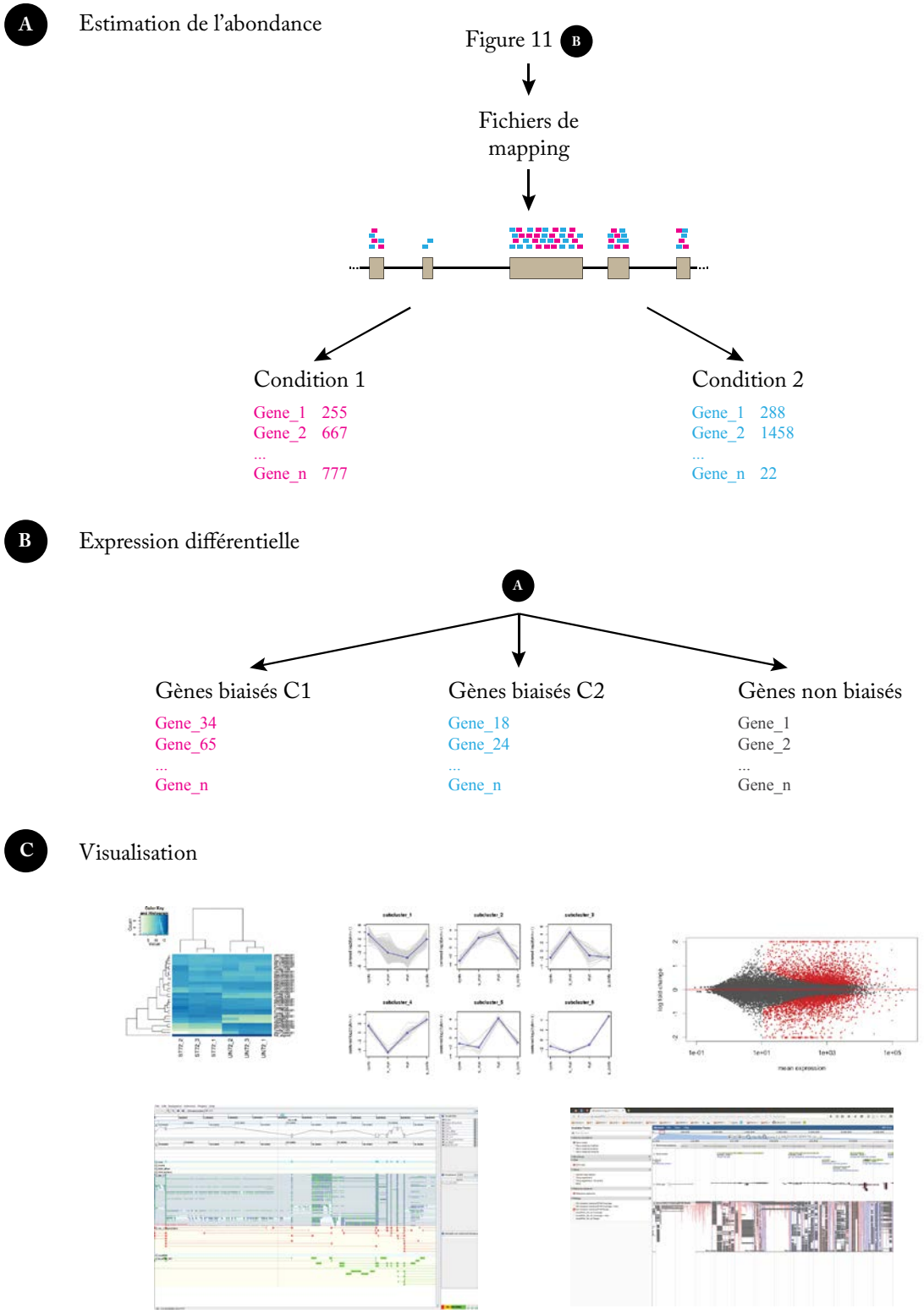


Figure 11 B

Figure 13 : Exemple de pipeline de l'analyse d'expression différentielle et de visualisation des données RNA-seq. Adapté de (Oshlack et al. 2010; Griffith et al. 2015). A) Le comptage du nombre de reads alignés contre le transcriptome permet d'estimer l'abondance de chacun des gènes. B) A partir du comptage du nombre de reads par gène et par condition, il est possible d'identifier les gènes différentiellement exprimés entre plusieurs conditions. C) Les données obtenues peuvent être visualisées de manière graphique via des packages R ou bien des outils spécifiquement développés à cette fin.

Une fois les données normalisées, il est possible de passer à l'étape d'analyses statistiques de l'expression différentielle entre différentes conditions (**Figure 13b**). A cette fin, plusieurs outils statistiques sont disponibles, principalement distribués via des packages R. Les différences entre les différents outils sont liées aux méthodologies statistiques employées afin d'identifier les gènes différentiellement exprimés entre conditions. Il existe par exemple des outils basés sur la méthode binomiale négative tels que DESeq (Anders and Huber 2010), edgeR (Robinson et al. 2010) et baySeq (Hardcastle and Kelly 2010) d'autres basés sur des tests non paramétriques comme NOIseq (Tarazona et al. 2011) ou encore des méthodes basées sur la détection au niveau du transcrit et ensuite rapportées au niveau du gène, comme Cuffdiff 2 (Trapnell et al. 2013) et EBSeq (Leng et al. 2013). Les analyses comparatives de ces outils ont montré un manque de consistance entre leurs résultats selon le type de données ou encore le nombre de réplicats (Kvam et al. 2012; Seyednasrollah et al. 2015). Il est donc particulièrement important d'adapter l'utilisation de l'outil aux données à analyser ou bien de mener des analyses comparatives sur les résultats obtenus afin d'assurer des résultats corrects (Seyednasrollah et al. 2015).

Dans le cadre de cette thèse, les différentes méthodologies bio-informatiques pour l'assemblage des données DNA-seq et RNA-seq ont permis d'identifier le chromosome femelle et les gènes associés, d'étudier la structure et l'évolution des chromosomes sexuels et de réaliser une étude des différences d'expression des gènes entre mâle et femelle à différentes étapes de la croissance d'*Ectocarpus*, résultats qui sont présentés dans les articles suivants. Ces méthodologies et résultats ont aussi été utilisés afin de réaliser une réannotation complète du génome d'*Ectocarpus*, incluant la prédiction et l'identification des miRNA et lncRNA (long non-coding RNA), résultats qui seront présentés dans la seconde partie du manuscrit.

Article 1 - A Haploid System of Sex Determination in the Brown Alga *Ectocarpus* sp.

Introduction

L'article présenté porte sur l'identification et l'analyse génomique de la SDR femelle chez l'algue brune *Ectocarpus*. Plus spécifiquement, ma contribution à cet article a été d'identifier le SDR femelle et de vérifier si l'identification des séquences de la SDR mâle était complète. A cette fin, plusieurs méthodologies, basées sur différents types de données de séquençage, ont été utilisées et combinées afin de pouvoir identifier l'ensemble des séquences composant les SDR.

La première approche a consisté à combiner l'utilisation de données RNA-seq d'individus mâles et femelles avec des données génomiques. Les données RNA-seq ont été obtenues à partir d'individus mâles (Ec603) et femelles (Ec602) de deux souches quasi isogéniques (**Figure S1**), avec un total de quatre librairies Illumina single-end, deux pour chaque sexe, obtenues à partir de gamétophytes matures. Les données génomiques comprenaient le génome mâle (Ec32) (Cock et al. 2010) et l'assemblage du génome femelle (Ec597) (**Figure S1**) (données non publiées). Dans le cas du génome mâle, les données Sanger ont été assemblées au VIB à Gand. Pour le génome femelle, quatre librairies Roche 454 single-end, deux librairies Ilumina mate-paire (avec une taille d'insert de 10 kb) et une librairie Illumina paired-end ont été séquencées et l'assemblage de ces données a été réalisé par le Génoscope. Les assemblages des données RNA-seq par sexe ont permis d'obtenir deux transcriptomes, un premier avec les données mâles et un second avec les données femelles. La comparaison des transcriptomes a permis d'identifier les transcrits spécifiques à chaque assemblage et donc à chacun des deux sexes. La recherche d'homologies de séquences avec Blast entre les transcrits sexe spécifiques et les deux génomes a permis d'identifier les séquences dans les assemblages des deux génomes correspondant à des loci spécifiques de chacun des sexes.

Afin de palier au fait que cette technique permet d'identifier uniquement les séquences génomiques possédant des gènes spécifiques à l'un des deux sexes, j'ai adapté l'approche YGS (Y

chromosome Genome Scan) (Carvalho and Clark 2013) aux données d'*Ectocarpus* afin de vérifier l'exhaustivité de l'identification précédemment réalisée.

Une fois les différentes séquences composant la SDR femelle identifiées et validées, une autre partie de ma contribution, réalisée en collaboration avec le VIB à Gand, a été d'annoter structuralement la SDR femelle dans le but d'obtenir la structure des différents gènes. La prédiction de gènes a été réalisée en utilisant le même protocole que lors de l'annotation du génome mâle, par l'utilisation de l'outil Eugène (Cock et al. 2010). De plus, afin de prendre en compte l'information fournie par les données RNA-seq, les prédictions obtenues avec Eugène ont été annotées de manière experte en les croisant avec les transcrits assemblés par Cufflinks et l'alignement des transcrits Trinity contre les séquences génomiques. Dans le cas du génome mâle, ce dernier ayant déjà été annoté lors de la publication du génome, une réannotation experte complète des gènes de la SDR a été réalisée pour vérifier la qualité des modèles de gènes. Une fois la prédiction et l'annotation experte des gènes des deux SDR réalisées, les données RNA-seq ont été mappées une seconde fois avec TopHat pour optimiser le mapping des reads sur ces gènes. Le niveau d'expression de l'ensemble des gènes a été mesuré à l'aide de Cufflinks/Cuffdiff afin de permettre la comparaison du niveau d'expression des gènes de la SDR avec les gènes présents dans les autosomes.

Le contig sctg_285 présente la particularité d'avoir une portion de sa séquence qui appartient à la SDR et l'autre à la PAR. La localisation précise de la zone de transition entre ces deux régions représente un point important. Les techniques de mesure du taux de recombinaison permettent de déterminer la zone de transition, mais avec une précision limitée. Une dernière partie de ma contribution a été de réaliser le mapping des données DNA-seq femelle sur ce contig d'origine mâle. Cela a permis d'améliorer considérablement la localisation de la zone de transition entre les deux régions par l'analyse du taux de couverture des reads le long du super-contig, le taux de couverture étant nul au niveau de la SDR mâle et d'une moyenne de 40x au niveau de la PAR.

Les résultats de ces travaux ont été intégrés avec d'autres analyses dans l'article suivant, publié dans Current Biology en septembre 2014.

Article

A Haploid System of Sex Determination in the Brown Alga Ectocarpus sp.

Auteurs : Ahmed S[1,2§], Cock JM[1§], Pessia E[3§], Luthringer R[1], Cormier A[1], Robuchon M[1,8], Sterck L[4], Peters AF[5], Dittami SM[1], Corre E[7], Valero M[8], Aury J-M[6], Roze D[8], Van de PeerY[4,9], Bothwell JH [2], Marais GAB [3], Coelho SM[1*]

Affiliations :

[1] Integrative Biology of Marine Models, CNRS UMR 8227, Sorbonne Universités, UPMC Université Paris 6, Station Biologique de Roscoff, CS 90074, 29688 Roscoff, France;

[2] Medical Biology Centre, Queens University Belfast, Belfast BT9 7BL, Northern Ireland, UK;

[3] Laboratoire de Biométrie et Biologie Evolutive, UMR 5558, Centre National de la Recherche Scientifique, Université Lyon1, 69622 Villeurbanne, France;

[4] Department of Plant Systems Biology (VIB) and Department of Plant Biotechnology and Bioinformatics (Ghent University), Technologiepark 927, 9052 Gent, Belgium;

[5] Bezhin Rosko, 29250 Santec, France;

[6] Commissariat à l'Energie Atomique (CEA), Institut de Génomique (IG), Genoscope, 91000 Evry, France;

[7] ABIMS Platform, FR2424, Station Biologique de Roscoff, CS 90074, 29688 Roscoff, France;

[8] Evolutionary Biology and Ecology of Algae, CNRS UMI 3604, Sorbonne Université, UPMC, PUCCh, UACH, Station Biologique de Roscoff, CS 90074, 29688 Roscoff, France;

[9] Genomics Research Institute, University of Pretoria, Hatfield Campus, Pretoria 0028, South Africa.

§égale contribution

Correspondance : coelho@sb-roscoff.fr

Article publié dans Current Biology

doi:10.1016/j.cub.2014.07.042

Current Biology 24, 1945–1957, September 8, 2014 ©2014 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2014.07.042

Article

A Haploid System of Sex Determination in the Brown Alga *Ectocarpus* sp.

Sophia Ahmed,[1,2,10] J. Mark Cock,[1,10] Eugenie Pessia,[3,10] Remy Luthringer,[1] Alexandre Cormier,[1] Marine Robuchon,[1,8] Lieven Sterck,[4] Akira F. Peters,[5] Simon M. Dittami,[1] Erwan Corre,[7] Myriam Valero,[8] Jean-Marc Aury,[6] Denis Roze,[8] Yves Van de Peer,[4,9] John Bothwell,[2] Gabriel A.B. Marais,[3] and Susana M. Coelho[1,*]

[1]Integrative Biology of Marine Models, CNRS UMR 8227, Sorbonne Universités, UPMC Université Paris 6, Station Biologique de Roscoff, CS 90074, 29688 Roscoff, France
[2]Medical Biology Centre, Queens University Belfast, Belfast BT9 7BL, Northern Ireland, UK
[3]Laboratoire de Biométrie et Biologie Évolutive, UMR 5558, Centre National de la Recherche Scientifique, Université Lyon 1, 69622 Villeurbanne, France
[4]Department of Plant Systems Biology (VIB) and Department of Plant Biotechnology and Bioinformatics (Ghent University), Technologiepark 927, 9052 Gent, Belgium
[5]Bezhin Rosko, 29250 Santec, France
[6]Commissariat à l'Energie Atomique (CEA), Institut de Génomique (IG), Genoscope, 91000 Evry, France
[7]ABiMS Platform, FR2424, Station Biologique de Roscoff, CS 90074, 29688 Roscoff, France
[8]Evolutionary Biology and Ecology of Algae, CNRS UMI 3604, Sorbonne Université, UPMC, PUCCh, UACH, Station Biologique de Roscoff, CS 90074, 29688 Roscoff, France
[9]Genomics Research Institute, University of Pretoria, Hatfield Campus, Pretoria 0028, South Africa

Summary

Background: A common feature of most genetic sex-determination systems studied so far is that sex is determined by nonrecombining genomic regions, which can be of various sizes depending on the species. These regions have evolved independently and repeatedly across diverse groups. A number of such sex-determining regions (SDRs) have been studied in animals, plants, and fungi, but very little is known about the evolution of sexes in other eukaryotic lineages.
Results: We report here the sequencing and genomic analysis of the SDR of *Ectocarpus*, a brown alga that has been evolving independently from plants, animals, and fungi for over one giga-annum. In *Ectocarpus*, sex is expressed during the haploid phase of the life cycle, and both the female (U) and the male (V) sex chromosomes contain nonrecombining regions. The U and V of this species have been diverging for more than 70 mega-annum, yet gene degeneration has been modest, and the SDR is relatively small, with no evidence for evolutionary strata. These features may be explained by the occurrence of strong purifying selection during the haploid phase of the life cycle and the low level of sexual dimorphism. V is dominant over U, suggesting that femaleness may be the default state, adopted when the male haplotype is absent.
Conclusions: The *Ectocarpus* UV system has clearly had a distinct evolutionary trajectory not only to the well-studied XY and ZW systems but also to the UV systems described so far. Nonetheless, some striking similarities exist, indicating remarkable universality of the underlying processes shaping sex chromosome evolution across distant lineages.

Introduction

Genetic determination of sex is mediated by sex-determining regions (SDRs) of various sizes or by sex chromosomes in a broad range of eukaryotes. Sex chromosomes have arisen independently and repeatedly across the eukaryotic tree, and comparative analysis of different sex-determination systems has provided insights into how these systems originate and evolve. A typical sex chromosome pair is thought to have derived from a pair of autosomes through the acquisition of genes involved in sex determination. If more than one locus involved in sex determination is located on the chromosome, recombination between loci is expected to be suppressed to avoid the production of maladapted individuals with a combination of male and female alleles of the sex-determining genes. This leads to the establishment of a nonrecombining region on the nascent sex chromosome, with important consequences for the evolution of this region of the genome [1]. For example, as a result of the suppression of recombination within the SDR, repetitive DNA tends to accumulate, leading to an increase in SDR size and degeneration of genes within the nonrecombining region. At a later stage, deletion of nonfunctional DNA from within the SDR may lead to a decrease in the physical size of the SDR.

There is also evidence that the nonrecombining region can progressively encroach on the flanking regions of the chromosome so that it encompasses an increasingly greater proportion of the sex chromosome. This process is thought to be driven by the recruitment of genes with differential selective benefits to the two sexes (sexually antagonistic genes) into the SDR [2] (but see [3]). Extension of the SDR in this manner can lead to the creation of "strata," which are regions of the SDR that have become nonrecombining at different points in evolutionary time [4–7].

The genetic mechanism of sex determination also influences how the sex chromosomes evolve. In organisms in which sex is expressed in the diploid phase, such as most animals and land plants, one sex is heterogametic (XY or ZW), whereas the other is homogametic (XX or ZZ). In these systems, only the Y or W contains nonrecombining regions because the X and Z recombine in the homogametic sex. In some algae and bryophytes, the male and female sexes are genetically determined after meiosis, during the haploid phase of the life cycle [8, 9]. This type of sexual system, termed UV to distinguish it from the XY and ZW systems described above [10], exhibits specific evolutionary and genetic properties that have no exact equivalent in diploid systems. In UV systems, the female and male SDR haplotypes function in independent, haploid, male and female individuals, and, consequently, there is no heterozygous sex comparable to XY males or ZW females. This difference between UV and XY/ZW systems should have important implications for SDR evolution [8, 9]. In particular, the female U and the male V are expected to be under similar evolutionary pressures not only because they function

[10]Co-first author
*Correspondence: coelho@sb-roscoff.fr

independently in different individuals but also because neither the U nor the V SDR haplotype recombines [8, 9]. As a result, both haplotypes are expected to exhibit the effects of loss of recombination, such as gene degeneration, to a similar extent. Gene degeneration is, however, expected to be limited in both the U and the V regions, provided they both contain genes that are essential during the haploid phase. It has also been suggested that changes in the size of the U or V involved principally additions of beneficial (but not essential) genes rather than gene losses [8, 9]. Some asymmetry may be expected between the U and V, however, if sexual selection is stronger in males [11] or if one of the chromosomes plays a more active role in sex determination. These verbal predictions of the characteristics of UV systems still need to be rigorously tested empirically.

Although eukaryotic species with UV systems may be as common as those with XY and ZW systems, very few of the former have been characterized, with detailed sequence data being available for only two members of the Archaeplastida lineage: the liverwort *Marchantia* (which has a fully sequenced V chromosome but a partially identified U chromosome) [12] and a UV pair of unknown age in the green alga *Volvox* [13], together with more fragmentary information recently obtained for the moss *Ceratodon* [14]. Clearly, additional detailed sequence information is required to fully test the predictions that have been made with respect to UV sex-determination systems and to evaluate the generality of these predictions in a broad phylogenetic context.

We report here the identification and the genetic and genomic characterization of the U and V sex-determining regions of the brown algal model *Ectocarpus* sp. (formerly included in *E. siliculosus*) [15, 16]. Brown algae belong to the Stramenopiles, a lineage very distantly related to animals, fungi, and green plants (the common ancestors dating back more than one giga-annum [Ga]). The brown algae are considered to possess sex chromosomes rather than mating-type chromosomes [17–19] for a number of reasons: (1) there is a strict correlation between gamete size and sex in anisogamous species; (2) all sexual brown algal species exhibit some form of sexual dimorphism [20, 21]; and (3) heteromorphic sex chromosomes have been identified in some species [22, 23]. Previous work has shown that sex is determined by a single, Mendelian locus in *Ectocarpus* sp. [24]. During the haploid-diploid life cycle of this organism, meiospores, produced by the sporophyte generation, develop into dioicous (separate male and female) gametophytes, which then produce either male or female anisogametes (Figure 1A).

We show here that the *Ectocarpus* sp. UV has features typical of sex chromosomes in other systems, such as low gene density and a large amount of repeated DNA. The male and female SDRs are extremely diverged, reflecting a long independent evolutionary history, which we estimated at more than 70 mega-annum (Ma). Despite its age, the SDR constitutes only one-fifth of the sex chromosome. A possible explanation for this observation was suggested by the low number of sex-biased genes, implying that sexual conflict may be insufficient in *Ectocarpus* sp. to drive extensive SDR expansion. Both the male and female SDR haplotypes showed signs of degeneration despite the action of purifying selection during the haploid phase of the life cycle. Analysis of expression data suggested that the genes escaped degeneration function during the haploid phase of the life cycle. The male SDR haplotype was dominant over the female haplotype, suggesting that the V chromosome determines maleness, with femaleness possibly being the default state when this chromosome is absent. A male-specific high mobility group (HMG) domain gene was identified as a candidate male sex-determining gene. Analysis of the *Ectocarpus* sp. SDR has underlined the universality of sex chromosome evolution across the eukaryotes and has provided important insights into sex chromosome evolution in UV sexual systems.

Results

Identification and Characterization of the *Ectocarpus* sp. SDR

The initial screen for SDR sequence scaffolds used comparative genome hybridization experiments [25] to identify three male-specific scaffolds. PCR-based markers were used to localize these scaffolds to linkage group 30 of the *Ectocarpus* sp. genetic map [26] (Figure 1B; Tables S1A–S1C available online). Searches for additional male SDR scaffolds were then carried out by searching for scaffolds carrying male-specific genes using male and female transcriptomic data and by adapting the Y chromosome genome scan (YGS) method, which uses short-read sequencing and k-mer comparison to identify sex-linked sequences [27] (see the Supplemental Experimental Procedures for further details). Together, these methods allowed the identification of two large sequence scaffolds corresponding to the male SDR haplotype. Sex linkage was systematically verified by genetic mapping (Tables S1B and S1C).

Further analysis of the segregation patterns of genetic markers corresponding to SDR scaffolds in a single family of 2,000 siblings detected no recombination events (Figure 1B). The SDR therefore behaves as a discrete, nonrecombining haplotype. This genetic analysis indicated that the male SDR extended over a region of approximately 920 kilobase pairs (kbp) (Figure 1C; Table 1).

To characterize the female haplotype of the sex locus, we sequenced the genome of a female *Ectocarpus* sp. strain that is closely related to the sequenced male strain (Figure S1A) [16]. Several strategies were used to identify candidate female SDR scaffolds (Supplemental Experimental Procedures and Tables S1E–S1H). These included searches for female orthologs of male SDR protein sequences, a search for scaffolds carrying female-specific genes based on male and female transcriptomic data, and the adaptation of the YGS method [27] to search for female rather than male scaffolds. The cumulative size of the female sex-linked scaffolds was 929 kbp. Assuming that the combination of approaches used here has provided a near-complete list of male and female SDR scaffolds, this indicates that the male and female SDR haplotypes are of similar size (Figure 1C; Table 1).

To confirm cosegregation of the SDR with sexual phenotype, 34 *Ectocarpus* strains of known sex from different geographical origins and species were genotyped with several sex locus markers, corresponding to both the male and female SDR haplotypes (Table S1D). In all cases, the SDR genotype correlated with sexual phenotype, confirming that this region is the sex-determining locus in *Ectocarpus*.

The SDR is flanked by two large recombining regions, which we refer to as pseudoautosomal (PAR) domains. Analysis of molecular marker segregation [26] indicates that these regions recombine during meiosis, unlike the SDR (Figure 1B). The PAR had gene density, intron length, and percent GC content intermediate between those of the autosomes and the SDR (Figure 1B; Table 1). These unusual features are characteristic

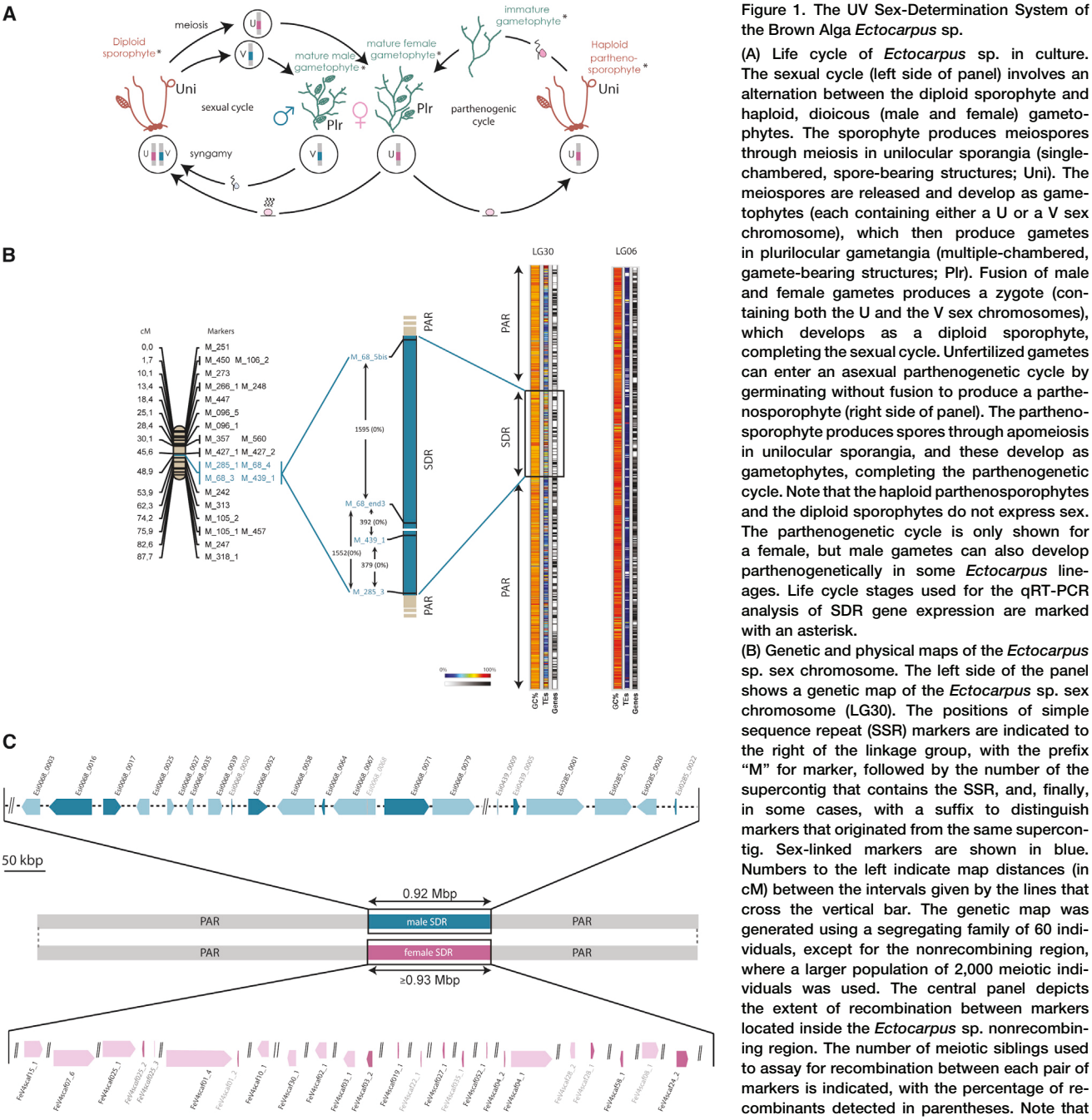


Figure 1. The UV Sex-Determination System of the Brown Alga *Ectocarpus* sp.

(A) Life cycle of *Ectocarpus* sp. in culture. The sexual cycle (left side of panel) involves an alternation between the diploid sporophyte and haploid, dioicous (male and female) gametophytes. The sporophyte produces meiospores through meiosis in unilocular sporangia (single-chambered, spore-bearing structures; Uni). The meiospores are released and develop as gametophytes (each containing either a U or a V sex chromosome), which then produce gametes in plurilocular gametangia (multiple-chambered, gamete-bearing structures; Plr). Fusion of male and female gametes produces a zygote (containing both the U and the V sex chromosomes), which develops as a diploid sporophyte, completing the sexual cycle. Unfertilized gametes can enter an asexual parthenogenetic cycle by germinating without fusion to produce a parthenosporophyte (right side of panel). The parthenosporophyte produces spores through apomeiosis in unilocular sporangia, and these develop as gametophytes, completing the parthenogenetic cycle. Note that the haploid parthenosporophytes and the diploid sporophytes do not express sex. The parthenogenetic cycle is only shown for a female, but male gametes can also develop parthenogenetically in some *Ectocarpus* lineages. Life cycle stages used for the qRT-PCR analysis of SDR gene expression are marked with an asterisk.

(B) Genetic and physical maps of the *Ectocarpus* sp. sex chromosome. The left side of the panel shows a genetic map of the *Ectocarpus* sp. sex chromosome (LG30). The positions of simple sequence repeat (SSR) markers are indicated to the right of the linkage group, with the prefix "M" for marker, followed by the number of the supercontig that contains the SSR, and, finally, in some cases, with a suffix to distinguish markers that originated from the same supercontig. Sex-linked markers are shown in blue. Numbers to the left indicate map distances (in cM) between the intervals given by the lines that cross the vertical bar. The genetic map was generated using a segregating family of 60 individuals, except for the nonrecombining region, where a larger population of 2,000 meiotic individuals was used. The central panel depicts the extent of recombination between markers located inside the *Ectocarpus* sp. nonrecombining region. The number of meiotic siblings used to assay for recombination between each pair of markers is indicated, with the percentage of recombinants detected in parentheses. Note that no recombination was detected between any of the sex locus markers. See Table S1B for the coordinate position of each marker on its respective scaffold. The right side of the panel shows a physical map of the sex chromosome and a heatmap of the GC percent, gene density, and TE density along the LG30 and along an autosome (LG06) for comparison. The heatmap was computed using a 4,000 base pair (bp) sliding window.

(C) Overview of the *Ectocarpus* sp. male and female SDR haplotypes. Genes are indicated by arrows, with the lighter colors corresponding to gametologs. Gene names (LocusIDs) are indicated, with pseudogenes in gray font and putative transposon remnants in gray italics. Putative transposon remnants were counted as protein-coding genes, but Esi0068_0068/FeV4scaf25_3 was not included in the set of gametolog pairs. The relative sizes of the male and female SDR genes are indicated, but they are not drawn to the same scale as the underlying scaffolds indicated by the dotted line and the scale bar. Only female SDR scaffolds carrying genes are represented. Scaffolds are separated by double diagonal lines, indicating that the relative positions of scaffolds within the SDR are unknown. Double-headed arrows indicate the estimated sizes of the SDR haplotypes. The gray bars indicate the sex chromosomes. SDR, sex-determining region; PAR, pseudoautosomal region. See also Figure S1.

of the entire recombining part of the chromosome and are not restricted to the regions closest to the SDR (Figure 1B). It is currently not clear why the PAR exhibits these structural differences compared to the autosomes.

Both the male and female SDR haplotypes are rich in transposable element sequences (Figure 1B; Figure 2A) and gene poor compared to the autosomes (Table 1), features typical of nonrecombining regions [1]. With only one exception (long

Table 1. Statistics for Several Features of the Male and Female *Ectocarpus* sp. SDR Compared with the PAR and the Complete Genome

	Male SDR	Female SDR	PAR	Genome
Total sequence (Mbp)	0.92	0.93	4.08	205.27
Genes (including pseudogenes)	20	24	228	15,779
Average gene length (bp)	25,710	18,836	8,188	6,974
Average CDS length (bp)	1,373	1,050	1,217	1,607
Average intron length (bp)	3,605	3,691	1,062	702
Average number of introns per gene	6.67	4.81	6.28	7.14
Gene density (genes per Mbp)	22.82	23.66	55.88	76.87
GC (%)	51.29	44.74	52.20	54.02

terminal repeat transposons in the female SDR), all transposable element (TE) classes were more abundant in the SDR and the PAR than they were in the autosomes, with the differences being particularly marked for both SDR haplotypes. When individual classes of transposable elements were considered, retrotransposons (which represent the least abundant transposon class in the *Ectocarpus* sp. genome as a whole) showed the most marked proportional enrichment in the SDR haplotypes compared to the autosomes (Figure S2A).

About 30% of the euchromatin of the male-specific (nonrecombining) region of the human Y chromosome consists of multiple, different "ampliconic sequences," which exhibit 99.9% identity within each set of repeated sequence. The identity between these sequences has been taken as evidence for a high level of gene conversion within this region [5, 30]. It was further suggested that gene conversion might "substitute" for interchromosomal recombination to some extent, counteracting the degenerative effects of reduced recombination within the SDR. Very little intrahaplotype sequence similarity was identified within either the male or the female *Ectocarpus* sp. SDR haplotypes (Table S1J). The total lengths of the repeated regions within the male and female SDRs were only 2.5% and 3.2%, respectively. It therefore seems unlikely that mechanisms similar to those proposed for the human Y chromosome have operated in this SDR, although it should be noted that large ampliconic repeats are difficult to assemble, and some sequences of this type may not have been identified, particularly for the female haplotype.

The male SDR haplotype contains 17 protein-coding genes and three pseudogenes, whereas 15 protein-coding genes and seven pseudogenes were found in the female haplotype (Figure 1C; Figure 3; Table S2). Eight of the female protein-coding genes and three of the pseudogenes are homologous to male SDR sequences ("gametologs"), consistent with the two SDR haplotypes having evolved from a common ancestral autosomal region. The classification of these genes as gametologs was supported by expression analysis, which showed that transcript abundances for gametolog pairs were strongly correlated (Figure S2B), and by their conserved intron and exon structures (Figure S3). This correlated expression pattern is consistent with the gametolog genes having been retained because they have non-sex-specific functions during the haploid phase of the life cycle. The genes and pseudogenes that were only found in one (male or female) haplotype may have been either acquired since the divergence of the U and the V regions or lost by the counterpart haplotype. Eighteen of the male and female genes and pseudogenes that were found in only one haplotype had homologs outside the SDR (including, in two cases, genes on linkage group 30; Figure 3 and Table S2). The high similarity between some of these SDR genes and their closest autosomal homologs would be consistent with these gene pairs having arisen from recent gene duplication events (i.e., since the divergence of the U and the V) that created either the SDR or the autosomal copy. The remaining two genes that were found in only one haplotype may represent cases of gene loss in the other haplotype, but they could also have resulted from gene relocation to the SDR. Testing these hypotheses will require comparison with a homologous gene from an outgroup species.

Genomic Degeneration of the SDR Region

Suppression of recombination across the SDR is expected to lead to genetic degeneration unless there is strong selection on gene function to counteract this effect. There are several indications that genetic degradation has occurred, at least to some degree, in the *Ectocarpus* sp. SDR. We identified a set of optimal codons for *Ectocarpus* sp. (Figures S2C and S2D). Selection on codon usage is known to be of weak intensity and particularly sensitive to loss of recombination [31, 32]. The coding sequences of SDR genes exhibited significant underrepresentation of optimal codons (Figure 2B). This suggests maladapted codon usage (although we cannot exclude that the underrepresentation is due, at least in part, to reduced rates of biased gene conversion [33] due to the loss of recombination within the SDR). In addition, transcripts of SDR genes tended to be less abundant on average than transcripts of autosomal genes, although note that codon usage and expression level are likely to be correlated, so these two parameters are not necessarily independent. Reduced transcript abundance was particularly marked for SDR genes that were exclusively present in one of the haplotypes (Figure 2C), and it may reflect degradation of the promoter and *cis*-regulatory sequences of these SDR genes. The same tendency was observed for the *Volvox* mating locus, where haplotype-specific genes were expressed at lower levels than genes that were part of a gametolog pair [13], suggesting that genetic degeneration of haplotype-specific SDR genes may be a general phenomenon. Note that expression analysis of the *Ectocarpus* sp. gametolog genes did not provide any evidence that these genes are degenerating.

SDR genes were found to be much longer on average than genes elsewhere in the genome, due principally to the presence of longer introns (Table 1). This difference was partly explained by the presence of a larger amount of inserted transposable element DNA (Figures 2A and S2E), which is typical of nonrecombining regions.

Although these various analyses provided some evidence for genomic degeneration in the SDR, the overall degree of degeneration was modest compared to previously characterized systems [34], perhaps because both the U and the V SDR haplotypes have essential functions during the haploid phase and are constantly exposed to selection (in contrast to Y or W chromosome genes, which are always heterozygous). An analysis of SDR gene expression supported this hypothesis: transcripts of SDR genes were consistently present during the haploid phase of the life cycle (Figure 4). Another potential explanation for the limited degree of degeneration is that the SDR is small compared to most previously characterized systems, and this may have limited the potential for Hill-Robertson interference among selected sites [35–37].

Predicted Functions of SDR Genes

Of the nine genes that were found in the male, but not the female, SDR haplotype, one was of particular interest because

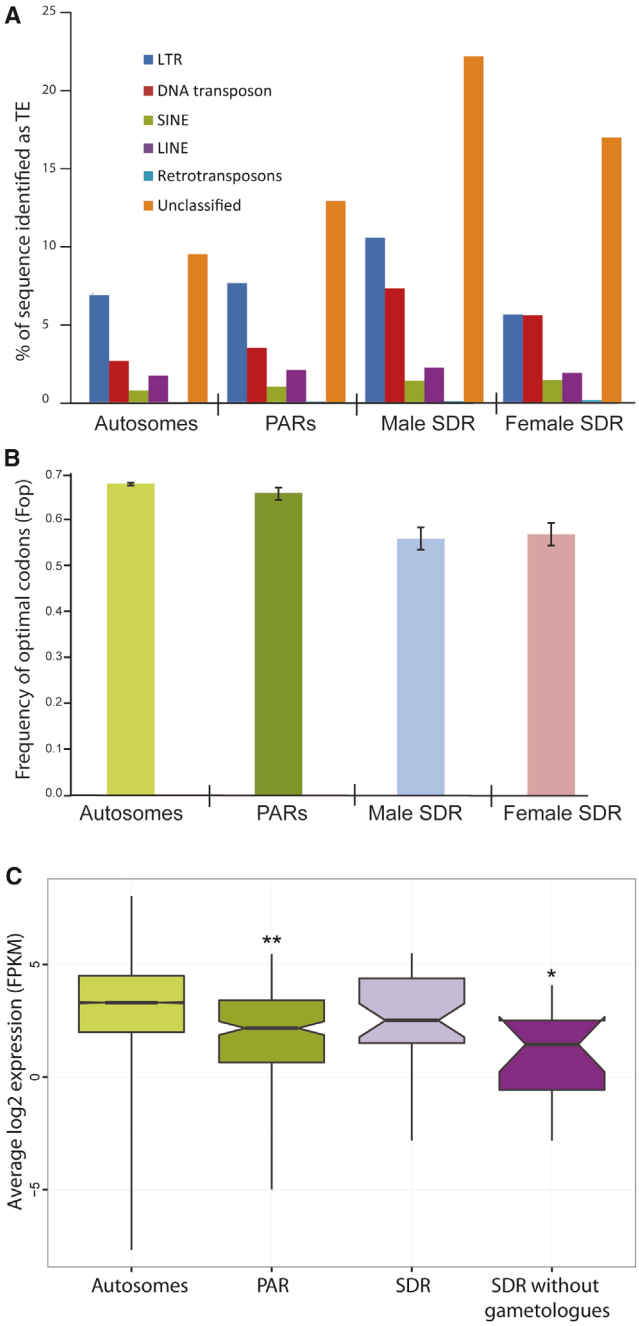


Figure 2. Comparison of Genomic Features of the SDR, PAR, and Autosomes

(A) Percentage of DNA corresponding to different classes of transposable elements (TEs) in different genomic fractions. Pairwise comparisons using a Fisher's exact test indicated that all of the sex chromosome compartments (PAR, male SDR, female SDR) were significantly different from the autosomal compartment ($p < 0.0001$).

(B) Median frequency of optimal codons in coding regions of autosomal, PAR, and male and female SDR genes. Error bars indicate 95% confidence intervals around the median. An analysis using the codon adaptation index (CAI, another codon usage index [28], which was computed using R and the seqinR package [29]) gave similar results.

(C) Mean transcript abundance in sexually mature, male and female gametophytes for genes in different genome fractions, determined by RNA-seq and expressed as fragments per kilobase per million reads (FPKM) mapped. The notched boxplot graph shows the means of autosomal genes (n = 14,677), PAR genes (n = 205), male and female SDR genes (n = 37), and SDR without gametolog genes (n = 16). Significant adjusted p values compared with autosomes, as calculated by Wilcoxon tests, are indicated by asterisks above each box (*$p < 0.01$, **$p < 0.001$).

it was predicted to encode an HMG domain protein (Figure S4A and Table S4A). This family of proteins has been implicated in sex or mating-type determination in both vertebrates and fungi [38, 39]. The SDR of the green alga *Volvox* also contains an HMG gene [13]. In addition, several of the genes that were found in both the male and female SDR haplotypes (gametologs) were predicted to encode potential signal transduction proteins (including a Ste20-like kinase, a casein kinase, a GTPase, a RING zinc-finger protein, and a MEMO domain protein; Table S2) and could potentially be involved in the regulation of sex determination.

An Ancient Sex-Determining Region

At the sequence level, the male and female haplotypes are extremely divergent. No large blocks of sequence similarity were found, and the only regions with a high level of similarity corresponded to gametolog exons (Figure S3). This divergence suggests that the male and female haplotypes have been evolving independently over a long period. Two phylogenetic trees were constructed based on sequences of either an SDR or an autosomal sequence from three *Ectocarpus* lineages and three distantly related brown algal species, *Scytosiphon lomentaria*, *Sphaerotrichia firma*, and *Laminaria digitata*. The topology of the phylogenetic tree based on the autosomal region was consistent with sequential speciation, with sequences from male and female strains of the same lineage grouping together (Figure 5A). In contrast, in the phylogenetic tree based on the SDR gene, sequences grouped together according to gender (Figure 5B). Note that we were not able to obtain sequence for this gene from female *L. digitata* individuals, suggesting that they may have lost the female gametolog. These data suggest that the SDR originated at least 70 million years ago and may be substantially older. The rate of synonymous site mutations (dS) in the coding regions of the 11 male and female gametolog pairs (Figure 5C) was used to independently evaluate the age of the SDR. The dS values for these gene pairs were compared with values for orthologous, autosomal gene pairs across 12 brown algal and diatom species for which divergence times had been estimated (Supplemental Experimental Procedures). The dS values for the SDR genes were remarkably high (mean value of 1.7, with most genes having $dS > 1$), and comparisons with values obtained for the pairs of autosomal orthologs indicated that the male and female haplotypes of the SDR stopped recombining more than 100 million years ago (Figure S5). Note, however, that the estimations based on genetic divergence are approximate because of saturation of synonymous site mutations at the evolutionary distances measured. These analyses suggest that the *Ectocarpus* sp. UV SDR is an old system, comparable to the *Drosophila* (60 Ma) [34] and mammalian (180 Ma) [41, 42] XY systems.

When dS values were calculated on an exon-by-exon basis, individual exons with a markedly lower dS value than those of the other exons within the gametolog gene pair were identified for 3 of the 11 gametolog pairs (Figure S3). The presence of these rare variant exon pairs suggests that gene conversion events affecting individual exons or small gene regions may have occurred since the divergence of the male and female SDR haplotypes, but more detailed studies are needed to address this possibility.

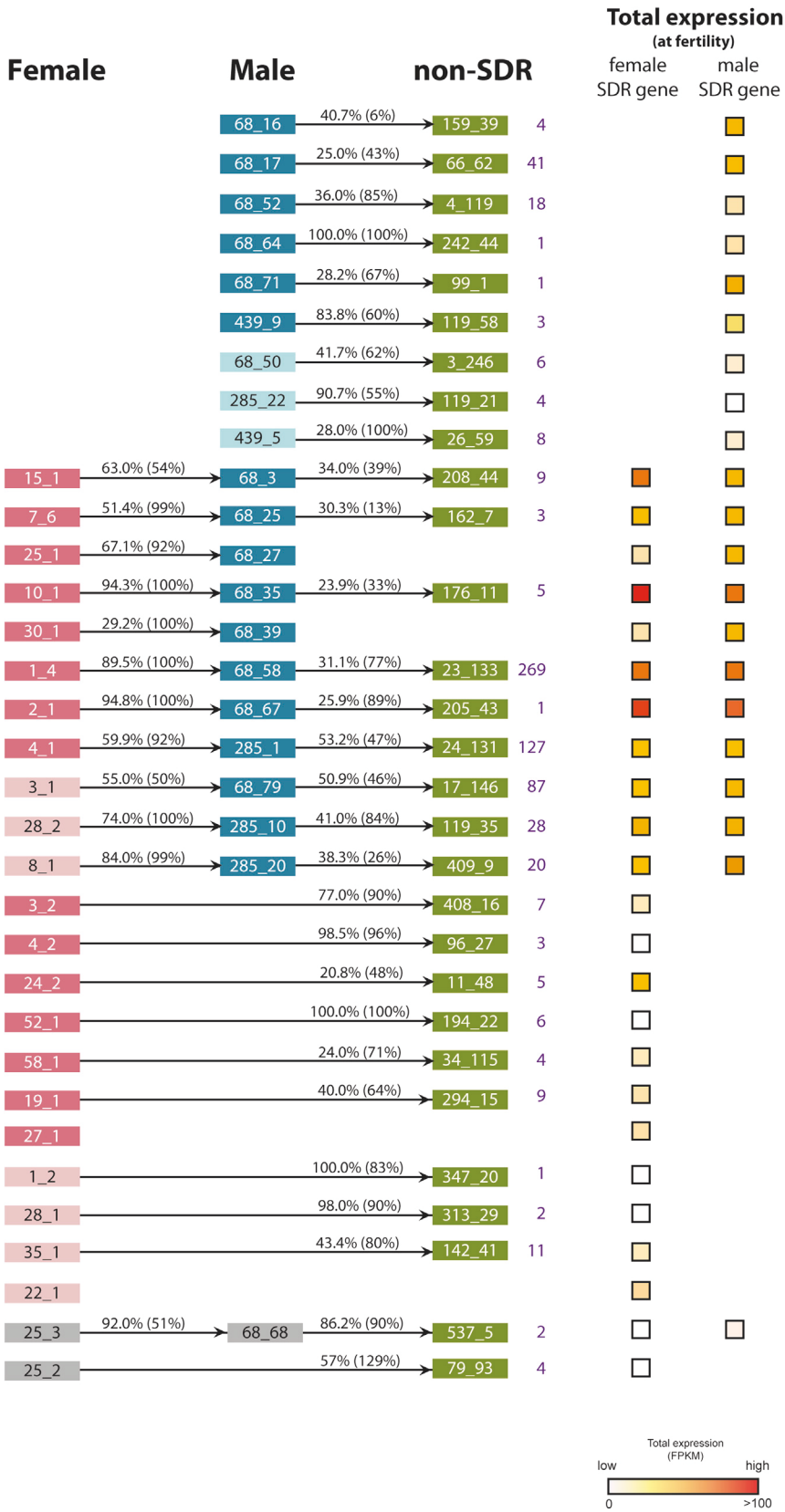


Figure 3. Relationships between SDR Genes and Autosomal Genes and Expression Patterns of the SDR Genes

Schematic diagram showing homology relationships between male and female SDR genes and autosomal genes. Autosomal or PAR (i.e., non-SDR) genes are shown in green; male and female SDR genes are shown in blue and pink, respectively, with putative functional genes in dark blue or dark pink and pseudogenes in light blue or light pink. Putative transposon remnants are shown in gray. A green box indicates the existence of at least one homolog outside the SDR, and the number to the right of the green box indicates the number of matches outside the SDR with an E value of less than 10^{-4}. Homology relationships were defined based on a BLASTP E value of less than 10^{-4} when predicted protein sequences were blasted against the complete set of *Ectocarpus* sp. predicted proteins. Percentage identity between predicted proteins is indicated above the arrows. The value in parentheses corresponds to the length of the matched region as a percentage of the total length of the protein to the left of the arrow. Gene abbreviations are as in the following examples: for male SDR or non-SDR genes, 68_16 indicates Esi0068_0016; for female SDR genes, 15_1 indicates FeV4scaf15_1. Note that the order of the genes is not intended to correspond to their locations in the genome. The right side of the panel depicts transcript abundances for each of the male and female SDR genes in male and female mature gametophytes, respectively, measured by RNA-seq and expressed as FPKM. See also Figure S2.

Limited Expansion of the *Ectocarpus* sp. SDR

Given its age and the prediction that an SDR should progressively enlarge over time to encompass a large part of its chromosome [1, 43], it is remarkable that the *Ectocarpus* sp. SDR accounts for only about one-fifth of linkage group 30 and extends over less than one megabase pair (Mbp). It is possible that the small size of the SDR is related to the low level of sexual dimorphism in *Ectocarpus* sp. because the recruitment of sexually antagonistic genes is believed to be an important driver of SDR expansion [1, 43]. Moreover, sexually antagonistic polymorphisms are predicted to be less stable in haploid systems than in diploid systems because dominance effects in XX (or ZZ) individuals are expected to favor allele maintenance in the latter [44, 45]. This effect may also limit expansion of the SDR by reducing the number of genes with sexually antagonistic polymorphisms available for recruitment into the SDR. Consistent with these hypotheses, comparison of the transcriptomes of male and female gametophytes indicated that only about 4% of *Ectocarpus* sp. genes showed sex-biased expression at the mature sexual stage of the life cycle (compared, for example, with up to 50%–75% in *Drosophila* [46, 47]; Table S4C).

SDR Gene Expression and Dominance

Quantitative PCR was used to measure the abundance of SDR gene transcripts in near-isogenic male and female strains (Figure 4) at different stages of the life cycle (Figure 1A). Whereas no clear pattern was observed for the female SDR genes, transcripts of two-thirds of the male SDR genes that were analyzed were most abundant in mature gametophytes (Figure 4), suggesting that these genes have a role in fertility. Interestingly, the transcript of the male gene that is predicted to encode an HMG domain protein was more than 10-fold more abundant in mature gametophytes than at the other stages assayed (Figure 4). The other fertility-induced genes included both additional male-specific genes (encoding conserved unknown proteins) and several gametolog pairs (predicted to encode, for example, a GTPase, a MEMO-like domain protein, a nucleotide transferase, and a homoaconitate hydratase; Table S2).

Diploid gametophytes bearing both the male and the female SDR haplotypes (UV) can be generated artificially, and these individuals are always phenotypically male, indicating that the male haplotype is dominant [24, 48]. This dominance relationship would be consistent with the existence of a master regulatory gene that determines maleness, carried by the V chromosome. To determine whether the dominance of the male haplotype is dose dependent, we used the life cycle mutant *ouroboros* [48] to construct 13 independent triploid (UUV) and tetraploid (UUUV) gametophytes (Figure S1A and Table S1I). All tested polyploids produced male gametes (as determined by genetic crosses with tester lines). Measurements of transcript abundances for 11 female SDR genes did not detect a marked downregulation of these genes in diploid heterozygous gametophytes compared to haploid gametophytes (Figures S4B and S4C). This suggests that the male haplotype does not silence female gene expression in this heterozygous context (although it was not possible to rule out that the expression of specific female haplotype genes was suppressed). It is likely, therefore, that gametophytes adopt the female developmental program by default, when the male SDR haplotype is absent.

Discussion

This study has demonstrated that sex is determined during the haploid phase of the brown alga *Ectocarpus* sp. by a nonrecombining region on linkage group 30 that extends over almost 1 Mbp. The male and female haplotypes of the SDR were of similar size but were highly diverged, with the only significant similarity being the presence of 11 gametologs, three of which were predicted to be pseudogenes in the female. Based on comparisons of these shared genes across diverse brown algal species, the SDR was estimated to be more than 100 million years old. Compared with previously characterized systems [49], the *Ectocarpus* sp. UV chromosomes can clearly be classed as an ancient (as opposed to a recently evolved) sex-determining system.

The brown algae belong to the Stramenopiles, which diverged from the lineages that led to green plants and animals more than one billion years ago [50]. This study therefore confirms that SDRs from diverse eukaryote groups share a number of fundamental features, such as stable maintenance of pairs of functional alleles (gametologs) over long periods of evolutionary time, suppressed recombination within the SDR, low gene density, and accumulation of transposable elements. The presence of 11 gametolog pairs provided unambiguous evidence that the *Ectocarpus* sp. UV pair is derived from an ancestral pair of autosomes, as has been observed for XY and ZW systems in animals and plants [1, 7, 43].

Analysis of the *Ectocarpus* sp. SDR has also allowed a number of predictions that specifically concern UV sexual systems [8, 9] to be tested. UV systems are not expected to exhibit the asymmetrical degeneracy of the sexual chromosomes (degeneracy of the Y and W chromosomes) observed in XY and ZW systems [34], and this supposition is supported by the similar estimated sizes of the male and female SDR haplotypes in *Ectocarpus* sp. Based on parameters such as transcript abundance and frequency of optimal codons, the *Ectocarpus* sp. SDR genes exhibit evidence of degeneration, but the degree of degeneration is modest compared to that observed for Y-located genes in XY systems of comparable age [34]. Because transcripts of all the SDR genes were detected in the gametophyte generation, the modest degree of degeneration is consistent with purifying selection acting to maintain gene functionality during the haploid phase, when the U and V chromosomes are found in separate male and female organisms. Selection is indeed expected to be stronger during the haploid phase, and it is expected to limit degeneration, as suggested for the V chromosome of *Marchantia* [12], another UV system, and by the low nonsynonymous to synonymous site mutation (dN/dS) ratios observed for sex-linked pollen-expressed genes in *Silene latifolia*, a plant with XY chromosomes [51]. The detection of modest levels of gene degeneration indicates that UV SDRs are nonetheless subject to the degenerating effects of suppressed recombination to some degree. Expression analysis indicated that in *Ectocarpus* sp., the SDR genes that escape degeneration belong principally to gametolog pairs, which presumably play a role during the haploid phase, or are male haplotype-specific genes, which are presumably required for male fertility. The *Ectocarpus* sp. SDR contains a large proportion of sex-specific genes (20 male and female sex-specific genes compared with only 11 gametolog pairs). This situation contrasts markedly with the UV system of *Volvox*, where the vast majority of the mating region genes are shared between haplotypes [13]. This difference in gene composition suggests that these two UV systems have had different evolutionary histories, perhaps having been affected in different ways by gene gain and gene loss events. Bull predicted that changes in the sizes of the U and V SDR haplotypes should be due to gain of genes beneficial to the gametophyte

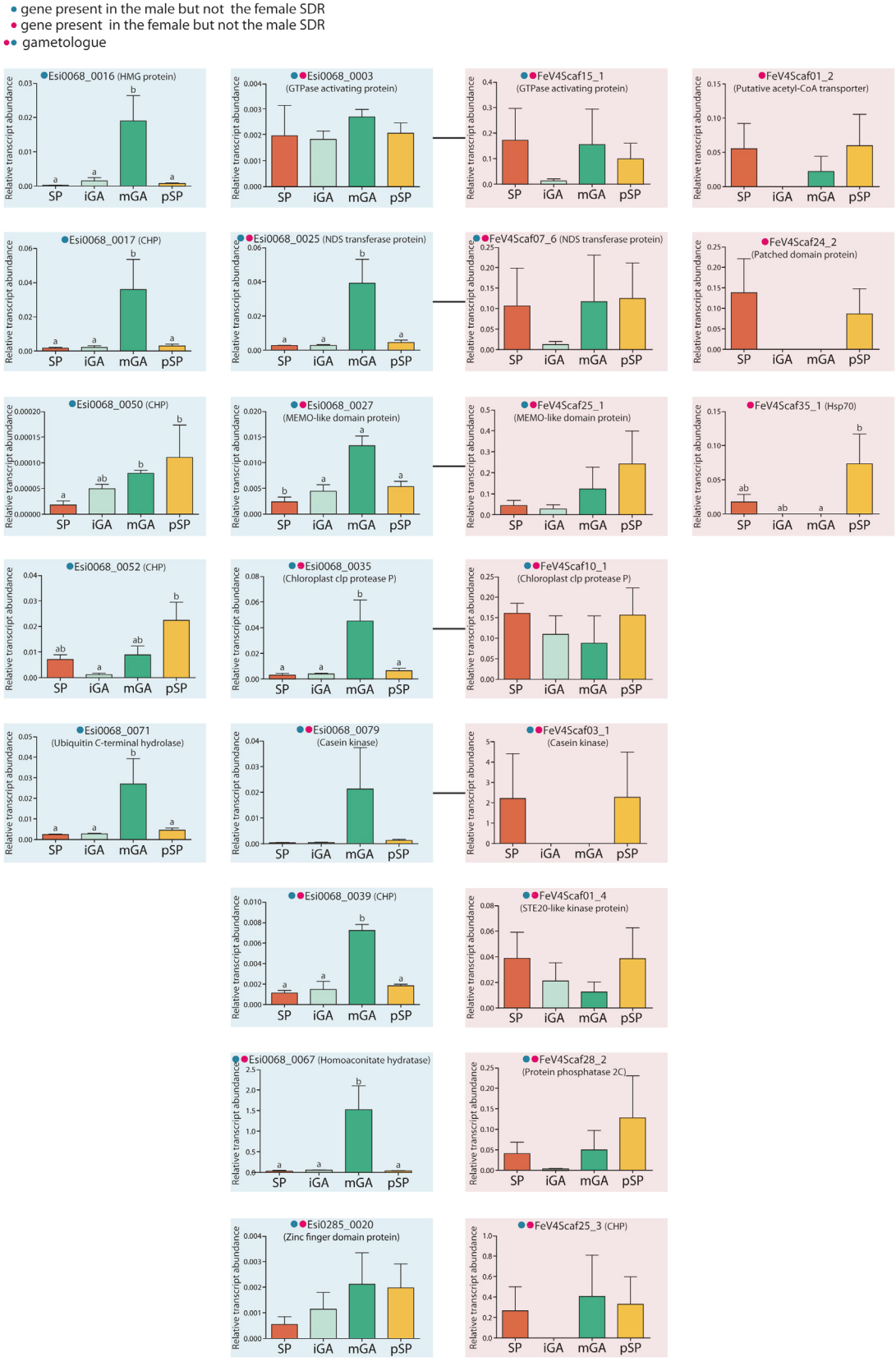


Figure 4. SDR Gene Expression during the Life Cycle

Male and female SDR gene expression during the life cycle of *Ectocarpus* sp. measured by qRT-PCR, relative to a housekeeping gene (EF1α). Gene annotations are indicated in parentheses (see Table S2 for further details). Abundances of transcripts for female and male SDR genes were measured using RNA from

(legend continued on next page)

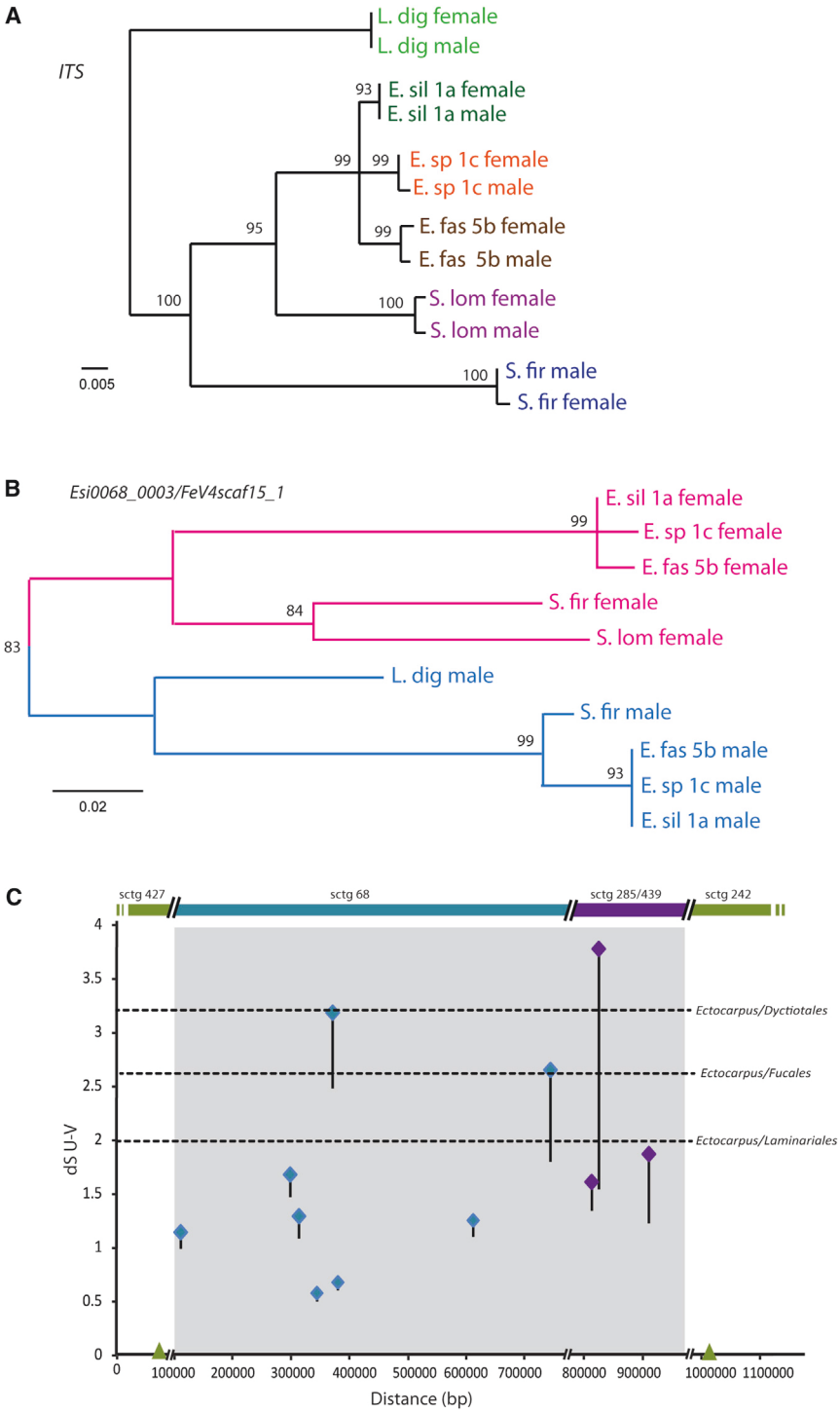


Figure 5. Estimation of the Age of the *Ectocarpus* sp. SDR

(A) Maximum likelihood tree created in MEGA5 [40] based on the Kimura 2-parameter model using sequence data amplified from 453 bases of the autosomal region ITS2 and adjacent 5′-LSU. The percentage of trees in which the associated taxa clustered together (bootstrap values from 1,000 resamplings) is shown next to the branches. Initial trees for the heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using the maximum composite likelihood (MCL) approach and by then selecting the topology with the best log likelihood value. A discrete gamma distribution was used to model evolutionary rate differences among sites (five categories, +*G*, parameter = 0.2094). Distinct lineages are indicated by different colors. Samples correspond to three different *Ectocarpus* lineages, *E. siliculosus* lineage 1a (E. sil 1a), *E. sp.* lineage 1c (E. sp 1c), and *E. fasciculatus* lineage 5b (E. fas 5b), and three distantly related brown algae, *Sphaerotrichia firma* (S. fir), *Scytosiphon lomentaria* (S. lom), and *Laminaria digitata* (L. dig). Lineage names and sex are indicated at the branch tips. The strains used are described in Table S1A.

(B) Maximum likelihood tree with equivalent parameters to those shown in (A) (gamma distribution, +*G*, parameter = 0.2868) for 148 bases of the sex-linked, exonic region of one gametolog pair (Esi0068_0003/FeV4scaf15_1). Pink and blue indicate sequences from female and male individuals, respectively.

(C) Plot of dS values of gametolog and PAR homologous pairs against gene distance, with gene order according to the male physical map. Blue and purple lozenges represent genes on the two male SDR scaffolds, sctg_68 and sctg_285and439, respectively. Green triangles at each end of the x axis represent two flanking PAR genes. One-sided SE bars represent half the SE of the estimation. Double diagonal bars indicate that the orientation of the locus relative to the flanking PAR is not known. Dotted lines indicate mean levels of synonymous site divergence between *Ectocarpus* sp. autosomal genes and autosomal genes of species from the brown algal groups indicated.

See also Figure S5.

rather than gene loss [8, 9]. The presence of a large proportion of haplotype-specific genes in the *Ectocarpus* sp. SDR, relative to the gametologs, and the expression patterns of many haplotype-specific genes, which indicate a role during fertility, would be consistent with his prediction. However, because there is an autosomal paralog for most of these haplotype-specific genes, it is also possible that functional redundancy of SDR genes and their autosomal paralogs allowed gene loss to occur. Future analysis of additional related SDRs, together with an outgroup species in which the region homologous to the *Ectocarpus* sp.

gametophytes and parthenosporophytes of strains carrying either the U or the V sex chromosome, respectively, and from diploid sporophytes (strains carrying both the U and the V). Bars with different letters are statistically different ($p < 0.05$). Details on the statistical analysis are presented in the Supplemental Experimental Procedures. The colored dots next to gene names indicate whether the gene is a gametolog (blue and pink dots) or whether it is only found in either the male or the female haplotype (blue or pink dot, respectively). Graphs corresponding to gametolog pairs are linked by a horizontal line. SP, diploid heterozygous sporophyte; iGA, immature gametophyte; mGA, mature gametophyte; pSP, parthenosporophyte; CHP, conserved hypothetical protein.

SDR is autosomal, may help to trace changes in SDR gene content over evolutionary time and determine the relative importance of gene gain and gene loss during the emergence of this system.

Despite being ancient, the *Ectocarpus* sp. SDR is quite small. Given the low level of sexual dimorphism in *Ectocarpus* sp. and the small number of genes that show sex-biased expression, both of which suggest that there is limited scope for sexual conflict, the small size of the SDR is consistent with the view that SDR expansion is driven by the evolution of genes with sexually antagonistic effects [1, 52]. In a number of sex chromosome systems, the expansion of the nonrecombining region of the Y (or W) has been shown to have proceeded through several events of recombination suppression, and these recombination events have formed regions with different degrees of X-Y (or Z-W) divergence (evolutionary strata) [4, 53] (reviewed in [1, 49]). The lack of detectable strata is consistent with the conclusion that this region has experienced limited expansion. However, given that strata may be extremely difficult to detect in ancient haploid systems (because both U and V can accumulate rearrangements), we cannot totally rule out the absence of these events. Indeed, recent evidence suggests the possible existence of at least two recombination suppression events in the UV system of the bryophyte *Ceratodon* [14], and therefore that UV systems may acquire evolutionary strata in some cases. Note also that the *Ectocarpus* sp. system provides independent evidence that the age of an SDR does not necessarily correlate perfectly either with its size or with the degree of heteromorphy (e.g., [54, 55]).

In *Ectocarpus* sp., the male SDR haplotype was dominant over the female haplotype, even when three copies of the female haplotype were present. It is therefore possible that femaleness may simply be the default state, adopted when the male haplotype is absent. This situation is comparable to that observed in diverse animal, fungal, and land plant sex-determination systems but differs from that observed with the UV systems of some mosses. In the latter, the male and female factors are codominant, leading to monoicy when both the male and female SDR haplotypes are present in the same gametophyte [56]. Functional differences can therefore be observed between different sex-determination systems, independent of the genetic nature of the system (XY, ZW, or UV).

The male-specific HMG gene is a good candidate for the gene that determines maleness in *Ectocarpus* sp. If this can be confirmed experimentally, it will raise important questions about the evolution of sex and mating-type-determination gene networks across the eukaryote tree, suggesting shared or convergent mechanisms in brown algae, fungi, and animals.

Experimental Procedures

Ectocarpus Culture

Ectocarpus strains were cultured as described [57].

RNA-Seq Transcriptome Data

RNA sequencing (RNA-seq) analysis was carried out to compare the abundances of gene transcripts in male and female mature gametophytes. Synchronous cultures of gametophytes of the near-isogenic male and female lines Ec603 and Ec602 (see Table S1A and Figure S1) were prepared under standard conditions [57] and frozen at maturity. Total RNA was extracted from 2 bulks of 400 male individuals and 2 bulks of 400 female individuals (two biological replicates for each sex) using the QIAGEN Mini kit (http://www.qiagen.com) as previously described [48]. For each replicate, RNAs were quantified, and cDNAs for transcriptome analysis were polythymine primed, fragmented, cloned, and sequenced by Fasteris. We used both de novo assembly (Trinity) (r2012-01-25) [58] and TopHat (v.2.0.8) [59, 60] and Cufflinks (v.2.1.1) [60, 61] algorithms. Statistical testing for sex-biased gene expression was performed using DEseq [62].

Identification and Mapping of the Male SDR

A comparative genome hybridization approach [25] identified several regions of the genome exhibiting polymorphisms between male (Ec32) and female (Ec568) strains. Primers were developed for these putative sex-linked regions, and mapping was performed by genotyping the 60 individuals of the mapping population [26]. Details of the PCR conditions are given in the Supplemental Experimental Procedures. The approaches used to improve the assembly of the male SDR and to verifiy the completeness of the male SDR using both an RNA-seq-based method and an approach based on the YGS method developed by Carvalho and Clark [27] are described in detail in the Supplemental Experimental Procedures.

Recombination Analysis

Recombination between sex locus markers was analyzed using a large segregating family of 2,000 meiotic individuals (Figure S1) derived from a cross between the male line Ec494 [48] and the female outcrossing line Ec568 [26].

Sequencing of a Female Strain and Identification and Assembly of the Female SDR

The genome of the female strain Ec597 (Table S1A and Figure S1A) was sequenced using a whole genome shotgun strategy that involved the implementation of both Illumina HiSeq 2000 technology and Roche 454 pyrosequencing. Velvet (v.1.1.05) was used to run several assemblies during the sequencing process, including the v.3 assembly (which used all the paired-end reads and reads from one of the mate-pair libraries) and the final v.4 assembly with the complete read data set (Table S1E). An independent de novo assembly was also carried out with the CLC assembler (http://www.clcbio.com/products/clc-assembly-cell) using only the paired-end Illumina data.

Female SDR scaffolds were identified using three different approaches. First, we blasted the deduced protein sequences of male SDR genes (all annotated genes on the two male SDR scaffolds sctg_68 and sctg_285and439) against the female genome assembly. Fourteen candidate female SDR scaffolds were identified in the V4 assembly using this approach. Second, we used an approach that employed RNA-seq transcriptome data. Third, we also adapted the YGS method [27] to identify female-linked sequences. These approaches are described in detail in the Supplemental Experimental Procedures. All putative female-specific scaffolds were verified by PCR using between 8 and 57 individuals. Several approaches were used to improve the assembly of the female SDR. Details are given in the Supplemental Experimental Procedures.

Annotation of SDR Scaffolds

The male SDR scaffolds had been annotated as part of the *Ectocarpus* sp. genome project [16], but the gene models were considerably improved by integrating transcript information derived from the RNA-seq analysis carried out as part of this study and by using comparisons of male and female gametolog gene models. The updated gene models can be accessed on the OrcAE database (http://bioinformatics.psb.ugent.be/orcae/overview/Ectsi) [63]. The female SDR scaffolds were annotated de novo by running the gene prediction program EuGène [64], which incorporated the signal prediction program SpliceMachine [65], using the optimized Markov models and SpliceMachine splice site predictions derived previously for the male genome sequence [16]. Gene prediction incorporated extrinsic information from mapping of the RNA-seq data onto the female-specific scaffolds. Both male and female SDR gene models were manually curated using the raw, mapped RNA-seq data, the Cufflinks and Trinity transcript predictions, and the comparisons between the male and female haplotypes.

Pseudogenes were identified manually by comparing SDR sequences with genes in the public databases. An additional screen for pseudogenes was carried out by blasting male protein sequences against the genomic sequence of the female SDR and vice versa. All sequences that had been annotated as "gene" or "TE" were excluded from this latter analysis using Maskseq and RepeatMasker, respectively.

Homologous genes present in both the male and female haplotypes of the SDR were considered to be gametologs if they were detected as matches in a reciprocal BLASTP search against the SDR scaffolds (E value cutoff: 10^{-4}). The same criterion was used to identify homologs of SDR genes located outside the SDR (Table S2).

Identification of Transposons and Other Repeated Sequences in the SDR

An *Ectocarpus*-specific TE library (described in [16]), which had been compiled with REPET [66], was used to annotate SDR transposons. TEs were also annotated by running the de novo annotation software Repclass [67] with default parameters. See the Supplemental Experimental Procedures for details.

Intrahaplotype Sequence Similarity

Analyses of sequence similarity within the male and female SDR haplotypes were performed using a custom Perl code [5]. By default, the threshold for sequence identity was fixed to 97%. When the threshold was reduced to 50%, the same result was obtained.

Quantitative RT-PCR Analysis of SDR Gene Transcript Abundances during the *Ectocarpus* sp. Life Cycle

The abundance of male and female SDR gene transcripts during the *Ectocarpus* sp. life cycle was assessed by quantitative RT-PCR (qRT-PCR). Primer pairs were designed to amplify regions of the 3′UTR or the most 3′ exon of the gene to be analyzed (Table S4D). In silico virtual PCR amplifications were carried out using the electronic PCR program [68] and both the male and female genome sequences to check the specificity of oligonucleotide pairs. qRT-PCR analysis was carried out for 13 male SDR genes and 11 female SDR genes (Figures S4A and S4B). The remaining SDR genes could not be analyzed either because they had very small exons, which posed a problem for primer design, or because it was not possible to obtain a single amplification product. RNA extraction and qRT-PCR were performed as previously described [48].

Construction of Phylogenetic Trees for an SDR and an Autosomal Gene

Exon sequences from an SDR and an autosomal sequence were amplified from three *Ectocarpus* lineages, from *S. firma* (E. Gepp) Zinova and *S. lomentaria* (Lyngbye) Link, distantly related brown alga within the order Ectocarpales, and from the kelp *L. digitata* (Hudson) J.V. Lamouroux. For the SDR gene, an exon region was amplified for the gametolog pair Esi0068_0003 (male) and FeV4scaf15_1 (female). Alignable sequence data from the internal transcribed spacer 2 (ITS2) nuclear autosomal region and adjacent large subunit (LSU) were obtained for the same strains. Sequences were edited using the Codon Code sequence aligner and aligned with Muscle in the program SeaView [69]. Evolutionary history was inferred using both the Neighbor-Joining (Figures 5B and 5C) and PhyML methods implemented in MEGA5 [40], with the same topology resolved by both methods. The strains and lineages used are described in Table S1A, and the primers are described in Table S3.

Synonymous Divergence

Pairwise alignments of the deduced protein sequences of gametolog gene pairs were performed in SeaView using Muscle with default parameters. Regions with poor alignments were further analyzed with Gblocks [70]. The aligned protein sequences were then back translated to coding sequence, and dS was calculated using Codeml within the suite of programs in PAML v.4 [71].

Estimating the Age of the *Ectocarpus* sp. SDR

Coding sequence data from 65 Stramenopile species, including two diatoms, were obtained from the Hogenom database v.6 and from GenBank [72]. Homologous genes were identified using a clustering approach. Orthologous sequences were identified and checked using phylogenetic information (described in the Supplemental Experimental Procedures). Coding sequences from other Phaeophyceae species were added to the cluster data, and further data cleaning was carried out so that only orthologous sequences were retained, as described in the Supplemental Experimental Procedures. A pairwise alignment of the *Ectocarpus* sp. genes with all of the identified orthologous genes from each cluster was then carried out using Prank [73], and alignments were improved using Gblocks [70, 71]. The programs Codeml and Yn00 from PAML v.4 [71] were then run on each gene pair in order to calculate pairwise dS values. The resulting dS values were plotted against the divergence times estimated by Silberfeld et al. [74] and Brown and Sorhannus [75].

Codon Usage Analysis

A set of 27 optimal codons was identified by comparing the codon usage of highly expressed genes (ribosomal genes) with the rest of the genome using the multivariate approach described in Charif et al. [29]. Fop values were correlated with RNA-seq expression levels (Figures S2C and S2D).

Sex Determination in Strains Carrying Different Numbers of U and V Chromosomes

Polyploid gametophytes were constructed using the *ouroboros* mutant [48] (Figure S1A). Details of genetic crosses and ploidy verification are given in the Supplemental Experimental Procedures.

Accession Numbers

The GenBank accession number for the raw sequence data (Supplemental Experimental Procedures) reported in this paper is ERP002539. The SRA accession numbers for the raw sequence data (Supplemental Experimental Procedures) reported in this paper are SRX468696 and SRX468697.

Supplemental Information

Supplemental Information includes Supplemental Experimental Procedures, five figures, and sixteen tables and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2014.07.042.

Author Contributions

S.M.C., D.R., J.M.C., and G.A.B.M. designed the research study. S.A., E.P., R.L., A.F.P., S.M.D., and M.R. performed the research study. S.M.C., J.M.C., S.A., A.C., R.L., E.P., G.A.B.M., J.B., L.S., and M.V. analyzed the data. J.-M.A., E.C., and Y.V.d.P. contributed analytic and computational tools. S.M.C. coordinated the research study. S.M.C., S.A., and J.M.C. wrote the manuscript, with input from all the authors.

Acknowledgments

The authors wish to thank Thomas Broquet, Veronique Storm, and Sylvain Mousset for advice on the statistical analysis; Aurélie Kapusta for help with Repclass; Emmanuelle Lerat for explanations about TE librairies; Thomas Bigot and Florent Lassalle for help with TPMS; Catherine Leblanc, Florian Weinberger, Gareth Pearson, and Olivier de Clerk for sharing unpublished RNA-seq data; and Helen Skaletsky for help with intrachromosomal similarity analyses. This work was supported by the Centre National de la Recherche Scientifique, the Agence Nationale de la Recherche (project SEXSEAWEED), the University Pierre and Marie Curie (Emergence program), the Interreg program France (Channel) – England (project Marinexus), and the Interreg IVB (project EnAlgae).

Received: October 11, 2013
Revised: February 11, 2014
Accepted: July 15, 2014
Published: August 28, 2014

References

1. Charlesworth, D., Charlesworth, B., and Marais, G. (2005). Steps in the evolution of heteromorphic sex chromosomes. Heredity (Edinb) *95*, 118–128.
2. Jordan, C.Y., and Charlesworth, D. (2012). The potential for sexually antagonistic polymorphism in different genome regions. Evolution *66*, 505–516.
3. Ironside, J.E. (2010). No amicable divorce? challenging the notion that sexual antagonism drives sex chromosome evolution. Bioessays *32*, 718–726.
4. Lahn, B.T., and Page, D.C. (1999). Four evolutionary strata on the human X chromosome. Science *286*, 964–967.
5. Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., et al. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature *423*, 825–837.
6. Lemaitre, C., Braga, M.D., Gautier, C., Sagot, M.F., Tannier, E., and Marais, G.A. (2009). Footprints of inversions at present and past pseudoautosomal boundaries in human sex chromosomes. Genome Biol. Evol. *1*, 56–66.
7. Wang, J., Na, J.K., Yu, Q., Gschwend, A.R., Han, J., Zeng, F., Aryal, R., VanBuren, R., Murray, J.E., Zhang, W., et al. (2012). Sequencing papaya

X and Yh chromosomes reveals molecular basis of incipient sex chromosome evolution. Proc. Natl. Acad. Sci. USA *109*, 13710–13715.

8. Bull, J.J. (1983). Evolution of Sex Determining Mechanisms (Menlo Park: Benjamin/ Cummings).
9. Bull, J. (1978). Sex chromosomes in haploid dioecy: a unique contrast to Muller's theory for diploid dioecy. Am. Nat. *112*, 245–250.
10. Bachtrog, D., Kirkpatrick, M., Mank, J.E., McDaniel, S.F., Pires, J.C., Rice, W., and Valenzuela, N. (2011). Are all sex chromosomes created equal? Trends Genet. *27*, 350–357.
11. Bachtrog, D. (2011). Plant sex chromosomes: a non-degenerated Y? Curr. Biol. *21*, R685–R688.
12. Yamato, K.T., Ishizaki, K., Fujisawa, M., Okada, S., Nakayama, S., Fujishita, M., Bando, H., Yodoya, K., Hayashi, K., Bando, T., et al. (2007). Gene organization of the liverwort Y chromosome reveals distinct sex chromosome evolution in a haploid system. Proc. Natl. Acad. Sci. USA *104*, 6472–6477.
13. Ferris, P., Olson, B.J., De Hoff, P.L., Douglass, S., Casero, D., Prochnik, S., Geng, S., Rai, R., Grimwood, J., Schmutz, J., et al. (2010). Evolution of an expanded sex-determining locus in *Volvox*. Science *328*, 351–354.
14. McDaniel, S.F., Neubig, K.M., Payton, A.C., Quatrano, R.S., and Cove, D.J. (2013). Recent gene-capture on the UV sex chromosome of the moss *Ceratodon purpureus*. Evolution *67*, 2811–2822.
15. Peters, A.F., Marie, D., Scornet, D., Kloareg, B., and Cock, J.M. (2004). Proposal of *Ectocarpus siliculosus* (Ectocarpales, Phaeophyceae) as a model organism for brown algal genetics and genomics. J. Phycol. *40*, 1079–1088.
16. Cock, J.M., Sterck, L., Rouzé, P., Scornet, D., Allen, A.E., Amoutzias, G., Anthouard, V., Artiguenave, F., Aury, J.M., Badger, J.H., et al. (2010). The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. Nature *465*, 617–621.
17. Billiard, S., López-Villavicencio, M., Devier, B., Hood, M.E., Fairhead, C., and Giraud, T. (2011). Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. Biol. Rev. Camb. Philos. Soc. *86*, 421–442.
18. Hood, M.E., Petit, E., and Giraud, T. (2013). Extensive divergence between mating-type chromosomes of the anther-smut fungus. Genetics *193*, 309–315.
19. Menkis, A., Jacobson, D.J., Gustafsson, T., and Johannesson, H. (2008). The mating-type chromosome in the filamentous ascomycete *Neurospora tetrasperma* represents a model for early evolution of sex chromosomes. PLoS Genet. *4*, e1000030.
20. Berthold, G. (1881). Die geschlechtliche Fortpflanzung der eigentlichen Phaeosporeen. Mitt. Zool. Stat. Neapel *2*, 401–413.
21. van den Hoek, C., Mann, D.G., and Jahns, H.M. (1995). Algae: An Introduction to Phycology (Cambridge: Cambridge University Press).
22. Evans, L.V. (1963). A large chromosome in the laminarian nucleus. Nature *198*, 215.
23. Lewis, R.J. (1996). Chromosomes of the brown algae. Phycologia *35*, 19–40.
24. Müller, D.G. (1975). Sex expression in aneuploid gametophytes of the brown alga *Ectocarpus siliculosus* (Dillw.) Lyngb. Arch. Protistenk. *117*, 297–302.
25. Dittami, S.M., Proux, C., Rousvoal, S., Peters, A.F., Cock, J.M., Coppée, J.Y., Boyen, C., and Tonon, T. (2011). Microarray estimation of genomic inter-strain variability in the genus *Ectocarpus* (Phaeophyceae). BMC Mol. Biol. *12*, 2.
26. Heesch, S., Cho, G.Y., Peters, A.F., Le Corguillé, G., Falentin, C., Boutet, G., Coëdel, S., Jubin, C., Samson, G., Corre, E., et al. (2010). A sequence-tagged genetic map for the brown alga *Ectocarpus siliculosus* provides large-scale assembly of the genome sequence. New Phytol. *188*, 42–51.
27. Carvalho, A.B., and Clark, A.G. (2013). Efficient identification of Y chromosome sequences in the human and Drosophila genomes. Genome Res. *23*, 1894–1907.
28. Sharp, P.M., and Li, W.H. (1987). The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. *15*, 1281–1295.
29. Charif, D., Thioulouse, J., Lobry, J.R., and Perrière, G. (2005). Online synonymous codon usage analyses with the ade4 and seqinR packages. Bioinformatics *21*, 545–547.
30. Rozen, S., Skaletsky, H., Marszalek, J.D., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K., and Page, D.C. (2003). Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature *423*, 873–876.
31. Bartolomé, C., and Charlesworth, B. (2006). Evolution of amino-acid sequences and codon usage on the *Drosophila miranda* neo-sex chromosomes. Genetics *174*, 2033–2044.
32. Bachtrog, D. (2003). Adaptation shapes patterns of genome evolution on sexual and asexual chromosomes in Drosophila. Nat. Genet. *34*, 215–219.
33. Pessia, E., Popa, A., Mousset, S., Rezvoy, C., Duret, L., and Marais, G.A. (2012). Evidence for widespread GC-biased gene conversion in eukaryotes. Genome Biol. Evol. *4*, 675–682.
34. Bachtrog, D. (2013). Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat. Rev. Genet. *14*, 113–124.
35. Hill, W.G., and Robertson, A. (1966). The effect of linkage on limits to artificial selection. Genet. Res. *8*, 269–294.
36. Charlesworth, B., and Charlesworth, D. (2000). The degeneration of Y chromosomes. Philos. Trans. R. Soc. Lond. B Biol. Sci. *355*, 1563–1572.
37. Bachtrog, D. (2008). The temporal dynamics of processes underlying Y chromosome degeneration. Genetics *179*, 1513–1525.
38. Idnurm, A., Walton, F.J., Floyd, A., and Heitman, J. (2008). Identification of the sex genes in an early diverged fungus. Nature *451*, 193–196.
39. Foster, J.W., Brennan, F.E., Hampikian, G.K., Goodfellow, P.N., Sinclair, A.H., Lovell-Badge, R., Selwood, L., Renfree, M.B., Cooper, D.W., and Graves, J.A. (1992). Evolution of sex determination and the Y chromosome: SRY-related sequences in marsupials. Nature *359*, 531–533.
40. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. *28*, 2731–2739.
41. Veyrunes, F., Waters, P.D., Miethke, P., Rens, W., McMillan, D., Alsop, A.E., Grützner, F., Deakin, J.E., Whittington, C.M., Schatzkamer, K., et al. (2008). Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. Genome Res. *18*, 965–973.
42. Potrzebowski, L., Vinckenbosch, N., Marques, A.C., Chalmel, F., Jégou, B., and Kaessmann, H. (2008). Chromosomal gene movements reflect the recent origin and biology of therian sex chromosomes. PLoS Biol. *6*, e80.
43. Bergero, R., and Charlesworth, D. (2011). Preservation of the Y transcriptome in a 10-million-year-old plant sex chromosome system. Curr. Biol. *21*, 1470–1474.
44. Fry, J.D. (2010). The genomic location of sexually antagonistic variation: some cautionary comments. Evolution *64*, 1510–1516.
45. Rice, W.R. (1984). Sex chromosomes and the evolution of sexual dimorphism. Evolution *38*, 735–742.
46. Ellegren, H., and Parsch, J. (2007). The evolution of sex-biased genes and sex-biased gene expression. Nat. Rev. Genet. *8*, 689–698.
47. Assis, R., Zhou, Q., and Bachtrog, D. (2012). Sex-biased transcriptome evolution in Drosophila. Genome Biol. Evol. *4*, 1189–1200.
48. Coelho, S.M., Godfroy, O., Arun, A., Le Corguillé, G., Peters, A.F., and Cock, J.M. (2011). *OUROBOROS* is a master regulator of the gametophyte to sporophyte life cycle transition in the brown alga *Ectocarpus*. Proc. Natl. Acad. Sci. USA *108*, 11518–11523.
49. Bergero, R., and Charlesworth, D. (2009). The evolution of restricted recombination in sex chromosomes. Trends Ecol. Evol. *24*, 94–102.
50. 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.
51. Chibalina, M.V., and Filatov, D.A. (2011). Plant Y chromosome degeneration is retarded by haploid purifying selection. Curr. Biol. *21*, 1475–1479.
52. Qiu, S., Bergero, R., and Charlesworth, D. (2013). Testing for the footprint of sexually antagonistic polymorphisms in the pseudoautosomal region of a plant sex chromosome pair. Genetics *194*, 663–672.
53. Ellegren, H., and Carmichael, A. (2001). Multiple and independent cessation of recombination between avian sex chromosomes. Genetics *158*, 325–331.
54. Stöck, M., Horn, A., Grossen, C., Lindtke, D., Sermier, R., Betto-Colliard, C., Dufresnes, C., Bonjour, E., Dumas, Z., Luquet, E., et al. (2011). Ever-young sex chromosomes in European tree frogs. PLoS Biol. *9*, e1001062.
55. Vicoso, B., Kaiser, V.B., and Bachtrog, D. (2013). Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. Proc. Natl. Acad. Sci. USA *110*, 6453–6458.
56. Allen, C.E. (1935). The genetics of bryophytes. Bot. Rev. *1*, 269–291.

57. Coelho, S.M., Scornet, D., Rousvoal, S., Peters, N.T., Dartevelle, L., Peters, A.F., and Cock, J.M. (2012). How to cultivate *Ectocarpus*. Cold Spring Harb Protoc *2012*, 258–261.
58. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.
59. Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.
60. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. *7*, 562–578.
61. Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.
62. Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.
63. Sterck, L., Billiau, K., Abeel, T., Rouzé, P., and Van de Peer, Y. (2012). ORCAE: online resource for community annotation of eukaryotes. Nat. Methods *9*, 1041.
64. Foissac, S., Gouzy, J.P., Rombauts, S., Mathé, C., Amselem, J., Sterck, L., Van de Peer, Y., Rouzé, P., and Schiex, T. (2008). Genome annotation in plants and fungi: EuGene as a model platform. Current Bioinformatics *3*, 87–97.
65. Degroeve, S., Saeys, Y., De Baets, B., Rouzé, P., and Van de Peer, Y. (2005). SpliceMachine: predicting splice sites from high-dimensional local context representations. Bioinformatics *21*, 1332–1338.
66. Flutre, T., Duprat, E., Feuillet, C., and Quesneville, H. (2011). Considering transposable element diversification in de novo annotation approaches. PLoS ONE *6*, e16526.
67. Feschotte, C., Keswani, U., Ranganathan, N., Guibotsy, M.L., and Levine, D. (2009). Exploring repetitive DNA landscapes using REPCLASS, a tool that automates the classification of transposable elements in eukaryotic genomes. Genome Biol. Evol. *1*, 205–220.
68. Schuler, G.D. (1997). Sequence mapping by electronic PCR. Genome Res. *7*, 541–550.
69. Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. *27*, 221–224.
70. Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. *17*, 540–552.
71. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591.
72. Penel, S., Arigon, A.M., Dufayard, J.F., Sertier, A.S., Daubin, V., Duret, L., Gouy, M., and Perrière, G. (2009). Databases of homologous gene families for comparative genomics. BMC Bioinformatics *10* (*Suppl 6*), S3.
73. Löytynoja, A., and Goldman, N. (2005). An algorithm for progressive multiple alignment of sequences with insertions. Proc. Natl. Acad. Sci. USA *102*, 10557–10562.
74. Silberfeld, T., Leigh, J.W., Verbruggen, H., Cruaud, C., de Reviers, B., and Rousseau, F. (2010). A multi-locus time-calibrated phylogeny of the brown algae (Heterokonta, Ochrophyta, Phaeophyceae): Investigating the evolutionary nature of the "brown algal crown radiation". Mol. Phylogenet. Evol. *56*, 659–674.
75. Brown, J.W., and Sorhannus, U. (2010). A molecular genetic timescale for the diversification of autotrophic stramenopiles (Ochrophyta): substantive underestimation of putative fossil ages. PLoS ONE *5*, e12759.

Discussion et perspectives

L'article présenté a permis de démontrer que les mécanismes d'évolution au sein des chromosomes sexuels U et V sont similaires à ceux des autres systèmes XY et ZW, avec une suppression de la recombinaison au niveau des SDR, une diminution de la densité de gènes, la présence accrue de pseudogènes et l'accumulation d'éléments transposables. Dans le cas d'*Ectocarpus*, les SDR possèdent un nombre similaire de gènes. Certains de ces gènes sont homologues entre les deux SDR et sont appelés des gamétologues. Une première analyse de l'évolution moléculaire des gènes a montré une pression évolutive similaire entre les gènes de la SDR mâle et femelle. L'analyse de la datation de la SDR a permis d'estimer son âge à plus de 100 millions d'années, résultat validé par des analyses complémentaires (Lipinska et al. in press). Contrairement à d'autres espèces comme l'homme, *Ectocarpus* ne montre pas de présence de strate évolutive au niveau des deux SDR. Cependant, cette absence de la présence de strate n'induit pas obligatoirement son absence, mais notre incapacité à pouvoir la détecter. En effet, contrairement au chromosome X qui continue de recombiner lorsque ce dernier est présent chez une femelle, l'absence de recombinaison des SDR dans le système UV induit que ces deux régions ont évolué indépendamment et simultanément, rendant impossible la détection de strates évolutives.

Le séquençage des génomes d'autres algues brunes et l'identification des SDR permettraient d'étudier l'évolution de cette région génomique au sein de ce groupe. L'identification complète de chaque chromosome sexuel permettrait une analyse plus fine de la dynamique de l'évolution de la taille des SDR et des PAR, une l'analyse du contenu en gènes des SDR, leurs fonctions et les mouvements entre les génomes ou encore de l'évolution moléculaire des gènes sexe spécifiques.

L'identification et la caractérisation des deux SDR, mâle et femelle, chez *Ectocarpus* ont permis une comparaison de ces deux régions. Cependant, la SDR femelle reste relativement fragmentée avec un total de 27 scaffolds, comparée aux 3 super-contigs de la SDR mâle. De plus, la proportion de nucléotides inconnus dans la SDR femelle approche les 30% tandis qu'elle n'est que de 5% chez la SDR mâle. Des efforts ont été menés pour tenter d'apporter une amélioration avec l'assemblage de la SDR femelle par l'utilisation de données PacBio (données non publiées). Des difficultés d'extraction de l'ADN et du protocole de séquençage ont cependant limité l'utilisation de ces données, ne permettant d'obtenir qu'une couverture de 1,5x, ne permettant pas d'améliorer

l'assemblage de la SDR femelle. D'autres séquençages de type PacBio sont prévus afin de palier à ce manque de couverture et permettre d'améliorer l'assemblage de la SDR femelle.

Article 2 - The Pseudoautosomal Region of the U/V Sex Chromosomes of the Brown Alga *Ectocarpus* Exhibit Unusual Features

Introduction

Les régions recombinantes du chromosome sexuel (PAR) ont un lien particulier avec la région non recombinante (SDR). La présence de ce lien fait qu'il a été prédit que la PAR présente des caractéristiques particulières au niveau structural et fonctionnel (Otto et al. 2011). Cet article présente les différents résultats sur l'analyse de la PAR chez *Ectocarpus*.

Ma contribution à cet article a été dans un premier temps de réannoter manuellement l'ensemble des super-contig composant le chromosome sexuel (groupe de liaison 30 – LG30) pour réaliser une comparaison la plus précise possible au niveau de la structure des gènes, mais aussi le groupe de liaison 4 (LG04) qui présente une taille similaire au LG30. En effet, plusieurs études suggèrent que la taille du chromosome doit être prise en compte avant de pouvoir réaliser des analyses comparatives des structures (Burt 2002; Jensen-Seaman et al. 2004). Les données de mapping des reads de dix librairies RNA-seq (quatre librairies provenant de gamétophytes immatures mâles et femelles, deux librairies par sexe ; quatre librairies provenant de gamétophytes matures mâles et femelles, deux librairies par sexe ; et deux librairies provenant d'un parthéno-sporophyte mature), obtenu avec TopHat2, ont été intégrées au sein de Orcae afin de faciliter le travail de réannotation réalisé avec l'outil GenomeView (Abeel et al. 2012).

Une fois le processus de réannotation réalisé, différentes caractéristiques structurales ont été comparées entre la PAR, le LG04, les autosomes et les SDR afin de déterminer si la PAR présente des particularités structurales. La densité des éléments transposables, la densité de gènes, le pourcentage de GC et de GC3 des gènes, la taille des gènes, la taille des introns et des CDS, et le nombre d'exons ont été comparés.

Une autre partie de ma contribution a été l'analyse de l'expression des gènes durant différentes phases du cycle de vie, et a été réalisée à l'aide de données RNA-seq, comprenant quatre librairies provenant de gamétophytes immatures mâles et femelles (deux librairies par sexe) ; quatre librairies provenant de gamétophytes matures mâles et femelles (deux librairies par sexe) ; et six librairies provenant de différents tissus du parthéno-sporophyte, afin de déterminer les profils d'expression des gènes dans la PAR lors de ces différents stades par rapport aux gènes autosomiques.

Les résultats de ces travaux ont été intégrés avec d'autres analyses dans l'article suivant, publié dans Molecular Biology and Evolution en août 2015.

Article

The Pseudoautosomal Region of the U/V Sex Chromosome of the Brown Alga *Ectocarpus* Exhibit Unusual Features

Auteurs : Luthringer R[1§], Lipinska A[1§], Roze D[2], Cormier A[1], Macaisne N[1], Peters AF[3], Cock JM[1], Coelho SM[1*]

Affiliations :

[1] Sorbonne Université, UPMC Univ Paris 06, CNRS, Algal Genetics Group, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, Roscoff, France;

[2] UMI 3614, Evolutionary Biology and Ecology of Algae, CNRS, Sorbonne Universités, UPMC, PUCCh, UACH, Station Biologique de Roscoff, Roscoff, France;

[3] Bezhin Rosko, Santec, France

§égale contribution

Correspondance : coelho@sb-roscoff.fr

Article publié dans Molecular Biology and Evolution

doi : 10.1093/molbev/msv173

MBE Advance Access published August 28, 2015

The Pseudoautosomal Regions of the U/V Sex Chromosomes of the Brown Alga *Ectocarpus* Exhibit Unusual Features

Rémy Luthringer,[†,1] Agnieszka P. Lipinska,[†,1] Denis Roze,[2] Alexandre Cormier,[1] Nicolas Macaisne,[1] Akira F. Peters,[3] J. Mark Cock,[1] and Susana M. Coelho*[,1]

[1]Sorbonne Université, UPMC Univ Paris 06, CNRS, Algal Genetics Group, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, Roscoff, France

[2]UMI 3614, Evolutionary Biology and Ecology of Algae, CNRS, Sorbonne Universités, UPMC, PUCCh, UACH, Station Biologique de Roscoff, Roscoff, France

[3]Bezhin Rosko, Santec, France

[†]These authors contributed equally to this work.

***Corresponding author:** E-mail: coelho@sb-roscoff.fr.

Associate editor: Stephen Wright

Abstract

The recombining regions of sex chromosomes (pseudoautosomal regions, PARs) are predicted to exhibit unusual features due to their being genetically linked to the nonrecombining, sex-determining region. This phenomenon is expected to occur in both diploid (XY, ZW) and haploid (UV) sexual systems, with slightly different consequences for UV sexual systems because of the absence of masking during the haploid phase (when sex is expressed) and because there is no homozygous sex in these systems. Despite a considerable amount of theoretical work on PAR genetics and evolution, these genomic regions have remained poorly characterized empirically. We show here that although the PARs of the U/V sex chromosomes of the brown alga *Ectocarpus* recombine at a similar rate to autosomal regions of the genome, they exhibit many genomic features typical of nonrecombining regions. The PARs were enriched in clusters of genes that are preferentially, and often exclusively, expressed during the sporophyte generation of the life cycle, and many of these genes appear to have evolved since the Ectocarpales diverged from other brown algal lineages. A modeling-based approach was used to investigate possible evolutionary mechanisms underlying this enrichment in sporophyte-biased genes. Our results are consistent with the evolution of the PAR in haploid systems being influenced by differential selection pressures in males and females acting on alleles that are advantageous during the sporophyte generation of the life cycle.

Key words: **pseudoautosomal region, sex chromosomes, UV sexual system, brown algae.**

Introduction

Sex chromosomes have commonly been found to possess strikingly distinctive features compared with autosomes, for example in terms of the content and density of genes and repeat sequences. These characteristics are thought to be a consequence of suppression of recombination between the sex chromosomes (X and Y or Z and W in diploid systems, or U and V in haploid systems; reviewed in Otto et al. [2011]). A broadly established model of sex chromosome evolution predicts gradual expansion of the region of suppressed recombination, driven by selection for linkage between the sex-determining region (SDR) and loci at which selection differs between males and females (Charlesworth et al. 2005; Immler and Otto 2015). Expansion of the SDR reduces the recombining portion of the sex chromosome, the so-called pseudoautosomal region (PAR). However, the recombining region is usually not lost completely and it is thought that most species retain a PAR because homologous recombination in this region plays a critical role in chromosomal pairing and segregation during meiosis (Rouyer et al. 1986; Shi et al. 2001). Moreover, there are situations where sexually antagonistic (SA) forces may be too weak to drive a marked expansion of the SDR, and an extensive PAR may be preserved. This may be expected to occur, for example, in organisms with a low level of phenotypic sexual dimorphism (e.g., Ahmed et al. 2014) or where SA selection has been resolved by alternative processes such as the evolution of sex-biased gene expression (Vicoso et al. 2013).

The evolutionary fate of PAR genes is expected to differ from that of either autosomal or fully sex-linked genes. In particular, sex differences in allele frequencies should be maintained more easily in the PAR, either due to SA polymorphisms (which are maintained under a wider range of conditions than on autosomes), or to other forms of selection, such as heterozygous advantage (Otto et al. 2011). These effects are expected to be strongest very near the SDR, and to decay as the genetic distance from the SDR increases (the rate of decay being inversely proportional to the strength of selection maintaining polymorphism; Charlesworth et al. 2014; Kirkpatrick and Guerrero 2014).

There has been little empirical work on PARs. Analyzes of the structure and genetic behavior of the PAR have mainly focused on organisms that have old, well-differentiated sex chromosomes such as humans and other mammals (Flaquer

et al. 2008; Raudsepp et al. 2012), and more recently birds (Smeds et al. 2014). These PARs have been shown to exhibit several unusual features compared with autosomes, including higher levels of recombination (Soriano et al. 1987; Lien et al. 2000; Kondo et al. 2001), greater abundance of repetitive DNA (Hinch et al. 2014; Smeds et al. 2014), and differences in GC content (Montoya-Burgos et al. 2003). These studies focused on organisms with diploid sexual chromosome systems (XY and ZW), whereas in a large number of taxa including many red, brown, and green algae, land plants and fungi, sex is determined during the haploid phase of the life cycle (UV systems; Bachtrog et al. 2011). Many of the theoretical predictions made for diploid sexual systems are also relevant to UV sexual systems, for example concerning the evolution of recombination suppression and the maintenance of sex differences in allele frequencies in the PAR (Immler and Otto 2015). Some effects, such as the potential of sex differences in selection to drive gene differentiation in the PAR, are expected to be stronger in UV systems because the U and V chromosomes occur only in females and males, respectively (in contrast with the X, for example, which can occur in males and females). At present however, few empirical data are available for haploid sexual systems to test these various predictions.

We have recently shown that the UV sex chromosomes of the brown alga *Ectocarpus* have a small nonrecombining SDR, despite being at least 70 My old, and that this region is bordered by two relatively large PARs (Ahmed et al. 2014). Here, we show that the PARs of these chromosomes recombine at a similar rate to autosomal regions of the genome and yet exhibit many features typical of nonrecombining regions. The PARs are enriched in physically linked clusters of genes that are preferentially, and often exclusively, expressed during the sporophyte generation of the life cycle and many of these genes appear to have evolved since the Ectocarpales diverged from other brown algal lineages. A model is presented that provides a possible mechanism for the accumulation of these sporophyte-biased genes on the PARs.

Results

The PAR of the *Ectocarpus* Sex Chromosome Exhibits Unusual Structural Features

The PARs of the *Ectocarpus* sex chromosome (linkage group 30, LG30) represent about 2 Mb of sequence on each side of the 1 Mb SDR. We have previously noted that the PARs exhibit a number of structural differences compared with the autosomes. For instance, values for gene density, mean intron length, and percentage of GC content are intermediate between those of the autosomes and the SDR (Ahmed et al. 2014).

Several studies (Burt 2002; Jensen-Seaman et al. 2004) suggest that chromosome size should be taken into account when comparative analyzes of chromosome structure are carried out. In *Ectocarpus*, transposable element (TE) content tends to be negatively correlated with linkage group physical size (Spearman's correlation test $\rho = -0.113$, $P = 0.598$) whereas gene density and GC percentage increase with chromosome size (Spearman's correlation test $\rho = 0.303$, $P = 0.151$ and $\rho = 0.284$, $P = 0.178$, respectively). Consequently, to analyze in detail the unusual structural features of the *Ectocarpus* PARs, we compared the sex chromosome not only to the autosomal regions as a whole (all chromosomes apart from the sex chromosome) but also with one specific chromosome, linkage group 4 (LG04), which is of similar size (5.028 Mb) to the sex chromosome. For this comparison, all genes on LG30 and LG04 were manually curated to produce high-quality annotations for both chromosomes. Comparison of these two genomic regions showed that the PARs contained more TE sequences and lower gene density than LG04, and that GC content and the size of coding regions were significantly lower for the PAR, compared with LG04 (fig. 1*A*–*D*). Moreover, PAR genes tended to have longer introns, and fewer and smaller exons on average than genes on LG04 (fig. 1*E*–*H*). All of these differences were also detected when the PARs were compared with the autosomes (fig. 1*A*–*D*; supplementary fig. S1, Supplementary Material online), confirming that the PARs are unusual. Moreover, the features that distinguish the PARs from the autosomes were not confined to the regions that were close to the SDR. The PARs exhibited some structural heterogeneity along their length, with, for example, a significant negative correlation between TE content and gene content (Pearson's correlation test $r = -0.606$, $P < 0.01$), but we found no evidence that the features that distinguish the PARs from the autosomes (gene structure, GC content, etc.) were more marked in the vicinity of the SDR (supplementary table S1, Supplementary Material online). These unusual structural features are therefore characteristic of the entire PARs.

Recombination along the Sex Chromosome

The structural analysis described above strongly indicated that the *Ectocarpus* PARs exhibit features resembling those of the nonrecombining SDR. Recombination is completely suppressed within the SDR of the *Ectocarpus* sex chromosome (Ahmed et al. 2014) but analysis of molecular marker segregation has confirmed that the PARs recombine during meiosis (Heesch et al. 2010). In order to build a more comprehensive recombination map of the *Ectocarpus* sex chromosome, an expanded segregating population of 280 individuals was genotyped with 23 LG30 markers. The average recombination rate for the PARs was 40 cM/Mb whereas the average recombination rate for autosomes was 23 cM/Mb. Comparisons of average rates between adjacent markers indicated that this difference was not significant (Mann–Whitney *U* test, $P = 0.28$). However, recombination events were unevenly distributed along the sex chromosome (fig. 2). Specifically, two regions of high recombination (one of them recombining at about ten times the genome average) were found on each side of the SDR. Recombination between markers within these peaks was significantly higher than the background recombination rate on the sex chromosome (Mann–Whitney *U* test, $P = 0.0038$). When markers within these recombination peaks were excluded, the PARs had an

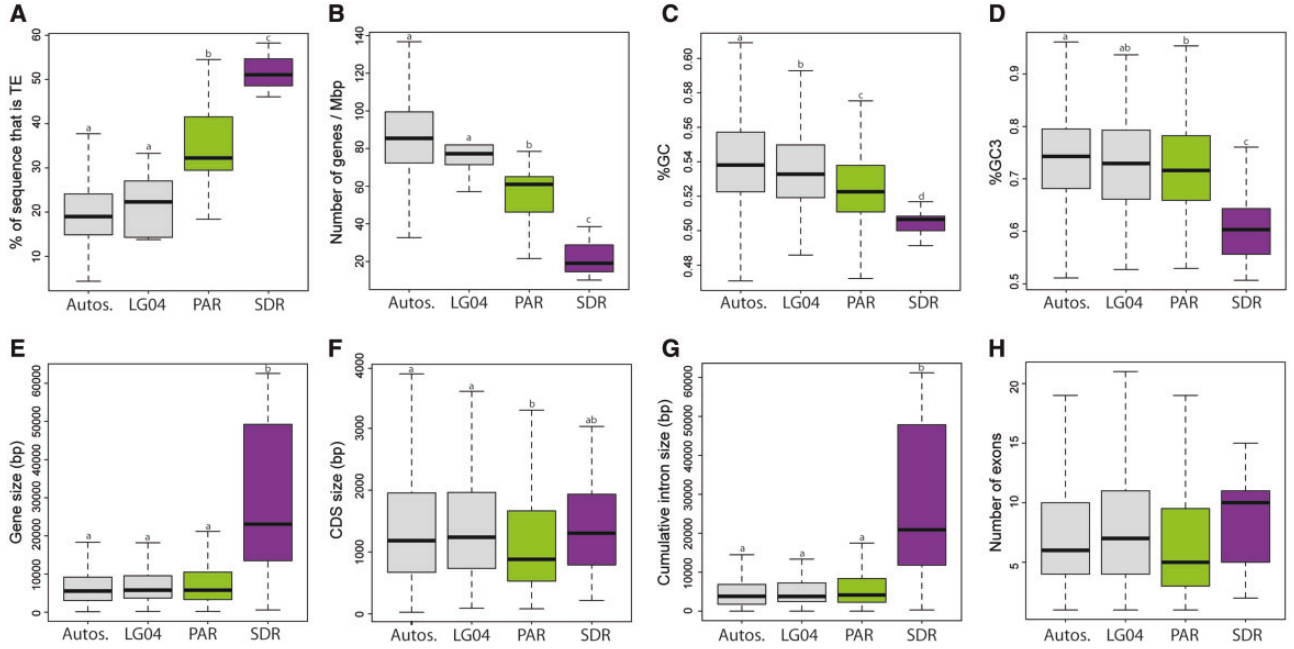


Fig. 1. Structural characteristics of the PAR compared with the SDR, LG04, and autosomes. (*A*) Percentage of TE calculated per supercontig; (*B*) gene density per supercontig; (*C*) percentage of GC per gene; (*D*) percentage of GC3 per coding sequence (CDS); (*E*) gene size; (*F*) CDS size per gene; (*G*) total intron length per gene; (*H*) number of exons per gene. Statistical differences were tested using pairwise Mann–Whitney *U* test. Letters shared in common between the groups indicate no significant difference.

average recombination rate of 15.3 cM/Mb, which was still not significantly different from the genome average (Mann–Whitney *U* test, $P = 0.388$). Globally, we found no significant correlation between recombination rate and TE or gene content (Pearson's correlation test, $P > 0.05$) along the length of the PARs, although there was a tendency for regions that exhibited higher recombination rates to have higher gene density and lower TE density (fig. 2).

Genetic recombination rates along the PARs were also studied in a segregating family generated from two parental strains of another *Ectocarpus* species, *Ectocarpus siliculosus* lineage 1a (Stache-Crain et al. 1997), demonstrating that the PARs are also a recombining region in this sister species (supplementary fig S2 and table S2, Supplementary Material online).

Expression Patterns of PAR Genes during the *Ectocarpus* Life Cycle

The PARs contain 209 protein-coding genes. We investigated their patterns of expression, using RNA-Seq, at several stages of the life cycle of *Ectocarpus*, including male and female immature and fertile gametophytes, and different tissues of the sporophyte generation. The PAR genes exhibited significantly lower mean expression levels than genes on LG04 (median 5.88 RPKM (reads per kilo base pair per million) for the PARs compared with 11.16 RPKM for LG04; Mann–Whitney *U* test, $P = 4.50 \times 10^{-10}$) and than autosomal genes in general (median 9.88 RPKM for all autosomes; Mann–Whitney *U* test, $P < 1.10 \times 10^{-7}$) (fig. 3*A*). This difference in transcript abundance was particularly marked during the gametophyte generation, and slightly less marked during the sporophyte generation.

A heatmap representing the expression levels of the PAR genes revealed a striking pattern (fig. 3*B*; supplementary fig. S3*A*, Supplementary Material online). Several clusters of genes had coordinated patterns of expression during the life cycle, including two clusters of PAR genes that were strongly upregulated during the sporophyte generation, and a cluster of genes that exhibited transcription below the detection limit (RPKM < 1), during both the gametophyte and the sporophyte generations. The sporophyte-biased gene clusters were localized in regions of the PAR that exhibited low levels of recombination (in supercontigs sctg_96 and sctg_266, fig. 2; supplementary table S3, Supplementary Material online). No other linkage group exhibited similar patterns of generation-biased gene clusters (supplementary fig. S3*B*, Supplementary Material online).

To further analyze the relationship between genomic location and life cycle expression pattern, we carried out a genome-wide analysis to identify genes that were differentially expressed during the alternation between the sporophyte and gametophyte generations of the life cycle. About 25% of the genes in the *Ectocarpus* genome were significantly differentially regulated between the generations (fold change [FC] ≥ 2, false discovery rate [FDR] < 0.1), with slightly fewer sporophyte-biased genes (~12% of the genome, 1,883 genes) than gametophyte-specific genes (~13%, 2,083 genes). The PAR was found to be significantly enriched in genes that are upregulated during the sporophyte generation (χ^2 test, $P_{adj} = 2.2 \times 10^{-7}$, Bonferroni correction) (fig. 3*C*), while none of the autosomes exhibited a significant enrichment in sporophyte-biased genes (supplementary figs. S3*B* and S4*C*, Supplementary Material online).

To examine the relationship between level of expression and degree of generation-bias, the sporophyte-biased genes

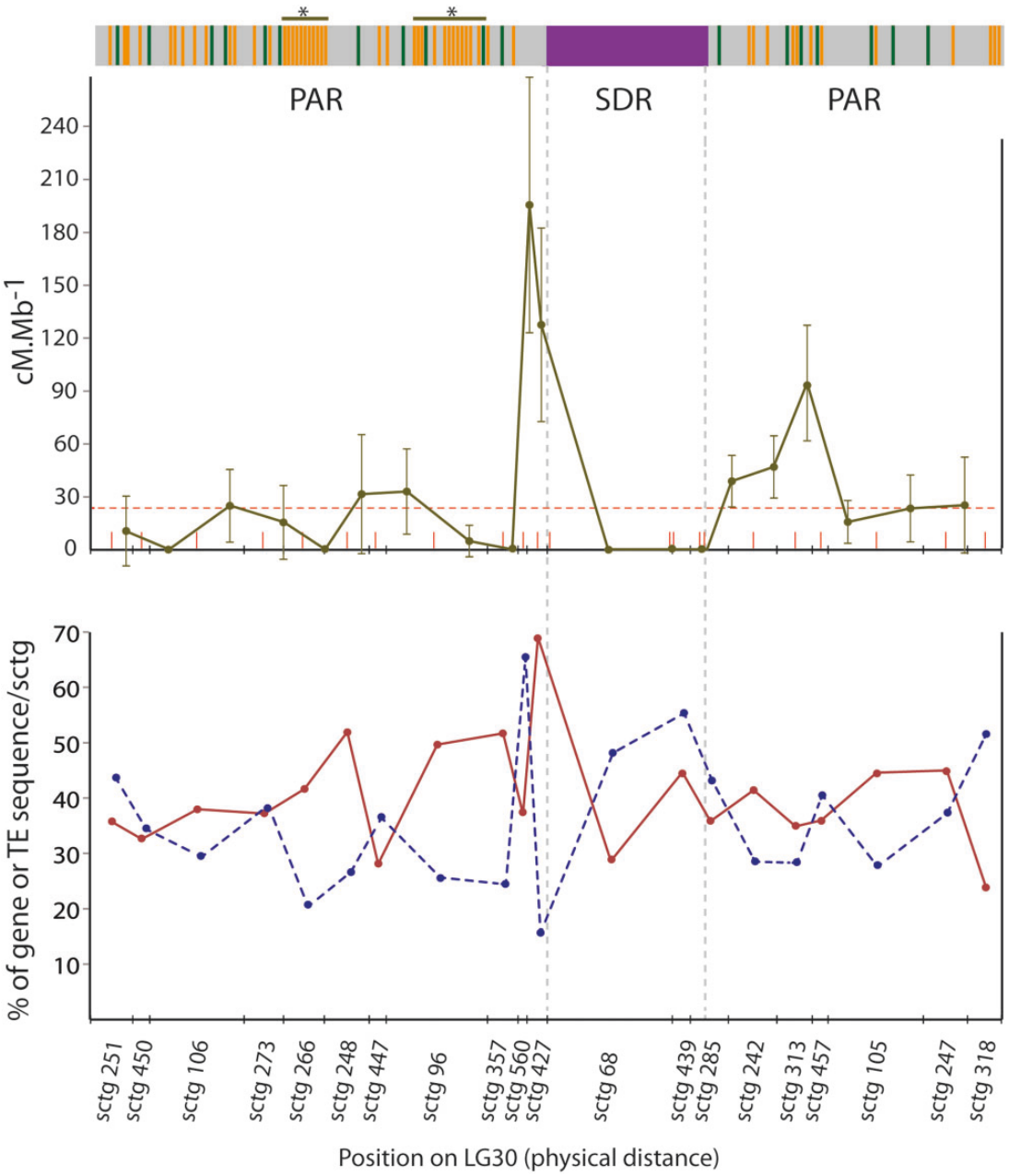


Fig. 2. Recombination frequency and distribution of TEs and gene density in the sex chromosome of *Ectocarpus*. The *x* axis indicates the physical position along the sex chromosome. Upper panel: *y* axis indicates the recombination rate (cM/Mb). Error bars represent 95% confidence intervals (CI). The recombination frequency around the SDR is unusually high. The average recombination between two adjacent markers on the PAR is 40.3 cM/Mb (18.8–66.9, 95% CI), compared with 217.2 and 95.0 cM/Mb for the peaks at the borders of the SDR. The red dashed line represents the average recombination frequency over the entire *Ectocarpus* genome. The black and red lines on the *x* axis indicate boundaries between supercontigs (sctgs) and the midpoints of supercontigs, respectively. Gray background rectangle above the upper graph indicates the distribution of generation-biased genes along the sex chromosome. Orange: sporophyte-biased genes; green: gametophyte biased genes. Horizontal bars and asterisks represent clusters of sporophyte-biased genes. See also figure 3. Gene and TE density along the *Ectocarpus* sex chromosome on the lower panel are represented by the solid red and dashed blue lines, respectively. Analysis of gene and TE density was performed by calculating the percentage of bases on each supercontig that are part of a gene or a TE, respectively. Vertical gray dashed lines indicate the boundaries between the PARs and the SDR. Note that the existing genetic map only allowed 70% of the genome sequence to be assigned to linkage groups (Heesch et al. 2010) and therefore we cannot exclude the possibility that missing scaffolds have led to an underestimation of the Mb/cM ratio in some regions of the sex chromosome.

on the PARs, on LG04, and on all autosomes were grouped according to fold-change in transcript abundance between the sporophyte and gametophyte generations, and the mean expression level (RPKM) of each group was plotted (fig. 3*D*). For LG04, and for autosomal genes in general, the degree of sporophyte-biased expression was determined by the level of expression in the gametophyte, so that their high fold difference correlated with low gametophyte expression. In contrast, all the sporophyte-biased genes on the PAR exhibited very low levels of expression in the gametophyte generation

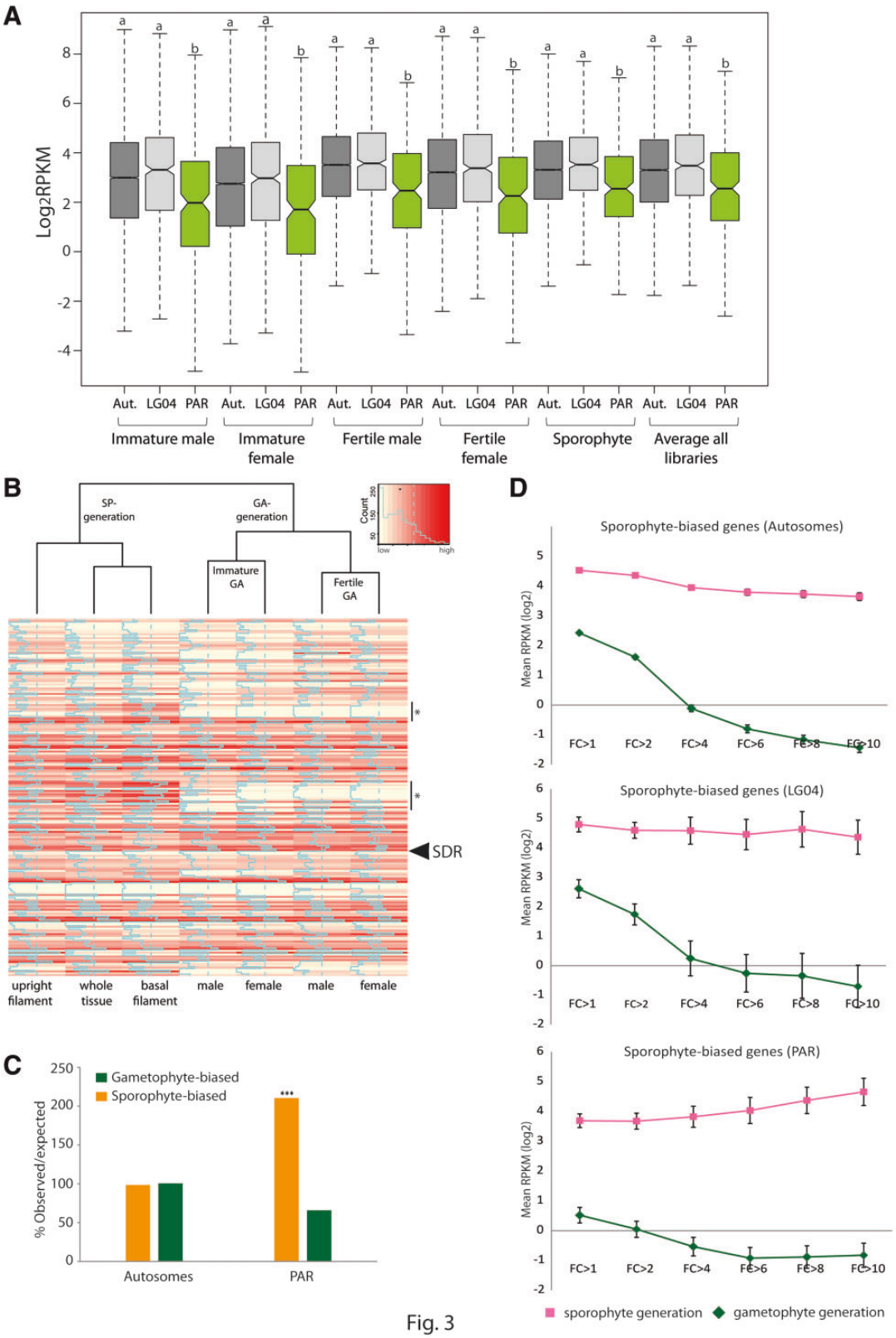


Fig. 3

Fig. 3. PAR gene expression during different life cycle stages. (*A*) Average gene expression (log_2RPKM) of all autosomes, LG04 (a linkage group of similar size to the sex chromosome) and PAR genes in male and female gametophytes (immature and fertile), and sporophytes. Letters shared between groups indicate no significant difference (Mann–Whitney *U* test, $P < 6.0 \times 10^{-5}$). (*B*) Heatmap showing the expression levels of PAR genes during different life cycle stages relative to their position on the sex chromosome (the SDR is excluded). Clusters of sporophyte-biased genes (also represented in fig. 2) are highlighted by asterisks. GA: gametophyte; SP: sporophyte (*C*) Enrichment of sporophyte-biased genes on the PAR compared with autosomes (χ^2 test with Bonferroni correction, $^{***}P_{adj} = 6.03 \times 10^{-5}$). (*D*) Expression of sporophyte-biased genes on autosomes, LG04 and PAR measured during the sporophyte (pink) and gametophyte (green) generations. Mean gene expression levels (log_2RPKM) at several degrees of generation-bias (from FC $>$ 1 to FC $>$ 10) are shown. Error bars represent standard errors of the mean.

and the degree of sporophyte-biased expression (fold change) was determined both by attenuation of expression during the gametophyte generation and by the strength of expression during the sporophyte generation.

Two types of measurement can be used to describe the expression of a gene in a multicellular organism: the level of gene expression in terms of the number of transcripts present in a particular tissue, and the breadth of expression (τ), which measures how often the gene is expressed through the life cycle and/or in how many different tissues it is transcribed (Lipinska et al. 2015).

We calculated the breadth of expression of *Ectocarpus* genes using gene expression data collected for two types of tissues and at different stages of the life cycle. Globally, PAR genes exhibited greater expression specificity than either LG04 genes or autosomal genes in general (Mann–Whitney *U* test, $P < 0.003$) (supplementary fig. S5*A*, Supplementary Material online). Sporophyte-biased PAR genes had τ values that were significantly higher than those of unbiased PAR or autosomal genes (Mann–Whitney *U* test, $P < 6.5 \times 10^{-5}$) (supplementary fig. S5*B*, Supplementary Material online).

Fifty-one sporophyte-biased and 18 gametophyte-biased genes were identified on the PARs (supplementary table S3, Supplementary Material online). A significant proportion (~50%) of the PAR sporophyte-biased genes were located in the two life cycle gene clusters mentioned above. In these clusters, 9 (sctg_266) and 13 out of 19 (sctg_96) contiguous genes exhibited sporophyte-specific expression (fig. 2; supplementary fig. S4*A*, Supplementary Material online). Clustering analysis confirmed that the distribution of sporophyte-genes on the PAR was not random (Runs test, $P < 2.2 \times 10^{-16}$). The sporophyte-biased genes in the two clusters included a duplicated pair of adjacent genes for which there was one copy in each cluster (supplementary table S3, Supplementary Material online). The regions corresponding to the clusters, which are not closely linked to the SDR (fig. 2), exhibit lower recombination rates (on average 9 cM/Mb) than the average PAR rate. However, genes located on the clusters did not exhibit different characteristics from other sporophyte-biased genes located outside the clusters and did not differ from unbiased PAR genes (supplementary fig. S4*B*, Supplementary Material online). The remaining sporophyte-biased genes were distributed along the PAR in triplets (1), pairs (5), or individually (16) (supplementary fig. S4*A*, Supplementary Material online). Neither functional domains nor orthologues in public databases were detected for most of these genes and it was therefore not possible to identify any enrichment with respect to function. However, possible roles in protein–protein interactions (leucine rich repeats, tetratricopeptide repeats, or ankyrin repeats motifs) were predicted for 7 of the 51 sporophyte-biased PAR genes.

Fewer than 12% of the genes in the *Ectocarpus* genome (i.e., 1,947 genes) exhibits sex-biased gene expression (Ahmed et al. 2014), including 31 that are located in the PAR (supplementary fig. S4*A* and table S3, Supplementary Material online). This latter set of genes did not display any unusual structural characteristics compared with unbiased PAR genes (supplementary fig. S4*B*, Supplementary Material online). There was also no significant tendency for generation-biased genes on the PAR to be also sex-biased (χ^2 test, $P = 0.25$). Nonetheless, 12 of the 69 generation-biased on the PAR exhibited both generation- and sex-bias and there was a marked correlation between the precise type of life cycle generation-bias and the type of sex-bias: all seven of the genes that were both gametophyte-biased and sex-biased were male-biased, whereas four out of five of the genes that were both sporophyte-biased genes and sex-biased were female-biased (supplementary table S3, Supplementary Material online).

The *Ectocarpus* PAR Is Enriched in Young Genes

Recently evolved genes (referred to as "orphan" genes) tend to exhibit similar features to those that we observed for the PAR genes, including shorter coding regions, fewer exons, lower expression, and weaker codon bias compared with older genes (Arendsee et al. 2014; Palmieri et al. 2014). We therefore investigated whether gene age might be one of the factors underlying the unusual features of the PAR. Complete genome resources are currently insufficient to identify orphan genes, which are defined as having evolved within a species or group of species, in *Ectocarpus* but we were able to distinguish "young genes" from "old genes" by carrying out BLASTp comparisons with other complete stramenopile genomes, including the recently published *Saccharina japonica* genome (Ye et al. 2015), and sequences in the public databases. Young genes were defined as having no BLASTp match (10^{-4} *E* value cutoff) with any of these other genomes (indicating that they are likely to have evolved since the split from the most recent common ancestor, about 100 Ma; Silberfeld et al. 2010). The PAR was significantly enriched in young genes compared with all the autosomal linkage groups (34% vs. 10%, χ^2 test with Bonferroni correction, $P = 1.5 \times 10^{-14}$). On average, young genes tended to be smaller and to have higher tissue specificity than old genes and their coding regions were smaller with lower codon adaptation index (CAI) and GC3 (Mann–Whitney *U* test, supplementary table S4, Supplementary Material online). When gene age was factored out by comparing only old genes or only young genes between the PAR and the autosomes, the PAR genes still exhibited higher percentage TE, lower GC content, longer gene size, shorter coding regions (significant for old genes only), shorter exons (significant for old genes only), and longer introns (Mann–Whitney *U* test, supplementary table S4, Supplementary Material online). Taken together, these analyzes indicated that the unusual features of the PAR genes could be explained in part by enrichment in young genes. However, when age is corrected for, PAR genes still exhibit markedly different features to autosomal genes. Interestingly, the proportion of young genes that showed generation-bias expression patterns was higher on the PAR than on the autosomes (52% vs. 28%, χ^2 test, $P = 4.18 \times 10^{-7}$).

Evolution of the PAR Genes

The rate and pattern of evolution of *Ectocarpus* genes was analyzed by comparing sequences from the reference strain (*Ectocarpus* sp. lineage 1c Peru) with orthologous sequences

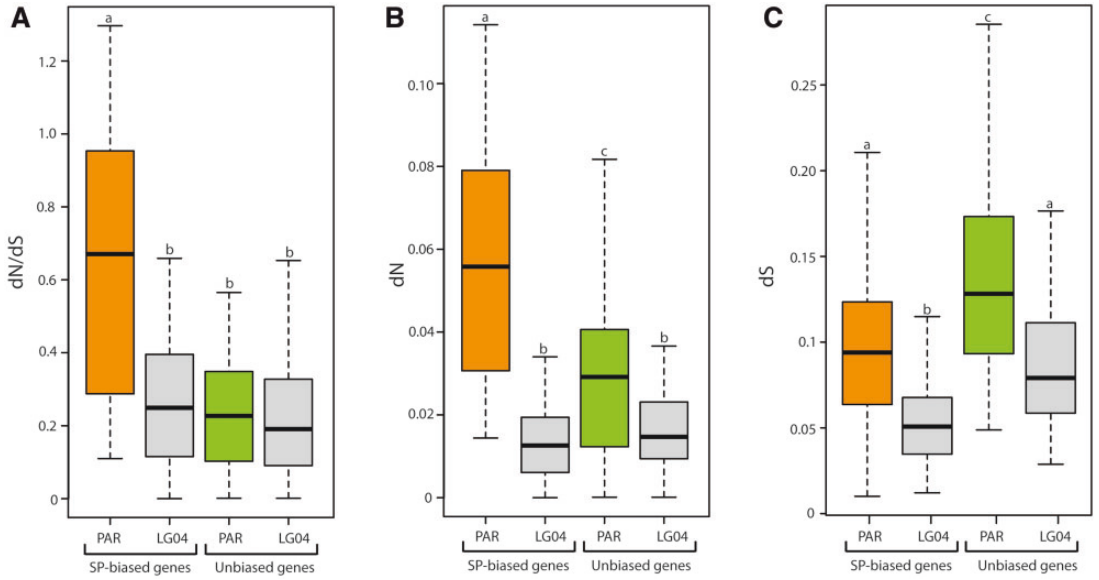


Fig. 4. Rates of evolution of PAR (generation-biased and unbiased) genes compared with autosomal genes (LG04). Pairwise dN, dS, and dN/dS ratios were calculated by comparing orthologous gene sequences from *Ectocarpus* sp. (lineage 1c Peru) and *Ectocarpus siliculosus* (lineage 1a). (*A*) Ratio of nonsynonymous to synonymous substitutions (dN/dS). (*B*) Nonsynonymous substitutions (dN). (*C*) synonymous substitutions (dS). Letters shared between groups indicate no significant difference (Mann–Whitney *U* test, $P < 0.01$).

from another *Ectocarpus* species (*E. siliculosus* lineage 1a). Compared with a set of 88 genes from LG04, the 84 PAR genes that were analyzed displayed, on average, significantly elevated values for nonsynonymous to synonymous substitution ratios (dN/dS) (Mann–Whitney *U* test, $P < 0.001$). However, when the generation-biased genes (40 genes) were removed from the data set, no significant difference in mean dN/dS ratios was detected between the PAR and autosomal gene sets (fig. 4*A*). Moreover, the sporophyte-biased PAR genes showed dN/dS ratios that were significantly higher than sporophyte-biased genes on LG04 (Mann–Whitney *U* test, $P = 2.268 \times 10^{-5}$), indicating that the increased evolutionary rates were related to the fact that these generation-biased genes were located on the PAR. The faster rate of evolution of the sporophyte-biased PAR genes was due to an increase in the rate of nonsynonymous substitutions (dN) and not to a decrease in the rate of synonymous substitution (dS) (fig. 4*B* and *C*) (Mann–Whitney *U* test, $P < 0.01$). Finally, note that although the average dN/dS ratio for unbiased PAR genes was similar to that of the autosomal gene set, the average values for both dN and dS were significantly greater than for the autosomal genes (Mann–Whitney *U* test, $P < 0.01$).

Interestingly, there was a weak, negative correlation between expression breadth and dN/dS for the PAR genes (Spearman's $\rho = 0.206$, $P = 0.0526$). In other words, PAR genes with higher dN/dS tended to exhibit a lower breadth of expression.

Of the 40 sporophyte-biased PAR genes analyzed, 24 had dN/dS ratios that were greater than 0.5, which could be an indication of adaptive evolution (Swanson et al. 2004). To perform a maximum likelihood analysis of positive selection (PAML), we searched for orthologues of the sporophyte-biased genes using transcriptome data for two additional *Ectocarpus* species (*Ectocarpus fasciculatus* lineage 5b and *Ectocarpus* sp. lineage 1c Greenland; supplementary table S5, Supplementary Material online). Complete sets of four orthologues from the four species were obtained for only seven of the sporophyte-biased PAR genes and the PAML analysis was therefore carried out using these sets. For one of these comparisons both pairs of models (M1a–M2a, M7–M8) suggested positive selection (Esi0096_0082, $\omega = 0.86$, $P < 0.05$).

Codon-usage bias has been observed in almost all genomes and is thought to result from selection for efficient and accurate translation of highly expressed genes (Kanaya et al. 2001). Optimal codons have been described for *Ectocarpus* (Ahmed et al. 2014) and a weak but significant correlation was noted between codon usage bias and gene expression levels (Wu et al. 2013). In accordance with these findings, the genes on the PARs, which were expressed at a lower level, on average, than autosomal genes (fig. 3*A*; Mann–Whitney *U* test, $P = 6.06 \times 10^{-5}$), showed significantly lower frequency of optimal codons (CAI) compared with autosomal genes (Mann–Whitney *U* test, $P = 2.0 \times 10^{-5}$). Interestingly, we found that genes in the regions close to the SDR tended to have higher CAI than more distal genes, although the significance is borderline (Spearman's $\rho = -0.15$, $P = 0.028$).

However, when codon usage analysis was carried out specifically for the groups of sporophyte-biased and unbiased genes, the CAIs were significantly lower only for the sporophyte-biased genes on the PAR, compared with all other genes (Mann–Whitney *U* test , $P < 0.004$) (supplementary fig S6, Supplementary Material online). Analysis of the *Drosophila* genome identified a positive correlation between codon bias and recombination rate (Haddrill et al. 2007;

Campos et al. 2012). *Ectocarpus* PAR genes located in regions with low recombination rates had significantly lower CAI (Mann–Whitney *U* test, $P = 0.01879$), but we found no significant difference in CAI for sporophyte-biased genes located in PAR with low versus average-to-high recombination rates. Therefore, the local recombination rate does not explain the low codon usage bias of the sporophyte-biased PAR genes (supplementary fig. S7, Supplementary Material online).

A Model for the Spread of Generation-Biased Alleles Located in the PAR

In XY or ZW systems, it has been argued that the excess of sex-biased genes often observed on X (or Z) chromosomes may result from SA selection (e.g., Vicoso and Charlesworth 2006). For example in XY systems, alleles with recessive or partially recessive effects that increase male fitness at a cost to female fitness are expected to spread more easily on the X than on autosomes; modifiers that decrease the expression of these genes in females may then spread, leading to an excess of male-biased genes on the X. We developed a theoretical model to explore whether a similar scenario (involving generation-antagonistic rather than SA selection) could explain the excess of sporophyte-biased genes observed on the PARs. This would imply that alleles increasing the fitness of sporophytes but with a fitness cost to gametophytes would spread more easily in the PAR than on autosomes.

The model (detailed in the supplementary material S1, Supplementary Material online) considers a selected locus located at a recombination distance r from the SDR of a UV sex-determination system, at which two alleles (denoted *A* and *a*) have different effects on the fitness of sporophytes, female gametophytes and male gametophytes. The different events of the life cycle are diploid selection, meiosis (recombination), haploid selection (within each sex), and fertilization (random union of gametes); the fitnesses of the different genotypes are given in table 1 (note that the results only depend on relative fitnesses within each ploidy phase and sex, as we assume that selection takes place independently among females, males, and sporophytes). The model is similar to that recently proposed by Immler and Otto (2015) but, while these authors explored conditions under which selection favors decreased recombination between a PAR locus and the SDR, we focus on the conditions for the spread of a rare allele (say allele *a*) at the selected locus, as a function of r, and the fitness effect of the allele on sporophytes (s_d) and on female (s_f) and male (s_m) gametophytes. We focus on generation-antagonistic alleles (s_d and $s_h = (s_f + s_m)/2$ have opposite signs), because the spread of such alleles may result in an increase in the frequency of genes that are differentially expressed in sporophytes and gametophytes (generation-biased genes) (table 1).

Overall, our analysis (fig. 5) shows that genomic localization has little effect on the spread of alleles when selection is similar in both sexes ($s_f \approx s_m$). However, when selection differs between the sexes (and in particular when the gametophyte-deleterious allele is neutral or slightly beneficial in one of the sexes), the model indicates that a sporophyte-beneficial allele benefits from sex-linkage, as this allows the allele to avoid being inherited by the sex where it is disfavored. Linkage to the SDR is also predicted to benefit the gametophyte-beneficial allele, but to a lesser extent since this allele still suffers from a fitness cost in the sporophytic generation. This can be seen on figures 5*B* and *C*: reducing the recombination rate between the selected locus and the sex-determining locus (from to solid curves for $r = 0.5$ to the dotted curves for $r = 0.01$) increases the parameter regions where alleles increase in frequency when rare, this effect being more marked for the sporophyte-beneficial allele (blue curves) than for the gametophyte-beneficial allele (red curves; note that the scale of the x axis is logarithmic). Therefore, taking into account the possibility of sex-differences in selection, being in the PAR expands the range of parameters allowing generation-antagonistic mutations to spread, but more so for sporophyte-beneficial, gametophyte-deleterious alleles than for gametophyte-beneficial, sporophyte-deleterious alleles. Again, this effect is generated by the fraction of generation-antagonistic mutations that is differentially selected in males and females. This model could thus explain the observed excess of sporophyte-biased gene expression in the PAR, assuming that reduced expression in gametophytes would have evolved secondarily to prevent the expression of alleles that are deleterious in at least one sex (note that complete linkage to the SDR would be another means to resolve this conflict).

Table 1. Fitnesses of the Different Genotypes at the Selected Locus.

	AA	Aa	Aa	A	a
Sporophyte	1	$1 + h\,s_d$	$1 + s_d$		
Female gametophyte				$1 + s_f$	1
Male gametophyte				$1 + s_m$	1

Discussion

The *Ectocarpus* PAR Does Not Exhibit an Increased Recombination Rate Compared with Autosomes but Does Exhibit Local Peaks of Recombination

PARs play a critical role in successful progression through meiosis in the heterogametic sex of most plant and animal species because at least one crossover is required for correct segregation of the sex chromosomes (e.g., Burgoyne et al. 1992; Wai et al. 2012), generating a strong selective force to maintain recombination in the PAR. Accordingly, in human males, PAR1 has a crossover rate that is 17-fold greater than the genome-wide average. In contrast, the recombination rate in females, where recombination is between homologous X chromosomes, is comparable to the genome-wide average (Page et al. 1987; Flaquer et al. 2008). In UV systems, meiosis occurs in the sporophyte and, consequently, there is no male or female meiosis and all meiotic events involve pairs of U and V chromosomes in which recombination can only occur in the PARs. This feature of UV systems might be expected to further increase overall recombination rates in the PAR, but measurement of the recombination rate along the *Ectocarpus* PAR indicated a mean rate that was not significantly different

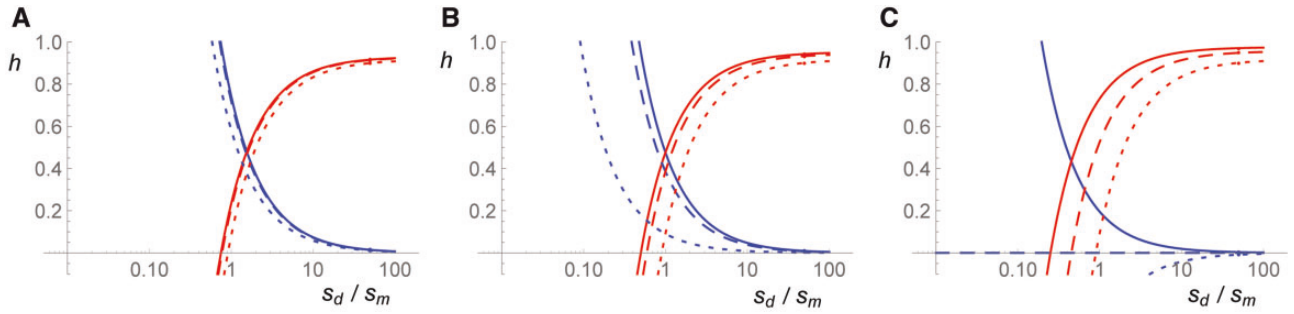


Fig. 5. Conditions for the spread of generation-antagonistic alleles. The sporophyte-beneficial allele (*a*) increases in frequency when rare above the blue curves, while the gametophyte-beneficial allele (*A*) increases in frequency when rare above the red curves. Solid curves: $r = 0.5$; dashed curves: $r = 0.1$; dotted curves: $r = 0.01$. The strength of selection in males s_m is fixed to 0.1, while the different panels correspond to different values of s_f: 0.05 (*A*), 0 (*B*), and − 0.05 (*C*). Note that swapping s_f and s_m would yield exactly the same results, as the model assumes that both sexes are equivalent.

from that of the rest of the genome. The absence of a detectable increase in recombination rate is probably explained by the large relative size of the PAR in *Ectocarpus*, which occupies approximately 80% of the sex chromosome (Ahmed et al. 2014). Similarly, the PAR of the blood fluke, and the PAR of the emu, which represent a high proportion (57% and 75%, respectively) of their sex chromosomes, both exhibit average recombination rates that are similar to those of autosomes (Criscione et al. 2009; Janes et al. 2009). Therefore, there appears to be a general tendency for PARs that constitute a large proportion of the physical size of the sex chromosome not to exhibit increased recombination rates compared with autosomes.

Although the mean recombination rate along the *Ectocarpus* PAR was comparable to that measured for autosomes, recombination mapping identified two peaks of elevated recombination rates flanking the SDR. Fine scale mapping of recombination rates along all the *Ectocarpus* linkage groups will be required to determine whether this type of recombination peak is a specific feature of the sex chromosome or if such peaks occur in autosomes (e.g., surrounding regions of reduced recombination such as the centromeres). Recombination hotspots at borders of SDRs have been described for species with XY or ZW sexual systems, including humans (Flaquer et al. 2008), mice (Soriano et al. 1987), blood flukes (Criscione et al. 2009), medaka fish (Kondo et al. 2001), emu (Janes et al. 2009), flycatcher birds (Smeds et al. 2014), *Populus* (Yin et al. 2008), and papaya (Wai et al. 2012). A similar phenomenon has also been observed in fungal mating type chromosomes (Hsueh et al. 2006).

The PARs Recombine at Similar Levels to the Rest of the Genome but Exhibit Structural Characteristics Typical of Nonrecombining Regions

The *Ectocarpus* PARs exhibit a number of features that are typical of genomic regions with reduced levels of genetic recombination (Charlesworth D and Charlesworth B 2005), including increased TE content, decreased gene density, smaller average coding sequence size, larger average intron size, higher gene GC content, higher rates of both synonymous and nonsynonymous substitution (and higher dN/dS ratios), and lower average gene expression levels compared with autosomes. Paradoxically, despite these features, the mean recombination rate measured for the PAR was not significantly different from that of the autosomal part of the genome. Moreover, we found no evidence that the majority of the PAR genes (excluding sporophyte-biased genes) contained higher levels of suboptimal codons than autosomal genes. However, note that PAR gene coding regions are significantly shorter than those of autosomal genes and this might counteract any tendency for suboptimal codons to accumulate, because selective pressures on codon usage are typically stronger for genes that encode short proteins (Duret and Mouchiroud 1999). The PAR was found to be enriched in young genes compared with autosomes but while the presence of these genes contributes to some extent to the unusual features of this region, this enrichment alone does not explain all the unusual structural features of the PAR.

We considered possible evolutionary mechanisms that might explain these unusual structural and functional features of the PAR and its constituent genes. Genetic linkage to the SDR is expected to influence the evolution of the PAR, but the effect should be limited to regions of the PAR that are very close to the SDR, unless selection is very strong (Charlesworth et al. 2014). This was not the case for the *Ectocarpus* PAR, as the unusual structural features were characteristic of the entire PAR and were not limited to regions adjacent to the SDR. To date, no mechanisms have been proposed which would allow the SDR to influence the evolution of linked, recombining regions over the distances observed here. It is not clear at present, therefore, whether the unusual structural features of the *Ectocarpus* PAR are related in some way to the presence of the SDR on the same chromosome or if they indicate that the evolutionary history of the PAR has been different from that of the other autosomes.

Preferential Accumulation of Sporophyte-Biased Genes on the PAR

The *Ectocarpus* PAR is enriched in sporophyte-biased genes compared with the autosomes and these sporophyte-biased genes appear to be evolving in a different manner to the other genes on the PAR. PAR genes in general showed elevated

levels of both synonymous and nonsynonymous mutations compared with autosomal (LG04) genes whereas the sporophyte-biased PAR genes showed highly elevated rates of nonsynonymous mutations but a similar synonymous mutation rate to unbiased autosomal (LG04) genes. The elevated rate of nonsynonymous substitutions could be indicative of adaptive evolution, and indeed a signature of positive selection was detected for one out of the seven sporophyte-biased PAR genes that could be analyzed for this feature. However, while positive selection may be driving the evolution of some of the sporophyte-biased genes, this is unlikely to be the case for all of them. The set of sporophyte-biased PAR genes had a reduced content of optimal codons compared with an autosomal gene set, suggesting that the majority of these genes are under relaxed purifying selection. One possible explanation for the accumulation of nonoptimal codons in these genes is that they may escape haploid purifying selection (Lewis and Benson-Evans 1960; Lewis 1961; Lewis and John 1968), as they are completely silent during the gametophyte generation. Consequently, alleles with suboptimal codons will be masked in diploid heterozygous individuals and will not be selected against during the haploid phase.

Another possibility is that the lack of expression of the sporophyte-biased PAR genes during the gametophyte generation leads to relaxed selection due to the reduced breadth of expression of these genes. Breadth of expression, that is, the degree of tissue or developmental stage specificity, is known to effect nonsynonymous substitution rates (Duret and Mouchiroud 2000). However, this hypothesis alone is not sufficient to explain the higher evolutionary rates of sporophyte-biased genes, because gametophyte-biased PAR genes, which also have a reduced breath of expression, had similar nonsynonymous mutation rates to an average PAR gene.

Mathematical modeling was used to identify evolutionary mechanisms that might explain the preferential accumulation of sporophyte-biased genes in the PAR. Consistent with a recent model proposed by Immler and Otto (2015), we show that generation-antagonistic alleles spread more easily on the PAR than on autosomes if selection differs between males and females. The model presented here may explain our empirical observations that generation-biased genes accumulate preferentially on the PAR, provided that differences in expression between generations result from generation-antagonistic selection. However, note that there is evidence that the relationship between sex-biased gene expression and SA selection is complex (Innocenti and Morrow 2010; Parsch and Ellegren 2013) and this is likely also to be the case for the relationship between generation-biased gene expression and generation antagonistic selection. Generation-biased gene expression may therefore only be an approximate proxy for generation-antagonism.

Our model also predicts that sporophyte-beneficial, gametophyte-detrimental alleles tend to benefit more from linkage to the SDR than gametophyte-beneficial, sporophyte-detrimental alleles, in situations where selection is much weaker in one sex than in the other. Such a process might explain the prevalence of sporophyte-specific genes in the *Ectocarpus* PAR. Although this phenomenon should occur predominantly in regions that are tightly linked to the SDR (as the influence of the SDR is predicted to decrease rapidly with genetic distance), it may extend over a larger proportion of the PAR if the reproductive system involves partial clonality or inbreeding (thereby reducing effective recombination rates). Note that a recent field study identified both sexual populations and populations that were reproducing asexually (Couceiro et al. 2015), consistent with significant levels of asexual reproduction occurring under some conditions.

Young genes are 3-fold more abundant in the PARs than in autosomes. This enrichment is likely to be due to a combination of factors. As new genes are often derived from TEs (Tautz and Domazet-Loso 2011; Arendsee et al. 2014) the higher density of TEs in the PARs may play a role by permitting a higher rate of creation of new transcribed loci. This hypothesis is supported by the fact that, compared with the young autosomal genes, a greater proportion of the young genes in the PARs share homology with elements in the repeated fraction of the *Ectocarpus* genome (46.3% compared with 31.7%, Mann–Whitney *U* test, $P = 0.038$). Note, however, that additional factors are likely to be operating because this mechanism does not explain why the young PAR genes are enriched 2-fold in sporophyte-biased genes compared with young autosomal genes. Novel, transcribed loci are thought to arise at a high frequency in the genome but most of these loci are thought to be subsequently lost unless they are stabilized by selective forces (Tautz and Domazet-Loso 2011; Arendsee et al. 2014). It is possible that the mechanism considered in our model (where the excess of sporophyte-biased genes on the PAR results from the spread of sporophyte-beneficial, gametophyte-detrimental alleles) promotes the emergence of new genes with sporophyte-biased expression in the PAR. However, this mechanism alone does not seem sufficient to explain the high proportion of young PAR genes that are generation-biased (52%), as it seems unlikely that such a high proportion of the selectively advantageous new genes have generation-antagonistic effects.

Sporophyte-Biased Genes in the PAR Occur in Clusters

Almost half of the sporophyte-biased PAR genes are located in two gene clusters that are highly enriched in sporophyte-biased genes. Clustering of genes with related functions does occur in eukaryotic genomes, although to a lesser extent than in prokaryotes (Williams and Hurst 2002; Mugford et al. 2013), but the *Ectocarpus* genome as a whole does not exhibit unusually high levels of functional clustering (Cock et al. 2010). At present it is not clear what mechanisms led to the formation of these gene clusters on the PAR. Gene duplication has not played a major role in the evolution of these clusters although there are paralogous pairs of two genes across the two clusters. The model presented in this manuscript provides a possible mechanism for the accumulation of sporophyte-biased genes near the SDR and this could lead to clustering. However, neither cluster is adjacent to the SDR, although it is possible that the clusters have translocated to

their current positions as a result of sex chromosome rearrangements.

Conclusion

We provide the first detailed analysis of the structural and evolutionary features of the PAR of a pair of UV sex chromosomes. We show that this PAR recombines at a rate that is not different from any other region of the genome, but remarkably, exhibits a number of structural and evolutionary features that are typically associated with regions of suppressed recombination. The PAR has significantly accumulated clusters of genes that are differentially expressed during the sporophyte versus gametophyte generation of the life cycle, and these generation-specific genes exhibit clear signs of accelerated evolution. We propose a mechanism that may explain some of the exceptional evolutionary features of these regions compared with autosomes.

Materials and Methods

Ectocarpus Strains and Culture Conditions

Ectocarpus strains were cultured as described (Coelho et al. 2012) and details are provided in supplementary material S1, Supplementary Material online. Supplementary table S5, Supplementary Material online, describes the *Ectocarpus* species used in this study. Note that, currently only three species are recognized within the genus *Ectocarpus* (*E. siliculosus*, *E. fasciculatus*, and *E. crouanorium*) (Peters et al. 2010) but there is increasing evidence that the taxa *E. siliculosus* represents a complex of several species. As the type specimen for *E. siliculosus* was isolated in England, we refer to the non-European strains related to *E. siliculosus* (such as the Peruvian and Greenland strains) as "*Ectocarpus* sp.". The *E.* sp. lineage 1c Peru is the reference species of *Ectocarpus* used for the genome sequencing project and genetic map (Cock et al. 2010; Heesch et al. 2010). To study PAR recombination in an additional *Ectocarpus* species we used *E. siliculosus* lineage 1a. *E.* sp. lineage 1c Greenland and *E. fasciculatus* lineage 5b were used to evaluation of rates of gene evolution.

Generation of a Fine Recombination Map

A segregating population of 60 individuals that had been used for the genetic map (Heesch et al. 2010) and additional 220 individuals from a segregating population derived from a cross between strains Ec494 (male) and Ec568 (female) (Ahmed et al. 2014) were used to more precisely estimate recombination frequencies across the PAR. Simple sequence repeat markers for each of the 23 supercontigs of the sex chromosome (LG30) have been described previously (Heesch et al. 2010; Ahmed et al. 2014), and additional markers are described in supplementary table S2, Supplementary Material online, and in supplementary material S1, Supplementary Material online.

RNA-Seq

RNA-Seq analysis was carried out to compare the relative abundances of PAR gene transcripts at several different developmental stages of the life cycle (immature and fertile male and female gametophytes and two tissues of the sporophyte generation, namely basal filaments and upright filaments). The RNA extractions and processing of sequenced reads were performed as previously described in Ahmed et al. (2014) and Lipinska et al. (2015) (see supplementary table S6, Supplementary Material online, for the sequencing and mapping statistics and supplementary material S1, Supplementary Material online, for details of the Materials and Methods).

Differential expression analysis between male and female gametophytes, as well as between gametophyte (male and female libraries as replicates) and sporophyte was performed with the DESeq package (Bioconductor) (Anders and Huber 2010) using an adjusted *P* value cutoff of 0.1 and a minimal fold-change of 2 (see supplementary material S1, Supplementary Material online, for more details). The PAR was also analyzed for the presence of duplicated genes. Duplicated gene pairs were detected as described in Cock et al. (2010).

Evaluation of Rates of Gene Evolution

To estimate evolutionary rates of PAR genes, we searched *E. siliculosus* lineage 1a genomic data for orthologues of *E.* sp. lineage 1c Peru genes by retaining best reciprocal BLASTn matches with a minimum *e* value of 10×10^{-10}. Sequences that produced a gapless alignment that exceeded 100 bp were retained for pairwise dN/dS (ω) analysis using PAML (codeml, F3x4 model, runmode = −2). To detect PAR genes under positive selection, we used transcriptomic and genomic data from four different *Ectocarpus* species (detailed in supplementary material S1 and table S5, Supplementary Material online). Nucleotide alignments (with a minimum length of 100 bp) were constructed using ClustalW implemented in Mega6 (Larkin et al. 2007; Tamura et al. 2013), curated manually when necessary and transformed to the PAML4 required format using perl fasta manipulation scripts (provided by Naoki Takebayashi, University Alaska Fairbanks). Nonsynonymous (dN) and synonymous (dS) rates were estimated by the maximum likelihood method available in CODEML program (PAML4 package). Effective number of codons and CAI were calculated using CAIcal server (http://genomes.urv.es/CAIcal/) (Puigbo et al. 2008).

Classification of "Old" and "Young" Genes

To determine the effect of gene age on various structural parameters, *Ectocarpus* genes were classified as "old genes" or as "young genes" based on the presence or absence, respectively, of homologous sequences in seven complete stramenopile genomes or in the NCBI database (excluding *Ectocarpus* sequences; February 2015). For the stramenopiles, BLASTp searches were carried out against the following complete deduced proteomes: *Thalassiosira pseudonana* (diatom; Thaps3 assembled and unmapped scaffolds, http://genome.jgi-psf.org/Thaps3/Thaps3.download.ftp.html; Armbrust et al. 2004), *Phaeodactylum tricornutum* (diatom; Phatr2 assembled and unmapped scaffolds, http://genome.jgi-psf.org/Phatr2/Phatr2.download.ftp.html; Bowler et al. 2008),

Aureococcus anophagefferens (Pelagophyceae; http://genome.jgi-psf.org/Auran1/Auran1.download.ftp.html; Gobler et al. 2011); *Nannochloropsis oceanica* (Eustigmatophyceae; https://bmb.natsci.msu.edu/BMB/assets/File/benning/genome_assembly.txt; Vieler et al. 2012), *Nannochloropsis gaditana* (Eustigmatophyceae; http://www.nature.com/ncomms/journal/v3/n2/full/ncomms1688.html; Radakovits et al. 2012), *Phytophthora capsici* (oomycete; http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3551261; Lamour et al. 2012), and *S. japonica* (http://124.16.129.28:8080/saccharina/; Ye et al. 2015). Recent estimates indicate that all these species diverged from the Ectocarpales lineage more than 100 Ma (Brown and Sorhannus 2010; Silberfeld et al. 2010). Genes were classified as old genes if their protein sequences detected a BLASTp match with an *E* value of less than 10^{-4} in any of the subject genomes.

Supplementary Material

Supplementary material S1, tables S1–S7, and figures S1–S7 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the Centre National de la Recherche Scientifique, the Agence Nationale de la Recherche (project SEXSEAWEED ANR12-JSV7-0008, project Bi-CYCLE ANR-10-BLAN-1727, and project IDEALG), the University Pierre and Marie Curie Emergence program, the Interreg program France (Channel)-England (project MARINEXUS), and the ERC (grant agreement 638240). The authors thank Lynn Delgat and Anne Vanderheyden for help with the alignments for the dN/dS analysis, Claire Gachon for sharing unpublished sequence data from *E. fasciculatus* and *E.* sp. lineage 1c Greenland, and Didier Jollivet and Thomas Broquet for helpful discussions. The authors also like to thank three anonymous reviewers for their comments and suggestions that helped improve the manuscript.

References

Ahmed S, Cock JM, Pessia E, Luthringer R, Cormier A, Robuchon M, Sterck L, Peters AF, Dittami SM, Corre E, et al. 2014. A haploid system of sex determination in the brown alga *Ectocarpus* sp. *Curr Biol.* 24:1945-1957.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol.* 11:R106.

Arendsee ZW, Li L, Wurtele ES. 2014. Coming of age: orphan genes in plants. *Trends Plant Sci.* 19:698-708.

Armbrust E, Berges J, Bowler C, Green B, Martinez D, Putnam N, Zhou S, Allen A, Apt K, Bechner M, et al. 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* 306:79-86.

Bachtrog D, Kirkpatrick M, Mank JE, McDaniel SF, Pires JC, Rice W, Valenzuela N. 2011. Are all sex chromosomes created equal? *Trends Genet.* 27:350-357.

Bowler C, Allen A, Badger J, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar R, et al. 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456:239-244.

Brown JW, Sorhannus U. 2010. A molecular genetic timescale for the diversification of autotrophic stramenopiles (Ochrophyta): substantive underestimation of putative fossil ages. *PLoS One* 5:e12759.

Burgoyne PS, Mahadevaiah SK, Sutcliffe MJ, Palmer SJ. 1992. Fertility in mice requires X-Y pairing and a Y-chromosomal "spermiogenesis" gene mapping to the long arm. *Cell* 71:391-398.

Burt DW. 2002. Origin and evolution of avian microchromosomes. *Cytogenet Genome Res.* 96:97-112.

Campos JL, Charlesworth B, Haddrill PR. 2012. Molecular evolution in nonrecombining regions of the *Drosophila melanogaster* genome. *Genome Biol Evol.* 4:278-288.

Charlesworth B, Jordan CY, Charlesworth D. 2014. The evolutionary dynamics of sexually antagonistic mutations in pseudoautosomal regions of sex chromosomes. *Evolution* 68:1339-1350.

Charlesworth D, Charlesworth B. 2005. Sex chromosomes: evolution of the weird and wonderful. *Curr Biol.* 15:R129-R131.

Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* 95:118-128.

Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, Amoutzias G, Anthouard V, Artiguenave F, Aury J, Badger J, et al. 2010. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature* 465:617-621.

Coelho SM, Scornet D, Rousvoal S, Peters NT, Dartevelle L, Peters AF, Cock JM. 2012. How to cultivate *Ectocarpus*. *Cold Spring Harb Protoc.* 2012:258-261.

Couceiro L, Le Gac M, Hunsperger HM, Mauger S, Destombe C, Cock JM, Ahmed S, Coelho SM, Valero M, Peters AF. 2015. Evolution and maintenance of haploid-diploid life cycles in natural populations: the case of the marine brown alga *Ectocarpus*. *Evolution* 69:1808-1822.

Criscione CD, Valentim CL, Hirai H, LoVerde PT, Anderson TJ. 2009. Genomic linkage map of the human blood fluke *Schistosoma mansoni*. *Genome Biol.* 10:R71.

Duret L, Mouchiroud D. 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc Natl Acad Sci U S A.* 96:4482-4487.

Duret L, Mouchiroud D. 2000. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. *Mol Biol Evol.* 17:68-74.

Flaquer A, Rappold GA, Wienker TF, Fischer C. 2008. The human pseudoautosomal regions: a review for genetic epidemiologists. *Eur J Hum Genet.* 16:771-779.

Gobler CJ, Berry DL, Dyhrman ST, Wilhelm SW, Salamov A, Lobanov AV, Zhang Y, Collier JL, Wurch LL, Kustka AB, et al. 2011. Niche of harmful alga *Aureococcus anophagefferens* revealed through ecogenomics. *Proc Natl Acad Sci U S A.* 108:4352-4357.

Haddrill PR, Halligan DL, Tomaras D, Charlesworth B. 2007. Reduced efficacy of selection in regions of the *Drosophila* genome that lack crossing over. *Genome Biol.* 8:R18.

Heesch S, Cho GY, Peters AF, Le Corguille G, Falentin C, Boutet G, Coedel S, Jubin C, Samson G, Corre E, et al. 2010. A sequence-tagged genetic map for the brown alga *Ectocarpus siliculosus* provides large-scale assembly of the genome sequence. *New Phytol.* 188:42-51.

Hinch AG, Altemose N, Noor N, Donnelly P, Myers SR. 2014. Recombination in the human Pseudoautosomal region PAR1. *PLoS Genet.* 10:e1004503.

Hsueh YP, Idnurm A, Heitman J. 2006. Recombination hotspots flank the *Cryptococcus* mating-type locus: implications for the evolution of a fungal sex chromosome. *PLoS Genet.* 2:e184.

Immler S, Otto SP. 2015. The evolution of sex chromosomes in organisms with separate haploid sexes. *Evolution*.

Innocenti P, Morrow EH. 2010. The sexually antagonistic genes of *Drosophila melanogaster*. *PLoS Biol.* 8:e1000335.

International Chicken Genome Sequencing Consortium. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695-716.

Janes DE, Ezaz T, Marshall Graves JA, Edwards SV. 2009. Recombination and nucleotide diversity in the sex chromosomal pseudoautosomal region of the emu, *Dromaius novaehollandiae*. *J Hered.* 100:125-136.

Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen CF, Thomas MA, Haussler D, Jacob HJ. 2004. Comparative

recombination rates in the rat, mouse, and human genomes. *Genome Res.* 14:528-538.
Kanaya S, Yamada Y, Kinouchi M, Kudo Y, Ikemura T. 2001. Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis. *J Mol Evol.* 53:290-298.
Kirkpatrick M, Guerrero RF. 2014. Signatures of sex-antagonistic selection on recombining sex chromosomes. *Genetics* 197:531-541.
Kondo M, Nagao E, Mitani H, Shima A. 2001. Differences in recombination frequencies during female and male meioses of the sex chromosomes of the medaka, *Oryzias latipes*. *Genet Res.* 78:23-30.
Lamour KH, Mudge J, Gobena D, Hurtado-Gonzales OP, Schmutz J, Kuo A, Miller NA, Rice BJ, Raffaele S, Cano LM, et al. 2012. Genome sequencing and mapping reveal loss of heterozygosity as a mechanism for rapid adaptation in the vegetable pathogen *Phytophthora capsici*. *Mol Plant Microbe Interact.* 25:1350-1360.
Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947-2948.
Lewis KR. 1961. The genetics of bryophytes. *Trans Br Bryol Soc.* 4:111-130.
Lewis KR, Benson-Evans K. 1960. The chromosomes of *Cryptothallbus mirabilis* (Hepaticae: Riccardiaceae). *Phyton* 14:21-35.
Lewis KR, John B. 1968. The chromosomal basis of sex determination. *Int Rev Cytol.* 17:277-379.
Lien S, Szyda J, Schechinger B, Rappold G, Arnheim N. 2000. Evidence for heterogeneity in recombination in the human pseudoautosomal region: high resolution analysis by sperm typing and radiation-hybrid mapping. *Am J Hum Genet.* 66:557-566.
Lipinska A, Luthringer R, Peters AF, Corre E, Gachon CMM, Cock JM, Coelho SM. 2015. Sexual dimorphism and the evolution of sex-biased gene expression in the brown alga *Ectocarpus*. *Mol Biol Evol.* 32:1581–1597.
Montoya-Burgos JI, Boursot P, Galtier N. 2003. Recombination explains isochores in mammalian genomes. *Trends Genet.* 19:128-130.
Mugford ST, Louveau T, Melton R, Qi X, Bakht S, Hill L, Tsurushima T, Honkanen S, Rosser SJ, Lomonossoff GP, et al. 2013. Modularity of plant metabolic gene clusters: a trio of linked genes that are collectively required for acylation of triterpenes in oat. *Plant Cell* 25:1078-1092.
Otto SP, Pannell JR, Peichel CL, Ashman TL, Charlesworth D, Chippindale AK, Delph LF, Guerrero RF, Scarpino SV, McAllister BF. 2011. About PAR: the distinct evolutionary dynamics of the pseudoautosomal region. *Trends Genet.* 27:358-367.
Page DC, Bieker K, Brown LG, Hinton S, Leppert M, Lalouel JM, Lathrop M, Nystrom-Lahti M, de la Chapelle A, White R. 1987. Linkage, physical mapping, and DNA sequence analysis of pseudoautosomal loci on the human X and Y chromosomes. *Genomics* 1:243-256.
Palmieri N, Kosiol C, Schlotterer C. 2014. The life cycle of *Drosophila* orphan genes. *Elife* 3:e01311.
Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression. *Nat Rev Genet.* 14:83-87.
Peters AF, van Wijk SJ, Cho GY, Scornet D, Hanyuda T, Kawai H, Schroeder DC, Cock JM, Boo SM. 2010. Reinstatement of *E. crouaniorum* Thuret in Le Jolis as a third common species of *Ectocarpus* (Ectocarpales, Phaeophyceae) in western Europe, and its phenology at Roscoff, Brittany. *Phycol Res.* 58:157-170.
Puigbo P, Bravo IG, Garcia-Vallve S. 2008. CAIcal: a combined set of tools to assess codon usage adaptation. *Biol Direct.* 3:38.
Radakovits R, Jinkerson RE, Fuerstenberg SI, Tae H, Settlage RE, Boore JL, Posewitz MC. 2012. Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropis gaditana*. *Nat Commun.* 3:686.
Raudsepp T, Das PJ, Avila F, Chowdhary BP. 2012. The pseudoautosomal region and sex chromosome aneuploidies in domestic species. *Sex Dev.* 6:72-83.
Rouyer F, Simmler MC, Johnsson C, Vergnaud G, Cooke HJ, Weissenbach J. 1986. A gradient of sex linkage in the pseudoautosomal region of the human sex chromosomes. *Nature* 319:291-295.
Shi Q, Spriggs E, Field LL, Ko E, Barclay L, Martin RH. 2001. Single sperm typing demonstrates that reduced recombination is associated with the production of aneuploid 24,XY human sperm. *Am J Med Genet.* 99:34-38.
Silberfeld T, Leigh JW, Verbruggen H, Cruaud C, de Reviers B, Rousseau F. 2010. A multi-locus time-calibrated phylogeny of the brown algae (Heterokonta, Ochrophyta, Phaeophyceae): Investigating the evolutionary nature of the "brown algal crown radiation". *Mol Phylogenet Evol.* 56:659-674.
Smeds L, Kawakami T, Burri R, Bolivar P, Husby A, Qvarnstrom A, Uebbing S, Ellegren H. 2014. Genomic identification and characterization of the pseudoautosomal region in highly differentiated avian sex chromosomes. *Nat Commun.* 5:5448.
Soriano P, Keitges EA, Schorderet DF, Harbers K, Gartler SM, Jaenisch R. 1987. High rate of recombination and double crossovers in the mouse pseudoautosomal region during male meiosis. *Proc Natl Acad Sci U S A.* 84:7218-7220.
Stache-Crain B, Muüller DG, Goff LJ. 1997. Molecular systematics of *Ectocarpus* and *Kuckuckia* (Ectocarpales, Phaeophyceae) inferred from phylogenetic analysis of nuclear and plastid-encoded DNA sequences. *J Phycol.* 33:152-168.
Swanson WJ, Wong A, Wolfner MF, Aquadro CF. 2004. Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. *Genetics* 168:1457-1465.
Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 30:2725-2729.
Tautz D, Domazet-Loso T. 2011. The evolutionary origin of orphan genes. *Nat Rev Genet.* 12:692-702.
Vicoso B, Charlesworth B. 2006. Evolution on the X chromosome: unusual patterns and processes. *Nat Rev Genet.* 7:645-653.
Vicoso B, Kaiser VB, Bachtrog D. 2013. Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. *Proc Natl Acad Sci U S A.* 110:6453-6458.
Vieler A, Wu G, Tsai CH, Bullard B, Cornish AJ, Harvey C, Reca IB, Thornburg C, Achawanantakun R, Buehl CJ, et al. 2012. Genome, functional gene annotation, and nuclear transformation of the heterokont oleaginous alga *Nannochloropsis oceanica* CCMP1779. *PLoS Genet.* 8:e1003064.
Wai CM, Moore PH, Paull RE, Ming R, Yu Q. 2012. An integrated cytogenetic and physical map reveals unevenly distributed recombination spots along the papaya sex chromosomes. *Chromosome Res.* 20:753-767.
Williams EJ, Hurst LD. 2002. Clustering of tissue-specific genes underlies much of the similarity in rates of protein evolution of linked genes. *J Mol Evol.* 54:511-518.
Wu X, Tronholm A, Caceres EF, Tovar-Corona JM, Chen L, Urrutia AO, Hurst LD. 2013. Evidence for deep phylogenetic conservation of exonic splice-related constraints: splice-related skews at exonic ends in the brown alga *Ectocarpus* are common and resemble those seen in humans. *Genome Biol Evol.* 5:1731-1745.
Ye N, Zhang X, Miao M, Fan X, Zheng Y, Xu D, Wang J, Zhou L, Wang D, Gao Y, et al. 2015. *Saccharina* genomes provide novel insight into kelp biology. *Nat Commun.* 6:6986.
Yin T, Difazio SP, Gunter LE, Zhang X, Sewell MM, Woolbright SA, Allan GJ, Kelleher CT, Douglas CJ, Wang M, et al. 2008. Genome structure and emerging evidence of an incipient sex chromosome in *Populus*. *Genome Res.* 18:422-430.

Discussion et perspectives

Les analyses menées sur l'étude des caractéristiques de la PAR ont montré que l'ensemble de cette région présente un profil particulier par rapport au reste du génome, avec des caractéristiques proches de celles de la SDR. Ainsi, la densité d'éléments transposables est supérieure à celle des autosomes, tandis que la densité des gènes est plus faible. De plus, la structure des gènes est altérée par rapport aux autosomes, avec une taille supérieure liée à une augmentation de la longueur des introns. En revanche, la taille de la séquence codante diminue, de même que le pourcentage de GC. Les mécanismes à l'origine de ces modifications et l'influence de la SDR sur la PAR restent encore à identifier.

En plus de présenter une modification de sa structure, la PAR présente aussi des modifications au niveau des profils d'expression des gènes. Globalement, le niveau d'expression des gènes par rapport au reste des autosomes est significativement plus faible, plus particulièrement durant la phase gamétophytique. Une caractéristique particulière de la PAR d'*Ectocarpus* est la présence de clusters de gènes ayant un profil d'expression similaire durant le cycle de vie, avec la présence de deux clusters fortement surexprimés durant la phase sporophytique et un cluster constamment sous-exprimé durant la phase gamétophytique.

L'analyse des gènes orphelins menée, même si incomplète du fait du manque de données génomiques disponibles chez les algues brunes, a permis d'identifier chez *Ectocarpus* les gènes « jeunes » et les gènes « ancestraux ». La différence entre les deux catégories est basée sur la présence/absence d'un gène d'une espèce dans les autres espèces proches (Tautz and Domazet-Lošo 2011). Par exemple, un gène d'une espèce non retrouvé, par recherche d'homologie de séquence, dans les autres espèces proches sera qualifié de « jeune ». Chez *Ectocarpus*, la PAR présente un enrichissement en gènes « jeunes » par rapport aux autosomes, enrichissement qui pourrait être expliqué par la plus forte présence d'éléments transposables dans la PAR, qui sont connus pour favoriser l'apparition de nouveaux gènes (Arendsee et al. 2014). Cette hypothèse est supportée par le fait que l'analyse des gènes « jeunes » a montré que ces derniers présentaient une homologie de séquence avec les éléments répétés identifiés dans le génome d'*Ectocarpus*. Cependant, d'autres mécanismes susceptibles d'expliquer l'apparition de nouveaux gènes ne peuvent être exclus.

La disponibilité prochaine de nouvelles données génomiques et transcriptomiques, pour d'autres algues brunes, permettra de disposer de la séquence nucléotidique des chromosomes sexuels d'autres espèces afin de pouvoir étudier les différences structurales et fonctionnelles, ainsi que l'histoire évolutive des PAR dans ce clade. Plus particulièrement, avec l'étude des mouvements de gènes entre la PAR et les SDR, mais aussi entre les autosomes et la PAR, la taille relative des PAR dans les chromosomes sexuels et potentiellement, déterminer l'influence de la SDR sur la dynamique évolutive de la PAR. Les données transcriptomiques permettront de déterminer si la présence du cluster de gènes spécifiques au stade sporophytique est une caractéristique propre à *Ectocarpus* ou se retrouve dans d'autres espèces. Enfin, la disponibilité de ces informations génomiques et transcriptomiques permettra d'identifier de manière précise les gènes orphelins, autorisant l'identification et la datation des évènements d'acquisitions et de pertes de gènes. L'étude des fonctions des gènes acquis à différentes périodes évolutives permettra de mieux appréhender l'histoire évolutive de ce groupe et voir l'adaptation des espèces à leur milieu.

Article 3 - Sexual Dimorphism and the Evolution of Sex-Biased Gene Expression in the Brown Alga *Ectocarpus*

Introduction

L'article présenté porte sur l'identification des gènes biaisés au niveau de leur expression entre les sexes à différents stades du cycle de vie, sur l'analyse des caractéristiques au niveau de leur profile d'expression, et sur l'évolution moléculaire de ces gènes.

Plusieurs séquençages RNA-seq ont été réalisés afin de mesurer l'abondance des gènes lors de différents stades de développement des gamétophytes. Des individus mâles (Ec603) et femelles (Ec602) de deux souches quasi isogéniques ont été séquencés à deux stades de développement, au niveau immature et mature. Pour chaque stade de développement, un total de quatre librairies ont été séquencées, avec deux réplicats biologiques pour chaque sexe.

Ma contribution à cet article a été de préparer les données puis les mapper avec TopHat2, en utilisant comme guide les annotations de références, et obtenir le comptage des reads pour chaque gène avec HTSeq-count. L'analyse de l'expression différentielle entre mâles et femelles a été réalisée en utilisant DESeq afin d'identifier les gènes présentant une différence significative d'expression entre les individus mâles et femelles. Les gènes biaisés par le sexe ont ensuite été analysés afin d'identifier la présence ou non de gènes dupliqués afin de déterminer si ces derniers peuvent être impliqués dans la résolution d'un antagonisme sexuel.

Une partie de ma contribution a été de réaliser l'analyse d'enrichissement fonctionnelle, basée sur l'utilisation des GO termes avec le logiciel Blast2GO, afin de déterminer si des fonctions métaboliques étaient surreprésentées dans les gènes identifiés comme différentiellement exprimés.

Enfin, ma dernière contribution a été de générer l'assemblage *de novo* du transcriptome de l'espèce sœur *Ectocarpus fasciculatus*, avec l'assembleur Trinity, à partir de données provenant de quatre librairies paired-end (deux par sexe – données non publiées) afin de pouvoir analyser la différence de taux d'évolution entre les gènes biaisés et non biaisés par le sexe.

Les résultats de ces travaux ont été intégrés avec d'autres analyses dans l'article suivant, publié dans Molecular Biology and Evolution en février 2015.

Article

Sexual Dimorphism and the Evolution of Sex-Biased Gene Expression in the Brown Alga *Ectocarpus*

Auteurs : Lipinska A[1§], Cormier A[1§], Luthringer R[1], Peters AF[2], Corre E[3], Gachon CMM[4], Cock JM[1], Coelho SM[1*]

Affiliations :

[1] Sorbonne Université, UPMC Univ Paris 06, CNRS, Algal Genetics Group, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688, Roscoff, France;

[2] Bezhin Rosko, Santec, France;

[3] Abims Platform, CNRS-UPMC, FR2424, Station Biologique de Roscoff, Roscoff, France;

[4] Microbial and Molecular Biology Department, Scottish Marine Institute, Scottish Association for Marine Science, Oban, United Kingdom

§égale contribution

Correspondance : coelho@sb-roscoff.fr

Article publié dans Molecular Biology and Evolution

doi : 10.1093/molbev/msv049

Sexual Dimorphism and the Evolution of Sex-Biased Gene Expression in the Brown Alga *Ectocarpus*

Agnieszka Lipinska,[†,1] Alexandre Cormier,[†,1] Rémy Luthringer,[1] Akira F. Peters,[2] Erwan Corre,[3] Claire M.M. Gachon,[4] J. Mark Cock,[1] and Susana M. Coelho[*,1]

[1]Sorbonne Université, UPMC Univ Paris 06, CNRS, Algal Genetics Group, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688, Roscoff, France

[2]Bezhin Rosko, Santec, France

[3]Abims Platform, CNRS-UPMC, FR2424, Station Biologique de Roscoff, Roscoff, France

[4]Microbial and Molecular Biology Department, Scottish Marine Institute, Scottish Association for Marine Science, Oban, United Kingdom

[†]These authors contributed equally to this work.

***Corresponding author:** E-mail: coelho@sb-roscoff.fr.

Associate editor: John Parsch

Abstract

Males and females often have marked phenotypic differences, and the expression of these dissimilarities invariably involves sex differences in gene expression. Sex-biased gene expression has been well characterized in animal species, where a high proportion of the genome may be differentially regulated in males and females during development. Male-biased genes tend to evolve more rapidly than female-biased genes, implying differences in the strength of the selective forces acting on the two sexes. Analyses of sex-biased gene expression have focused on organisms that exhibit separate sexes during the diploid phase of the life cycle (diploid sexual systems), but the genetic nature of the sexual system is expected to influence the evolutionary trajectories of sex-biased genes. We analyze here the patterns of sex-biased gene expression in *Ectocarpus*, a brown alga with haploid sex determination (dioicy) and a low level of phenotypic sexual dimorphism. In *Ectocarpus*, female-biased genes were found to be evolving as rapidly as male-biased genes. Moreover, genes expressed at fertility showed faster rates of evolution than genes expressed in immature gametophytes. Both male- and female-biased genes had a greater proportion of sites experiencing positive selection, suggesting that their accelerated evolution is at least partly driven by adaptive evolution. Gene duplication appears to have played a significant role in the generation of sex-biased genes in *Ectocarpus*, expanding previous models that propose this mechanism for the resolution of sexual antagonism in diploid systems. The patterns of sex-biased gene expression in *Ectocarpus* are consistent both with predicted characteristics of UV (haploid) sexual systems and with the distinctive aspects of this organism's reproductive biology.

Key words: **sex-biased gene expression, haploid–diploid life cycle, brown algae, UV sex chromosomes.**

Introduction

In many animal and plant species, males differ markedly from females in morphology, physiology, and behavior. Most of these phenotypic differences are mediated by differential gene expression in the two sexes (Ellegren and Parsch 2007) and this differential gene expression may involve a significant proportion of the genome, as much as 75% in *Drosophila* for example (Assis et al. 2012). These sexually dimorphic patterns of gene expression evolve as a consequence of different selection pressures acting on males and females.

The advent of new generation sequencing has allowed comparative transcriptomic studies of males and females from a range of different species with separate sexes including *Drosophila* (e.g., Perry et al. 2014), birds (e.g., Pointer et al. 2013; Uebbing et al. 2013), cichlid fishes (Bohne et al. 2014), guppies (Sharma et al. 2014), nematodes (Albritton et al. 2014), moths (Smith et al. 2014), the pea aphid (Jaquiery et al. 2013), and brown algae (Lipinska et al. 2013; Martins et al. 2013). A general theme that has emerged from these studies across diverse species is that a significant proportion of the genes in the genome exhibit sex-biased expression, indicating that the expression of sexual dimorphism is associated with marked genetic reprogramming. In most cases, however, there are marked morphological differences between male and female individuals of the species that were studied and analyses of species displaying different degrees of sexual dimorphism would be useful to test the correlation between this character and level of sex-biased gene expression.

Studies such as those listed above are starting to provide a comprehensive overview of sex-biased gene expression in a broad range of species, but the evolutionary causes and consequences underlying the patterns of sex-biased gene expression have been examined in only a small subset of these systems. Most of our knowledge of how sex-biased genes

Article

evolve comes from work with *Drosophila* and birds (reviewed in Parsch and Ellegren 2013), although some studies have also looked at hermaphrodite species and have provided evidence for sexual selection in these systems (Whittle and Johannesson 2013; Gossmann et al. 2014). Evolutionary analyses have identified several unusual features of sex-biased genes. For example, in gonochoristic/dioecious systems, male-biased genes typically evolve more rapidly at the protein level than female-biased or unbiased genes (e.g., Zhang et al. 2004; Haerty et al. 2007; Assis et al. 2012; reviewed by Ellegren and Parsch 2007; see also Mank et al. 2007). This is believed to result from sex differences in selective pressures on genes; the rapid divergence of male-biased genes resulting from sexual selection due to male–male competition or female choice, natural selection, and/or relaxed purifying selection arising from gene dispensability or reduced functional pleiotropy (Ellegren and Parsch 2007; Mank and Ellegren 2009; Parsch and Ellegren 2013).

The genetic nature of the sexual system can also have an influence, both on the distribution of sex-biased genes in the genome and on their patterns of evolution. In XY sexual systems, for example, X chromosomes spend twice as much time in females as they do in males. Depending on the dominance/recessivity of the male-beneficial allele, this can lead to demasculinization of (i.e., loss of male-biased genes from) the X chromosome (e.g., Arunkumar et al. 2009; Bachtrog et al. 2010; Leder et al. 2010) or to enrichment of male-specific genes on the X (e.g., Khil et al. 2004; Bellott et al. 2010; Jaquiery et al. 2013). Moreover, adaptive fixation of recessive beneficial mutations in X-linked genes (Charlesworth et al. 1987), mutational biases associated with dosage compensation (Begun et al. 2007), or the smaller effective population size (Ne) of sex chromosomes (Vicoso and Charlesworth 2009) may cause genes located on the X (and Z) to evolve more rapidly, the so called faster X effect, and this phenomenon has been observed experimentally, at least in some systems (Presgraves 2008; Mank et al. 2010; Kayserili et al. 2012; Meisel et al. 2012; Avila et al. 2014; Campos et al. 2014; Kousathanas et al. 2014).

These latter effects have not yet been investigated in so-called UV sexual systems, commonly found in mosses and many algae, in which sexuality is expressed during the haploid phase of the life cycle (Bachtrog et al. 2011). There are several important differences between UV systems and the more intensely studied XY and ZW systems and these are expected to have consequences for the evolution of sex-biased genes and for the expression patterns of these genes. For example, in XY and ZW systems recombination is suppressed only for the Y or W chromosome. The X and Z chromosomes can recombine because they are homozygous in one of the sexes. In contrast, in UV systems neither the U nor the V recombines. Moreover, despite the fact that they do not recombine, U and V chromosomes are expected to degenerate less markedly than Y and W chromosomes because they function in a haploid context where both the U and the V are directly exposed to purifying selection (Bull 1978). Finally, the effective population sizes of sex chromosomes differ across different sexual systems and this can have a marked effect on the evolution of the genes carried by these chromosomes. Both the U and the V chromosome have half the effective population size of autosomes (all else being equal) whereas in XY and ZW systems the Y/W and X/Z chromosomes have a quarter or three quarters the population size of autosomes, respectively. As far as sex-biased genes are concerned, masculinization or feminization of sex chromosomes is expected in UV systems only at regions very closely linked to the nonrecombining region because the male and female sex-determining region (SDR) haplotypes function in independent, haploid, male and female individuals. Similarly, a phenomenon similar to the faster X effect is not expected because there is no equivalent of the X chromosome, which recombines but is hemizygous in half of the individuals. Moreover, recent transcriptomic studies from a diverse range of species and tissues (reviewed in Mank 2013) suggest that incomplete or imperfect dosage compensation may be responsible for an important proportion of sex-biased gene expression. Dosage compensation is not expected to occur in UV systems because the U and V chromosomes determine sex during the haploid phase and thus gene dosage is the same for the sex chromosomes and the autosomes.

On the other hand, other features are anticipated to be shared by both diploid (XY and ZW) and haploid (UV) sex-determination systems. For example, in any sexual system resolution of sexual antagonism is expected to be one of the processes that lead to the emergence of sex-biased gene expression. Theoretical models predict that sexually antagonistic alleles should accumulate in the pseudoautosomal regions (PARs) of sex chromosomes, because even partial linkage to the SDR can be adaptive, allowing alleles to be at least partially restricted to the sex for which they are best adapted (Otto et al. 2011; Charlesworth et al. 2014). This effect is expected not only for the PARs of Y and W chromosomes but also for the PARs of U and V chromosomes. This accumulation of sexually antagonistic genes (Charlesworth et al. 2014; Kirkpatrick and Guerrero 2014) might be expected to lead to the PARs becoming enriched in sex-biased genes, although note that there is evidence that the relationship between sexual antagonism and sex-biased gene expression may be quite complex (Innocenti and Morrow 2010; Parsch and Ellegren 2013).

This study focused on sex-biased gene expression in the model brown alga *Ectocarpus*. Brown algae are a group of multicellular photosynthetic organisms that have been evolving independently of both animals and green plants for more than a billion years (Cock, Coelho, et al. 2010). As a group, the brown algae are of considerable interest for investigating the origins and evolution of sexual systems because they have a remarkable variety of levels of sexual dimorphism, reproductive system, types of life cycle, and sex chromosome systems. *Ectocarpus* is a small, filamentous alga that exhibits limited levels of sexual dimorphism, male and female individuals of the sexual phase of its haploid–diploid life cycle, the gametophyte, are morphologically similar organisms and both produce small flagellated gametes (Luthringer et al. 2015). Sex determination in this organism was recently

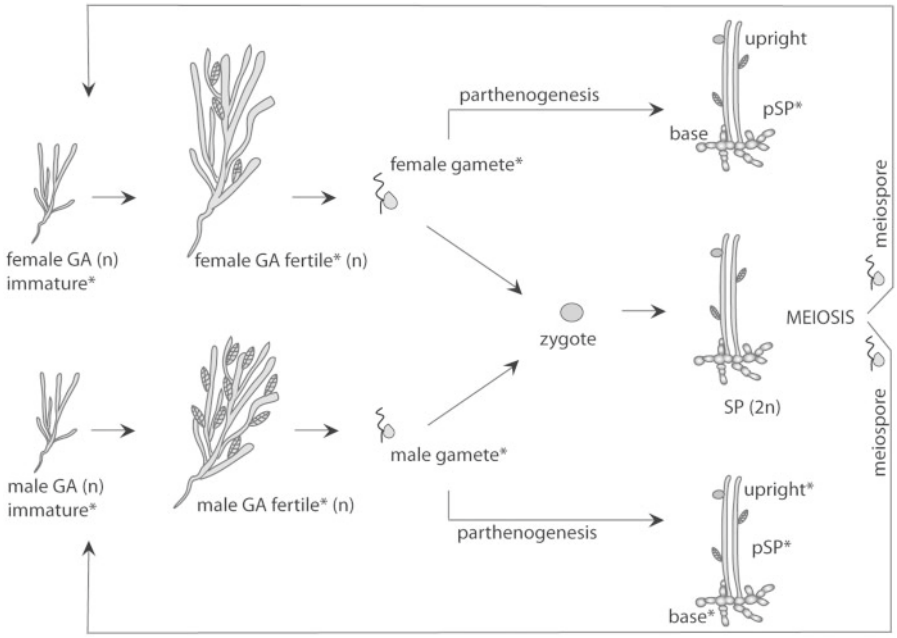


Fig. 1. The *Ectocarpus* life cycle. The life cycle of *Ectocarpus* sp. involves alternation between two independent multicellular generations, the gametophyte (GA) and the sporophyte (SP). Sporophytes produce meiotic spores (meiospores) that develop into haploid gametophytes, which are either male or female (dioecious). After approximately 3 weeks, gametophytes become fertile and produce gametes in reproductive structures (plurilocular gametangia). After release into the water column, male and female gametes strongly differ in their behavior and physiology. Female gametes settle rapidly and release a pheromone to attract male gametes, which then fuse with the female gametes to form zygotes (syngamy). Zygotes develop to produce diploid sporophytes, completing the cycle. Gametes that fail to fuse are able to develop parthenogenetically into haploid parthenosporophytes (pSP). Parthenogenesis is depicted for both male and female gametes. This is observed in some strains but in the majority of *Ectocarpus* species only the females are capable of parthenogenesis. Partheno-sporophytes are morphologically and functionally indistinguishable from diploid sporophytes. Life cycle stages used for transcriptomic analysis are marked with an asterisk.

shown to involve a UV sex chromosome system (Ahmed et al. 2014). In the present study, the level of sexual dimorphism in *Ectocarpus* was precisely quantified using morphometric methods and RNA sequencing (RNA-seq) was used to characterize sex-biased expression. Several unusual features were noted, compared with previously characterized sexual systems. First, fewer than 12% of *Ectocarpus* genes exhibited sex-biased expression, consistent with the low level of sexual dimorphism in this species. Second, both male and female sex-biased genes showed accelerated rates of evolution compared with unbiased genes, with male- and female-biased genes evolving at a similar pace. This balanced rate of evolution is also consistent with the low level of sexual dimorphism, which presumably provides limited scope for asymmetric sexual selection. Gene duplication has played a significant role in the generation of sex-biased genes in *Ectocarpus* and the evolution of these genes has been shaped by both positive selection and relaxation of purifying selection. We identified no clear effects of the UV sexual system on the genomic distribution of sex-biased genes but the PAR was found to be enriched in female-biased genes expressed during the mature gametophyte stage.

Results

Ectocarpus Exhibits a Low Level of Sexual Dimorphism

Sex is determined genetically during the haploid gametophyte generation of the *Ectocarpus* haploid–diploid life cycle (fig. 1) by a UV sexual system (Müller 1975; Ahmed et al. 2014). Meiosis occurs during the sporophyte generation, producing meiospores, which develop into either male or female gametophytes. The gametophyte generation produces either male or female gametes, depending on its sex, in sexual structures called plurilocular gametangia.

Morphometric analysis showed that male gametophytes were significantly smaller than female gametophytes at fertility but that they produced significantly more reproductive structures (plurilocular gametangia) despite their smaller size (fig. 2*A*, Student's *t*-test, $P < 0.0001$). Consequently, male gametophytes presumably produce more gametes than females, because they produce a larger number of plurilocular gametangia per individual.

Ectocarpus gametes have been described as being morphologically isogamous and physiologically anisogamous (Schmid et al. 1994). The physiological anisogamy refers to the behavior of the two types of gamete during the fertilization process. The female gametes settle rapidly after release from the plurilocular gametangia, retract their flagella, and then produce a pheromone to attract male gametes. Male gametes swim for longer and are attracted to the immobile female gametes by the pheromone. We used flow cytometry to precisely measure male and female gamete size in three different species of *Ectocarpus*. This analysis, based on measurements of more than 1,000 gametes, showed that male gametes not only exhibit physiological and behavioral differences

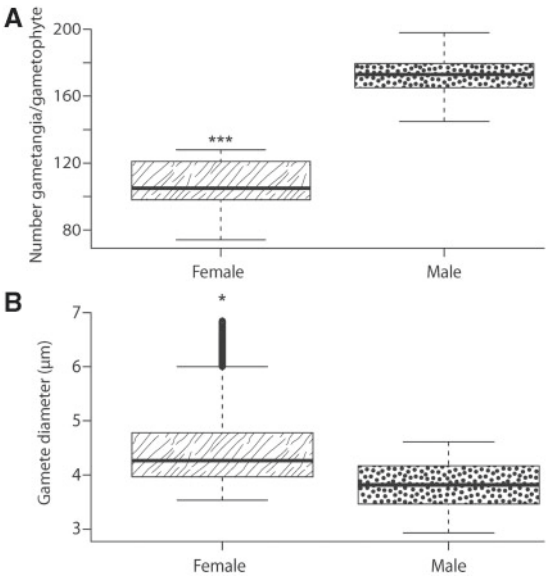


FIG. 2. Sexual dimorphism in *Ectocarpus* gametophytes. (*A*) Number of reproductive structures (plurilocular gametangia) per female ($n = 6$) and male ($n = 8$) gametophyte. Males produced significantly more reproductive structures (Student's *t*-test, $P < 0.0001$). Error bars show standard errors. The number of plurilocular gametangia for each female gametophyte was 128, 109, 74, 121, 101, 98 and for each male gametophyte 176, 145, 198, 178, 169, 170, 181, 161. (*B*) Mean diameters (µm) of female ($n = 5{,}668$) and male ($n = 5{,}619$) gametes. Female gametes (mean diameter 4.46 µm) were significantly larger (Mann–Whitney *U* test, $P < 0.0001$) than male gametes (mean diameter 3.83 µm). Error bars show standard errors. Mean gamete sizes for male and female individuals of other *Ectocarpus* species are provided in supplementary figure S5, Supplementary Material online.

compared with female gametes but they are also slightly, but significantly, smaller (fig. 2*B*, Mann–Whitney *U* test, $P < 0.0001$).

Taken together, these analyses identified sexual dimorphisms at both the gametophyte and gamete stages that had not been previously described. *Ectocarpus* therefore clearly exhibits sexual dimorphism, but the differences between males and females are subtle.

Analysis of Gene Expression during the Development of the Sexual Generation, the Gametophyte

Gene expression patterns during sexual differentiation were measured by deep sequencing (RNA-seq) of cDNA from haploid male and female gametophytes of *Ectocarpus* at two different sexual developmental stages: In juvenile immature gametophytes before the formation of the sexual structures (approximately 10 days after meiospore settlement) and at sexual maturity, when sexual structures were visible (fig. 1). Transcript abundances, measured as RPKM (reads per kilobase per million mapped sequence reads), were strongly correlated between biological replicates of each sex and life cycle stage, with *r* ranging from 0.91 to 0.99 ($P < 2\,e^{-16}$).

Counts of expressed genes (RPKM $>$ 1) identified 13,102 and 12,660 genes that were expressed at the immature stage (male and female, respectively) and 13,941 and 13,663 genes that were expressed at maturity (male and female, respectively). This indicates that about 88% of the protein-coding genes in the genome are transcribed during the gametophyte generation (supplementary fig. S1, Supplementary Material online).

Sex-Biased Gene Expression

Fewer than 12% of *Ectocarpus* genes showed sex-biased expression during the gametophyte generation (including both immature and fertile stages). This is considerably less than the numbers identified in previously characterized systems with more marked morphological sexual dimorphism such as *Drosophila* (e.g., Jiang and Machado 2009) and birds (Pointer et al. 2013) but coherent with the low level of morphological sexual dimorphism in *Ectocarpus*.

Unexpectedly, the number of genes that were differentially transcribed between males and females was higher during the immature gametophyte stage than at gametophyte fertility (fig. 3*A* and *B*). Male-biased genes were more numerous than female-biased genes at both developmental stages, although the numbers for the most strongly differential genes (fold change [FC] $>$ 10) were comparable for the two sexes (fig. 3*A* and *B* and supplementary table S1, Supplementary Material online). The majority of the sex-biased genes showed significant sex-biased expression in only one of the two developmental stages analyzed; only 12% of the male- and 3% of the female-biased genes were differentially expressed in both immature and fertile gametophytes (supplementary fig. S2, Supplementary Material online). Moreover, 3% of the genes that showed male-biased expression in the immature gametophytes were female-specific at maturity. Transitions from female-biased to male-biased were not detected. As females produce fewer plurilocular gametangia than males, we cannot exclude that differences in tissue complexity between male and female fertile gametophytes explain, at least in part, the slight excess of male-biased to female-biased genes (supplementary figs. S1 and S2, Supplementary Material online). Note, however, that comparison of immature gametophytes (where reproductive structures are absent) also identified a slight excess of male-biased over female-biased genes.

To examine the relationship between degree of sex-biased expression and transcript abundance (expression level), the sex-biased genes were grouped according to the FC difference between male and female samples and mean expression level in males and in females plotted for each group (fig. 3*C*). This analysis indicated that when genes exhibited a high degree of female-biased expression, this was predominantly due to downregulation of these genes in males. This was observed at both immature and fertile gametophyte stages. The results obtained for male-biased gene were more complex. In immature gametophytes, the situation was similar to that observed for the female-biased genes in that a high degree of male-biased expression appeared to be correlated with downregulation in females. In contrast, in mature gametophytes, when genes exhibited a high degree of male-biased expression this appeared to be due to a combination of both decreased expression in females and upregulation in males. We also noted

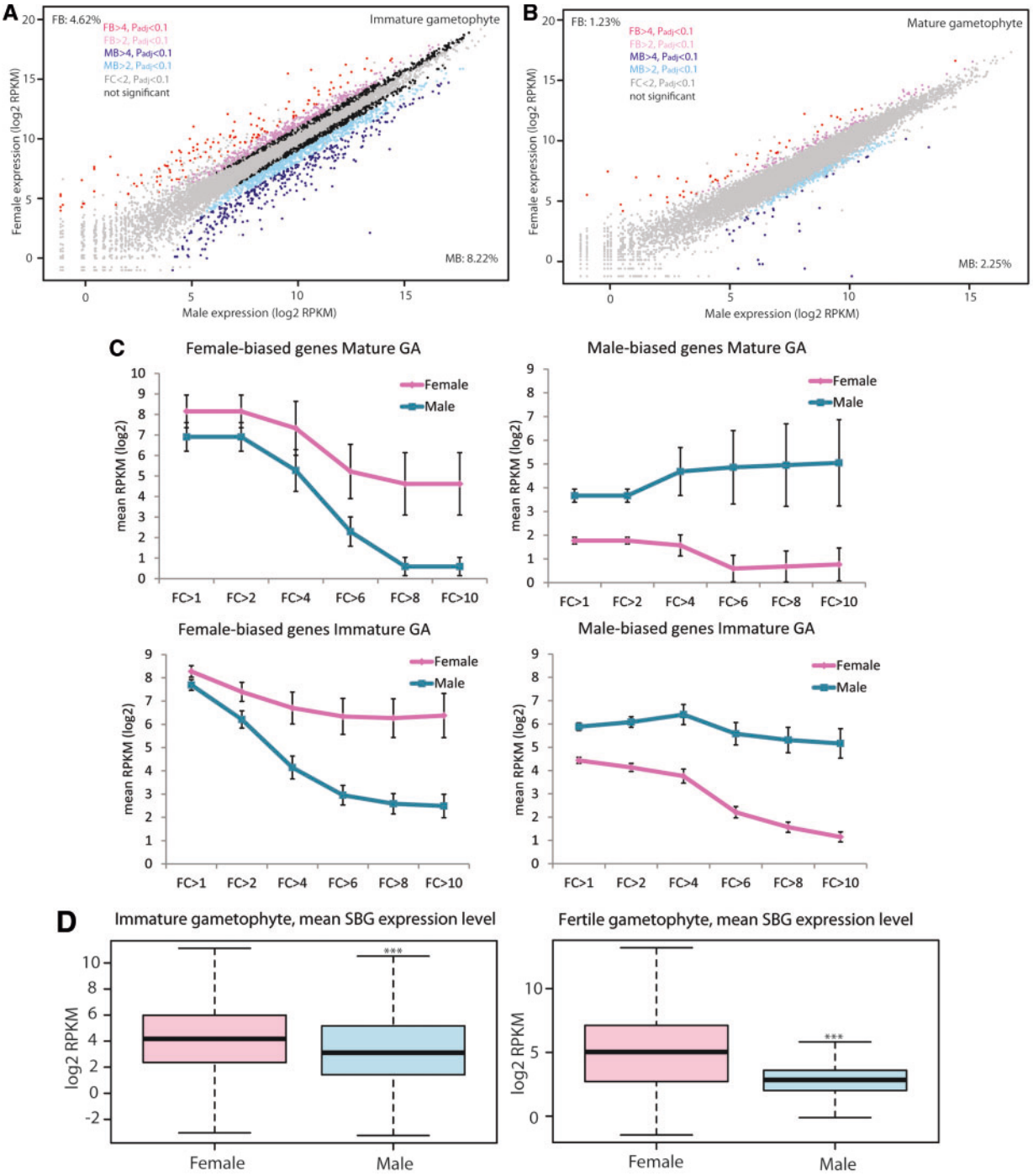


Fig. 3. Sex-biased gene expression. (*A*) Comparison of gene expression levels in male and female immature gametophytes. (*B*) Comparison of gene expression levels in male and female mature gametophytes. Colored dots indicate genes that exhibited significantly different levels of transcript abundance (sex-biased genes). Percentages in each panel indicate genes that were at least 2-fold female-biased (FB; upper left) and male-biased (MB; lower right). FC (fold change); *P*-adjusted (P_{adj}). Unbiased genes were defined as $P_{adj} > 0.1$ or less than 2-fold difference between the sexes. See also table 1. (*C*) Mean gene expression levels (RPKM) at several degrees of sex-bias (from FC > 1 to FC > 10) for female- (pink) and male-biased (blue) genes in fertile and immature gametophytes. Genes located in the SDR were excluded from this analysis. Error bars represent standard errors. (*D*) Boxplot showing the mean expression levels (RPKM) of female- and male-biased genes for immature and fertile gametophytes.

that, on average, female-biased genes were expressed at significantly higher levels than male-biased genes in both fertile and immature gametophytes (Mann–Whitney *U* test, $P < 2\,e^{-16}$) (fig. 3*D*).

Breadth of Expression of Sex-Biased Genes

The breadth of expression of a gene, that is, the extent to which its expression is limited to specific tissues or developmental stages, is a key determinant of its speed of evolution

Table 1. Relative Gene Expression for Male and Female Gametophytes.

		No. Genes	% of Expressed Genes
Immature gametophytes			
Female-biased ($P_{adj} < 0.1$)	**FC > 2**	**585**	**4.62%**
	FC > 4	**131**	**1.03%**
	FC > 10	**68**	**0.54%**
	Total expressed genes (RPKM > 1)	**12,661**	
Male-biased ($P_{adj} < 0.1$)	**FC > 2**	**1,077**	**8.22%**
	FC > 4	**295**	**2.25%**
	FC > 10	**78**	**0.60%**
	Total expressed genes (RPKM > 1)	**13,102**	
Fertile gametophytes			
Female-biased ($P_{adj} < 0.1$)	**FC > 2**	**168**	**1.23%**
	FC > 4	**61**	**0.45%**
	FC > 10	**29**	**0.21%**
	Total expressed genes (RPKM > 1)	**13,660**	
Male-biased ($P_{adj} < 0.1$)	**FC > 2**	**314**	**2.25%**
	FC > 4	**54**	**0.39%**
	FC > 10	**32**	**0.23%**
	Total expressed genes (RPKM > 1)	**13,937**	

NOTE.—Categories of immature or fertile gametophyte sex-biased genes with different levels of FC between the two sexes indicated both as number of genes (N. genes) and as a percentage of the total number of genes expressed (% of expressed genes) in the immature or fertile gametophyte of the corresponding sex.

(Duret and Mouchiroud 2000; Zhang et al. 2004). In the moss *Funaria hygrometrica*, which also has a haploid–diploid life cycle, the effect of breadth of expression was shown to be stronger than the masking effect associated with expression during the diploid phase (Szovenyi et al. 2013). In organisms with haploid–diploid life cycles, the breadth of expression of sex-biased genes is restricted because they tend to be preferentially expressed during the haploid phase (sexuality is only expressed during this phase of the life cycle). This restricted pattern of expression is expected to have a significant effect on their evolutionary rates.

When determining the breadth of expression of *Ectocarpus* genes, we integrated both spatial (tissue) and temporal (developmental and/or life cycle stage) information to obtain meaningful estimates because this species exhibits only a limited level of tissue differentiation during development. We determined the breadth of expression of the sex-biased genes using the specificity index (τ) (see Materials and Methods) and gene expression data collected both for different tissues (upright filaments vs. prostrate tissues during the sporophyte generation; fig. 1) and for different stages of the life cycle (parthenosporophyte, immature and fertile gametophyte and gamete stages; fig. 1). Male and female sex-biased genes had significantly higher τ values compared with unbiased genes, indicating that the former have a greater tendency to be expressed specifically in particular tissues or stages of the life cycle. However, no difference in breadth of expression was observed when the male- and female-biased gene sets were compared with each other (fig. 4). Note that the decrease in the breadth of expression of SBGs was not solely due to their sex-biased pattern of expression; when τ was calculated with a data set in which the male and female samples had been pooled, the male and female SBGs still showed a significantly lower breadth of expression than unbiased genes (Kruskal–Wallis test, $P < 10^{-4}$).

Functional Analysis of Sex-Biased Genes

An analysis of gene ontology (GO) terms associated with the sex-biased genes was carried out using BLAST2GO (Conesa and Gotz 2008) to search for enrichment in particular functional groups and to relate gene function to phenotypic sexual dimorphisms. Significant enrichment of specific GO categories was only detected for fertile male gametophyte and immature female gametophyte sex-biased genes. The set of male-biased genes in mature gametophytes was enriched for "microtubule" and "calcium binding-related" processes. These genes may be involved in the production of flagellated gametes inside plurilocular gametangia. Note that the same GO categories were enriched in the set of sex-biased genes expressed in male gametes identified by Lipinska et al. (2013). The set of female-biased genes in juvenile gametophytes was enriched for "photosynthesis" GO terms, consistent with the more extensive growth phase in the female gametophyte.

A test was also carried out to identify GO terms enriched in the expressed gene sets of the immature compared with the fertile developmental stage of the gametophyte, irrespective of sex. Genes involved in posttranslational regulation of gene expression, cellular component biogenesis, and photosynthesis were significantly enriched in immature compared with fertile gametophytes (FDR < 5%), whereas genes predicted to be involved in signaling, microtubule-based processes, and energy metabolism were significantly enriched in mature compared with immature gametophytes (FDR < 5%) (supplementary table S2, Supplementary Material online). The enriched gene GO terms were coherent overall with the transition from vegetative growth to reproductive function, particularly the production of flagellated gametes, between these two stages of development.

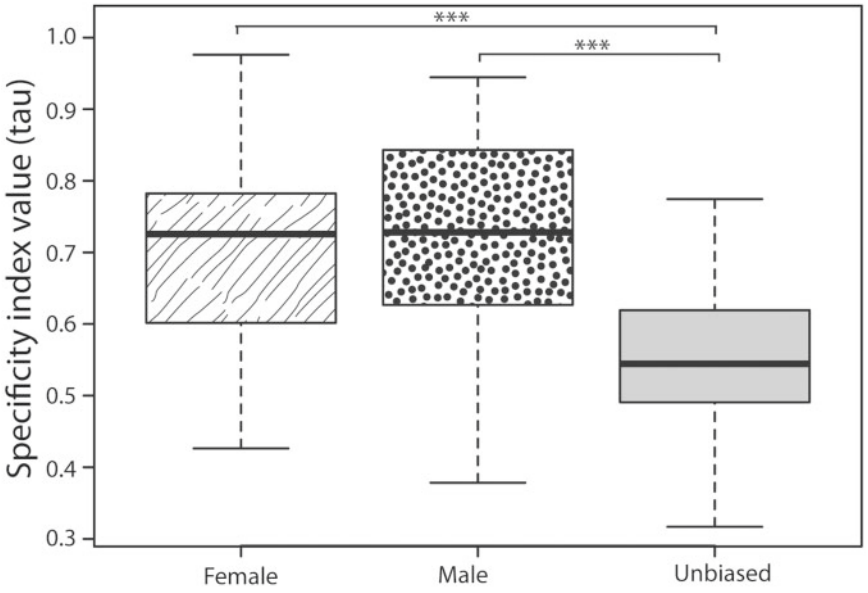


Fig. 4. Breadth of expression of the sex-biased genes as determined using the specificity index. Comparison of specificity index values (τ) for unbiased and for male- and female-biased genes. Male- and female-biased genes had significantly larger specificity index values (i.e., lower breadth of expression) compared with unbiased genes (Kruskal–Wallis test, $P < 10^{-5}$).

Genomic Locations of Sex-Biased Genes

An analysis of the genomic distribution of sex-biased genes expressed in fertile gametophytes found that the PAR region of the sex chromosome was enriched in female-biased genes expressed at this stage compared with the rest of the genome (Chi-squared test, $P < 0.01$) (supplementary fig. S3, Supplementary Material online). Moreover, when RPKM values were used to determine the ratios of transcript abundances in fertile female gametophytes compared with fertile male gametophytes for all the PAR genes, a significant bias toward expression in the female was detected, compared with all the autosomal genes (Kruskal–Wallis, $P < 0.001$) (fig. 5). These tendencies were not observed for sex-biased genes expressed in immature gametophytes. These observations suggest that the PAR and the autosomes are not evolving under the same selection pressures during the fertile gametophyte stage of the life cycle.

Evidence of a Role for Gene Duplication in Resolving Sexual Antagonism

Gene duplication is thought to have played a significant role in the evolution of sex-biased gene expression in *Drosophila* (Connallon and Clark 2011; Wyman et al. 2012). Duplication of a gene can release one or both of the duplicated products from selective constraints allowing the evolution of modified patterns of expression or of new gene functions. Gene duplication therefore represents a potential means to resolve sexual antagonism. The simplest mechanism would be the generation, after duplication, of one male- and one female-biased gene with male- and female-optimized functions, respectively. Other alternatives are possible, however. For example, it may be sufficient for only one member of a duplicated pair to evolve sex-specific functions to resolve a

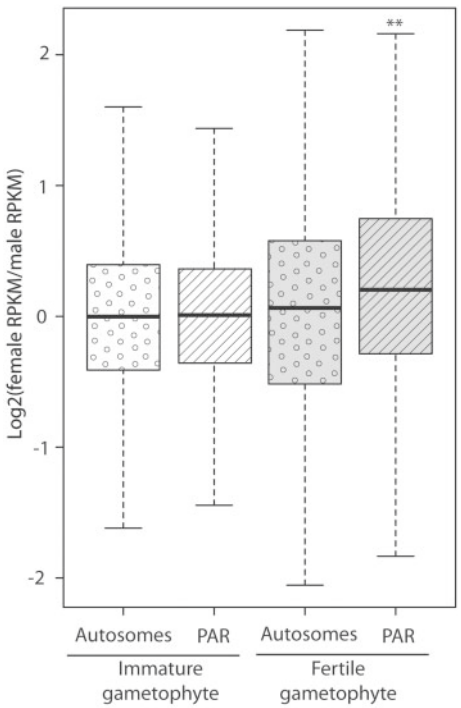


Fig. 5. Ratios of female-to-male expression level in immature and fertile gametophytes for genes on autosomes and genes on the PAR. The figure shows log2 of female/male RPKM ratios for autosomal and PAR genes during the immature and fertile gametophyte stages. Outliers were removed from the plot.

sexual antagonism. In such cases, gene duplication could help resolve sexual conflict for genes with ontogenetic or pleiotropic constraints by allowing one of the duplicated paralogs to evolve sex-biased expression whereas other maintains a general, sex-independent function (Gallach and Betran 2011;

Wyman et al. 2012). It is also possible that duplication of a gene that is already sex-biased may allow one of the duplicates to evolve an even stronger sex-biased function (Wyman et al. 2012).

Genome-wide analysis detected a total of 879 duplicated gene pairs in *Ectocarpus*. Of these, 174 pairs included at least one sex-biased gene. Only 3 of these 174 pairs included both a male-biased and a female-biased gene. These three duplicated gene pairs were autosomal and sex-biased expression was detected during the immature gametophyte stage. Comparisons with sequence data sets for other Ectocarpales species identified orthologs for only one of the genes from these three autosomal gene pairs (Esi0002_0006) but this locus did not show any signatures of positive selection. The other sex-biased, duplicated gene pairs included 143 pairs in which only one member of the pair exhibited sex-biased expression and 28 pairs where both members exhibited sex-biased expression, but in the same sex. The 143 duplicated gene pairs in which only one member exhibited sex-biased expression potentially correspond to events where gene duplication has released one member of the gene pair from selective constraints allowing it to evolve a sex-specific function. This hypothesis is supported by the fact that the specificity index (τ) values for the non-sex-biased members of these pairs are significantly lower than those of the sex-biased members (Kruskal–Wallis test with Dunn's posttest, $P < 10\,e^{-8}$) and are not significantly different from values for randomly selected single copy unbiased genes (fig. 6*A* and *B*).

No evidence has been found for whole genome duplication events having occurred in the lineage leading to *Ectocarpus* (Cock, Sterck, et al. 2010), suggesting that the 879 duplicated gene pairs in the genome of this species arose as a result of small-scale duplication events. When the proportion of the genome corresponding to sex-biased genes is taken into account (1,947 of 16,262 genes), duplicated gene pairs containing at least one sex-biased gene are overrepresented in the total set of 879 duplicated gene pairs (Chi-squared test, $P = 1.5\,e^{-12}$). This overrepresentation was also detected if only male-biased (Chi-squared test, $P = 8.77\,e^{-6}$) or only female-biased genes (Chi-squared test, $P = 2.47\,e^{-5}$) were considered. The results of these tests suggest that the resolution of sexual conflict was one of the forces driving gene duplication in this genome and support a role for gene duplication in the generation of sex-biased genes in this species.

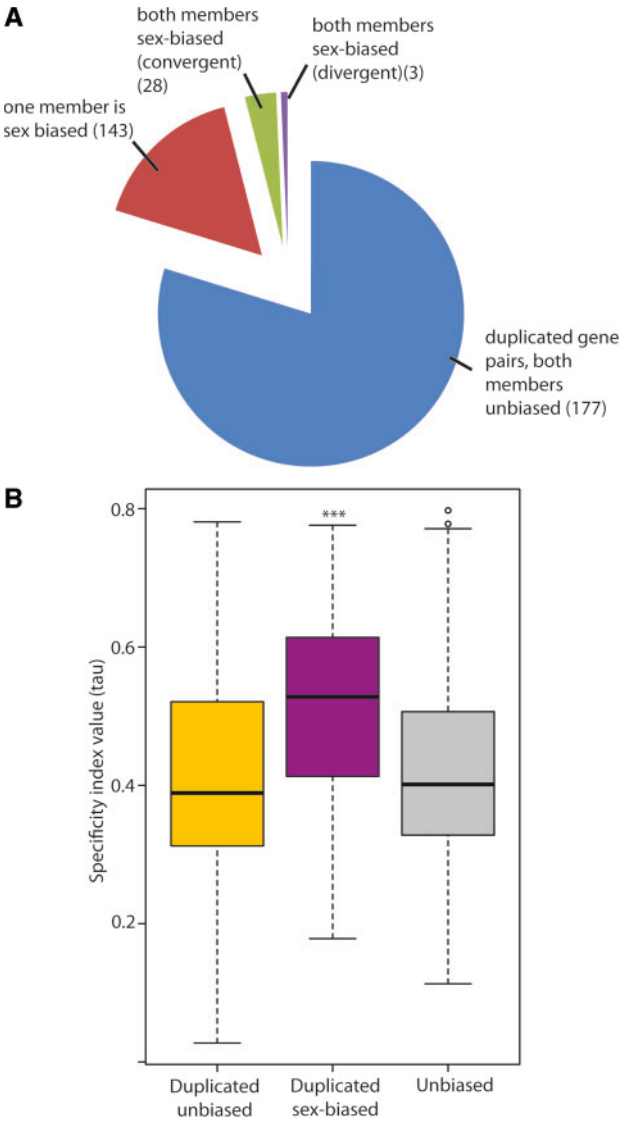


Fig. 6. Duplicated sex-biased genes in *Ectocarpus*. (*A*) Distribution of sex-biased genes among the duplicated gene pairs. (*B*) The sex-biased members (Duplicated sex-biased) of the 143 duplicated gene pairs that include one sex-biased and one unbiased member have a narrower breadth of expression than the unbiased members of these pairs (Duplicated unbiased). A random sample of unbiased single copy genes (Unbiased) is included for comparison. Comparison of breadth of expression is presented using the specificity index (τ). The median for unbiased members of duplicated pairs was significantly lower than the median for sex-biased paralogs (Kruskal–Wallis test with Dunn's posttest, $P < 10\,e^{-8}$) but was not significantly different from the median for single copy unbiased genes.

Sex-Biased Genes Are Evolving More Rapidly

To test for differences in rates of evolutionary divergence between different categories of sex-biased and unbiased genes, we calculated levels of nonsynonymous (d*N*) and synonymous (d*S*) substitution using pairwise comparisons with orthologs from the sister species *Ectocarpus fasciculatus*.

The results of this analysis indicated that genes that exhibited sex-biased expression patterns (either male- or female-biased expression) in fertile gametophytes had evolved significantly faster (i.e., had higher d*N*/d*S* values) than had unbiased genes (Mann–Whitney *U* test, $P < 0.01$). A similar, but weaker, pattern was observed for genes that were male-biased in immature gametophytes (Mann–Whitney *U* test; $P < 0.01$) but the rates of evolution of female-biased genes identified at this developmental stage were not significantly different from those of unbiased genes (fig. 7*A*). Therefore, although the evolution rates of male and female sex-biased genes were similar overall, differences were detected when the developmental stage at which the genes were expressed was taken into account. These differences suggest not only that the average selection pressure

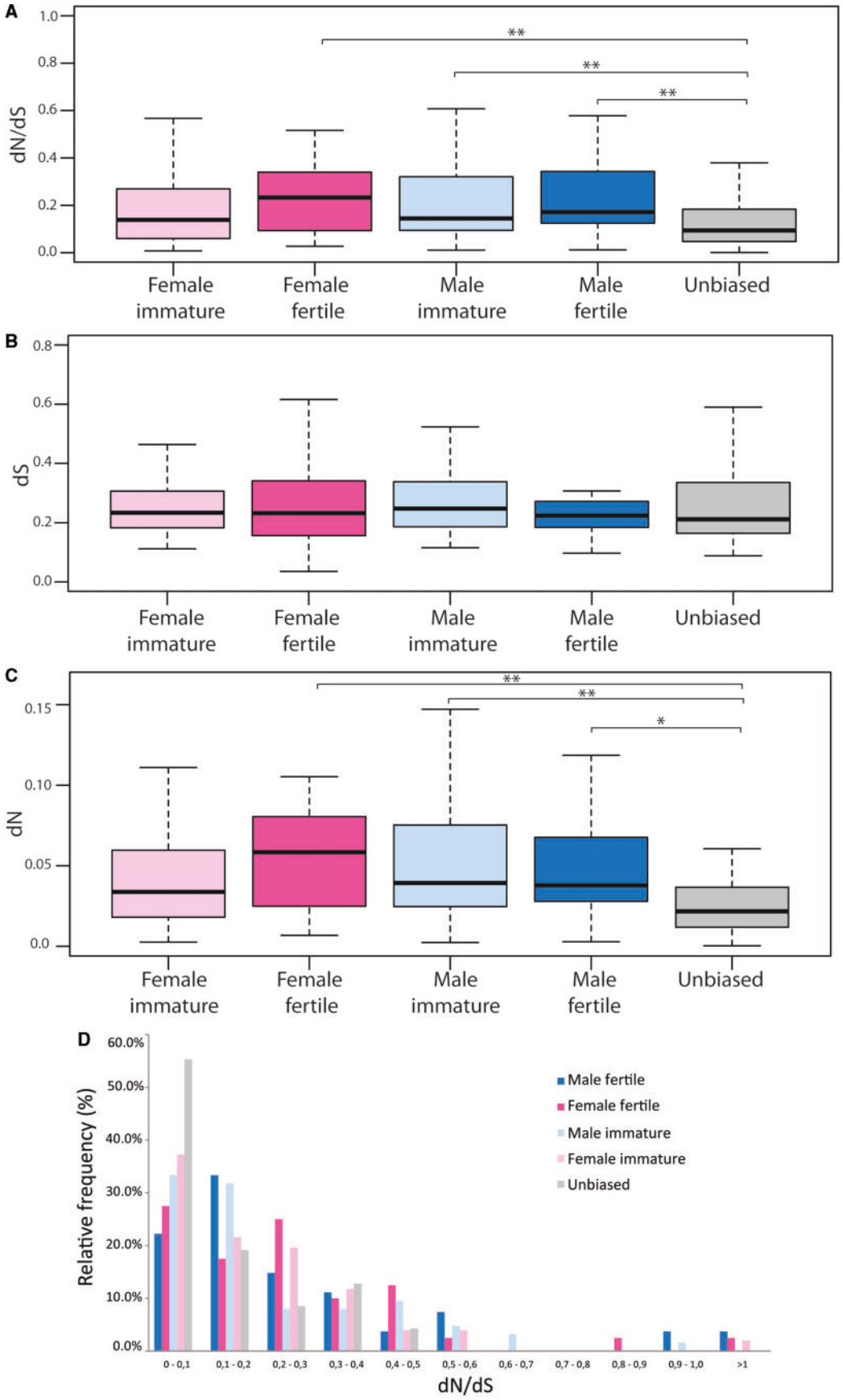


Fig. 7. Rates of evolution of female-biased, male-biased, and unbiased genes. Pairwise d*N*, dS, and d*N*/dS ratios were calculated by comparing orthologous gene sequences from *Ectocarpus* sp. lineage 1c Peru and *Ectocarpus fasciculatus*. (*A*) Ratio of nonsynonymous to synonymous substitutions (d*N*/dS). (*B*) and (*C*) Nonsynonymous substitutions (d*N*) and synonymous substitutions (dS). (*D*) Frequency of classes of d*N*/dS ratio in unbiased genes

(continued)

may vary during development but also that there may be some asymmetry in the evolution rates of male- and female-biased genes that are expressed at particular developmental stages. Concerning this latter point, however, it is possible that the stage at which the comparison was carried out is not directly comparable in males and females because the immature females delay reproduction in order to prolong growth. The comparison is therefore between a stage in males where there may already have been a cryptic transition toward the reproductive phase, as indicated by the greater overlap between the male-biased gene sets identified in immature and fertile individuals, and a stage in females which is equivalent in terms of timing but which corresponds to a continuation of the prereproductive growth phase.

The elevated d*N*/d*S* values for sex-biased compared with unbiased genes were due to significantly higher levels of nonsynonymous substitution (Mann–Whitney *U* test, $P < 0.05$) and not to a reduction in the synonymous substitution rate (fig. 7*B* and *C*). Analysis of the distribution of d*N*/d*S* values indicated that the different groups of sex-biased genes (i.e., male- or female-biased, expressed in immature or fertile gametophyte) tended to be enriched in genes with high d*N*/d*S* values, including values of 1 or more, and to contain fewer genes under strong selective constraint (d*N*/d*S* < 0.1) compared with the group of unbiased genes (fig. 7*D*). No correlation was detected between the degree of sex-bias (FC calculated by DESeq) and the rate of evolution (d*N*/d*S*) of the tested genes (Spearman's $\rho = 0.166$, $P = 0.0516$).

Analysis of specificity index (τ) values indicated that the rates of evolution of the sex-biased genes were only weakly correlated with breadth of expression (Spearman's $\rho = 0.1395$, $P = 0.0229$). This suggests that the effect of sex-biased expression on evolution rate was not solely an indirect effect of restricting gene expression patterns.

Expression bias in sexual tissues has been associated with optimal codon usage, a feature that promotes efficient translation (Duret 2000; Duret and Mouchiroud 2000). For instance, optimal codons occur less frequently in male-biased than in female-biased sexual genes in *Drosophila* (Hambuch and Parsch 2005), suggesting that adaptive protein evolution has modified selection on codon usage. Calculations of the Effective Number of Codons (ENC) and the Codon Adaptation Index (CAI) indicated that selection to maintain codon usage bias in *Ectocarpus* sex-biased genes is globally preserved (supplementary fig. S4*A* and *B*, Supplementary Material online).

As expected, codon usage bias was strongly correlated with the level of gene expression in *Ectocarpus* (CAI vs. $\log_2$RPKM, Spearman's $\rho = 0.623$, $P = 3.76\,e^{-06}$). A slight decrease in CAI was observed in female-biased compared with unbiased genes (Mann–Whitney *U* test, $P = 0.02$) but there was no significant difference in codon usage parameters (CAI and ENC) either between the male-biased genes and unbiased genes or between male and female sex-biased genes.

Evidence for Positive Selection of Sex-Biased Genes

To assess whether differences in divergence rates were due to increased positive selection or relaxed purifying selection, we used sequence data from several Ectocarpales species (supplementary table S3, Supplementary Material online) to estimate direction of selection. We tested 137 sex-biased genes (65 female-biased and 72 male-biased; including 12 genes with d*N*/d*S* > 0.5) and 137 randomly selected unbiased genes using the paired nested site models (M1a, M2a; M7, M8) implemented in PAML4 (CODEML) (Yang 2007). The second model in each pair (M2a and M8) is derived from the first by allowing variable d*N*/d*S* ratios between sites to be greater than 1, making it possible to detect positive selection at critical amino acid residues. This analysis detected evidence of positive selection for 5 of the 12 sex-biased genes with d*N*/d*S* values of greater than 0.5, including both male- and female-biased genes. Moreover, evidence of positive selection was also found for 12 of the remaining 125 sex-biased genes with lower d*N*/d*S* values based on either one or both pairs of models (M1a–M2a, M7–M8) (supplementary table S4, Supplementary Material online). In contrast, only 5 of the 137 unbiased genes had signatures of adaptive evolution, indicating that the set of sex-biased genes was significantly enriched in genes that were under positive selection (Fisher's exact test, $P = 0.0149$).

Discussion

A Complex Relationship across Sexual Species between the Proportion of the Transcriptome Showing Sex-Biased Expression and the Degree of Sexual Dimorphism

Analyses of sex-biased gene expression in *Drosophila* have shown that a large proportion of the transcriptome is differentially expressed in the two sexes (Ellegren and Parsch 2007; Jiang and Machado 2009; Assis et al. 2012). A similar observation was made for turkeys, where it was further shown that male-biased gene expression is significantly enhanced, across the genome, in dominant compared with subordinate males (Pointer et al. 2013). Given that dominant males exhibit stronger secondary sexual characteristics than subordinates, these studies indicate a correlation between the degree of sex-biased gene expression and the extent of sexual dimorphism. However, there is also evidence that the relationship between the level of sex-biased gene expression and the degree of sexual dimorphism may be more complicated. For example, in *Drosophila* more sex-biased genes were detected during the

Fig. 7. Continued

and male- and female-biased genes expressed in immature and fertile gametophytes. Outliers were removed from the plot. Pairwise statistical significance between the four groups of sex-biased genes on the one hand and the unbiased genes on the other was calculated for panels (*A*)–(*C*), only statistically significant differences are indicated (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$).

juvenile stage than in adults, despite the lower degree of observable sexual dimorphism during the former phase of development (Mank et al. 2010; Perry et al. 2014). Further studies are therefore required to investigate the exact relationship between these two parameters.

Ectocarpus represents an interesting system in this respect because the studies that have been carried out to date have focused on species that exhibit very marked sexual dimorphism. In contrast, we show here that this brown alga exhibits a limited degree of sexual dimorphism, restricted to subtle growth-habit and fertility differences during the gametophyte stage and a small difference in male and female gamete sizes. Accordingly, less than 12% of the genes in the genome were found to be differentially regulated between sexes, supporting the hypothesis that the overall degree of sex-biased gene expression and the level of phenotypic sexual dimorphism are correlated.

Analysis of the expression of *Ectocarpus* sex-biased genes during development revealed a more complex relationship between the expression patterns of these genes and the manifestation of sexually dimorphic traits. As observed with *Drosophila*, more sex-biased genes were detected during the sexually immature stage than in fertile, sexually mature individuals, despite the fact that the former exhibited less marked sexual dimorphism. Similarly, male and female gametes have been shown to exhibit high levels of sex-biased expression despite limited phenotypic sexual dimorphism (Lipinska et al. 2013). Thus, there is evidence in both *Drosophila* and *Ectocarpus* that the correlation between the level of sex-biased gene expression and the level of observed sexual dimorphism breaks down to some extent when the relationship is examined over the course of development. As *Ectocarpus* and *Drosophila* are two phylogenetically distant organisms with very marked differences in their levels of sexual dimorphism, these observations suggest that the lack of correlation between sex-biased gene expression and sexual dimorphism in immature individuals may be a general feature of sexual systems, but further studies on diverse sexual organisms are required to confirm this. One possible reason for this could be that part of the sex-biased gene expression is related to differences at the cellular level that do not have any effect on morphology.

Analysis of predicted gene functions indicated that about 12% of the male-biased genes expressed during the immature stage were also expressed in fertile gametophytes, but there was less overlap between female-biased genes expressed at the two stages (3% of the female-biased genes). This suggests that immature females were principally carrying out processes unrelated to those engaged at maturity, such as filamentous growth for example, whereas reproductive processes were already initiated to some extent in immature males, before any phenotypic change could be detected. Somewhat paradoxically, therefore, one of the roles of sex-biased genes in females may be to suspend reproductive functions to allow more extensive vegetative growth during the juvenile phase.

As far as the mechanism of evolution of the sex-biased genes in *Ectocarpus* is concerned, the set of sex-biased genes in this species is enriched in genes that are members of duplicated pairs indicating that neo- or subfunctionalization following gene duplication is one of the mechanisms through which sex-biased genes evolve in this brown alga. Gene duplication has been proposed to be one of the means of resolving sexually antagonistic conflict in other systems (Connallon and Clark 2011; Gallach and Betran 2011; Wyman et al. 2012).

Symmetrical Evolution Rates of Male- and Female-Biased Genes in *Ectocarpus*

In general, sex-biased genes tend to evolve at faster rates than unbiased genes and this effect is usually significantly more marked for male-biased genes than for female-biased genes (reviewed in Ellegren and Parsch [2007]). The faster evolution rate is thought to be due, at least in part, to positive selection acting on the sex-biased genes, the most likely underlying causes being sexual selection and/or sexual antagonism. The sex-biased genes in *Ectocarpus* also exhibit faster evolution rates than unbiased genes but this system is unusual in that, overall, male- and female-biased genes have evolved at similar rates. There are several possible explanations for this symmetry. The most obvious explanation, which is consistent with the low level of sexual dimorphism in this system, is that male- and female-biased genes are under similar levels of sexual selection. Both male and female gametes are small, motile cells that are produced in large numbers in plurilocular gametangia by male and female gametophytes, respectively. It is not known whether gamete competition occurs during fertilization under natural conditions but, if it does occur, the mechanism involved affords scope for both male and female competitions. Male gametes may compete to find and fertilize the settled female gametes, but the abundant female gametes may compete for optimal niches in which to settle and then compete with each other to attract male gametes through pheromone production. It is therefore quite possible that selection pressures on males and females are very similar in this organism.

Sex-biased genes in *Ectocarpus* are expressed during the haploid phase of the cycle and therefore directly exposed to purifying selection (Kondrashov and Crow 1991; Orr and Otto 1994; Gerstein et al. 2011). Another possible explanation for the symmetric evolution rates of male- and female-biased genes in *Ectocarpus* may be that haploid phase purifying selection is strong enough to mask any effects of sexual selection or sexual antagonism. This seems unlikely, however, as land plants also possess a haploid gametophyte generation and selection-driven evolution suggestive of sexual selection has been detected in this group of organisms (Arunkumar et al. 2013; Gossmann et al. 2014).

Another possible factor affecting evolution rate is breadth of expression pattern, as broadly expressed genes tend to be more constrained and therefore to evolve less rapidly than genes with restricted patterns of expression (Hastings 1996; Duret and Mouchiroud 2000). In *Drosophila* one of the reasons that female-biased genes evolve less quickly than male-biased genes may be that, in general, they tend to have broader patterns of expression (e.g., Meisel 2011; Grath and

Parsch 2012). Our analysis, based on RNA-seq analysis of multiple life cycle stages and tissues, indicated that, in contrast, both male- and female-biased genes in *Ectocarpus* tend to have restricted patterns of expression compared with unbiased genes (fig. 4). This parallel reduction in breadth of expression may be one of the factors underlying the symmetrical accelerated evolution of male- and female-biased genes in this species. However, we noted that there was only a weak positive correlation between expression breadth (τ) and evolutionary rate (d*N*/d*S*), suggesting that other factors have also influenced evolutionary rates.

In summary, therefore, possible explanations for the symmetrical rates of evolution of male- and female-biased genes in *Ectocarpus* include limited sexual selection, impacting similarly males and females, due to a low level of sexual dimorphism and comparable levels of breadth of expression pattern.

Sexual Selection Is One of the Forces that Drives the Evolution of Male- And Female-Biased Genes in *Ectocarpus*

The mean d*N*/d*S* value for sex-biased genes in *Ectocarpus* was more than twice as high as that of unbiased genes. This difference, which was particularly marked for genes expressed in fertile gametophytes, was due to a significantly higher rate of nonsynonymous changes compared with the unbiased genes. A test for adaptive evolution detected evidence for positive selection in a significant proportion of the sex-biased genes with the highest d*N*/d*S* values (>0.5). Similar observations have been made for sperm-specific genes in *Arabidopsis thaliana* (Arunkumar et al. 2013) and for gametophyte-specific genes in the moss *Funaria hygrometrica* (Szovenyi et al. 2013). The evidence that positive selection acts on a considerable number of *Ectocarpus* sex-biased genes indicates that sexual selection may be one of the forces driving their evolution. Note however that positive selection only affects a subset of the *Ectocarpus* sex-biased genes and a significant proportion appear to be under relaxed selection. One important consideration in this respect is that a gene that is expressed in only one sex will experience half as much purifying selection because selection can only act on the gene when it is in the appropriate sex (Barker et al. 2005).

Patterns of Genomic Distribution of Sex-Biased Genes

In XY and ZW systems, the pattern of segregation of the sex chromosomes can have a measurable influence on the distributions of sex-biased genes on this linkage group. For XY systems, for example, X chromosomes spend twice as much time in females as they do in males. Male beneficial mutations can either accumulate or be purged from this chromosome depending on whether they are recessive or dominant (Rice 1984). There is no equivalent to this phenomenon in UV systems because the sex chromosomes function in the haploid generation. However, UV systems may share other features with XY and ZW systems that affect the distribution of sex-biased genes. In particular, even partial linkage to the SDR can be beneficial for genes with sexually antagonistic alleles, allowing alleles to segregate preferentially to the sex for which they are most adaptive (Otto et al. 2011; Jordan and Charlesworth 2012). This is predicted to lead to the accumulation of sexually antagonistic genes in the PAR, which in turn could lead to an accumulation of sex-biased genes in this region because sex-biased expression is one of the possible mechanisms of resolving sexual antagonism. There is some experimental evidence for this mechanism from work on the ZW sexual system of the emu, which has shown that the PARs of the homomorphic sex chromosomes of this species are enriched in male-biased genes (Vicoso et al. 2013). As expected, this effect was most pronounced for genes expressed in older embryos with fully developed gonads.

For UV systems, in the absence of any additional selective pressure favoring genes of one sex or the other, this effect of linkage to the SDR would not be expected to lead to a preferential accumulation of male-biased genes compared with female-biased genes or vice versa, but it might be expected to result in a general excess of sex-biased genes in the PAR. We did not observe any such excess in *Ectocarpus*, the proportion of sex-biased genes in the PAR was not significantly different from the proportion in the autosomes. However, compared with the autosomes, the *Ectocarpus* PAR was found to be significantly enriched in genes that exhibited female-biased expression during the fertile gametophyte stage. One possible explanation for this enrichment in female-biased genes may be a combination of an effect of linkage to the SDR together with stronger selection for female-biased genes during the fertile gametophyte stage.

There is accumulating evidence that gene duplication has played a significant role in the evolution of sex-biased genes in animals (Connallon and Clark 2011; Gallach and Betran 2011; Wyman et al. 2012) and the data presented here indicate that this has also been the case for *Ectocarpus*, suggesting that similar mechanisms may be operating to generate sex-biased genes across diverse eukaryote sexual systems.

Materials and Methods

Biological Material

Ectocarpus strains were cultured at 13 °C in autoclaved natural sea water (NSW) supplemented with half-strength Provasoli solution (PES; Starr and Zeikus 1993) with a light:dark cycle of 12:12 h (20 µmol photons m^{-2} s^{-1}) using daylight-type fluorescent tubes. All manipulations were performed under a laminar flow hood in sterile conditions. Near-isogenic lines, Ec602 female and Ec603 male, were prepared by crossing brothers and sisters for eight generations. This produced male and female strains with essentially identical genetic backgrounds apart from the sex locus. Supplementary table S3, Supplementary Material online, describes the *Ectocarpus* species used in this study. Note that currently only three species are recognized within the genus *Ectocarpus* (*E. siliculosus, E. fasciculatus,* and *E. crouanorium;* Peters et al. 2010) but there is increasing evidence that the taxon *E. siliculosus* represents a complex of several species. As the type specimen for *E. siliculosus* was collected in England, we prefer to refer to the non-European strains related to

E. siliculosus (such as the Peruvian and Greenland strains) as "*Ectocarpus* sp."

Male and female gametophytes of *Scytosiphon lomentaria* were collected at Asari, Japan in March 2012. *Scytosiphon lomentaria* has been described as exhibiting near-isogamy, with the male gametes being slightly smaller than the female gametes (Nagasato and Motomura 2002). The male and female gametophytes are morphologically similar and no sexual dimorphism has been described at this stage. *Scytosiphon lomentaria* was cultured in NSW with full strength PES. Two different light conditions were required to complete the life cycle. Short-day conditions, with a light:dark cycle of 10:14 h (20 µmol photons m^{-2} s^{-1}), were used to produce unilocular sporangia from a diploid sporophyte. After a month approximately 100 young gametophytes were isolated. The gametophytes were then subjected to long-day conditions with a cycle of 14:10 h to induce gametophyte maturation. Gametophytes became fertile after approximately 4 weeks and were frozen in liquid nitrogen. Each individual was sexed by crossing with male and female tester lines.

Measurement of Gamete Size

Male and female gamete size was measured in three different *Ectocarpus* species (see Stache-Crain et al. 1997 for a description of the lineage structure of the genus *Ectocarpus*): Isogenic male and female strains of *Ectocarpus* sp. lineage 1c Peru (Ec602 and Ec603), *E. siliculosus* lineage 1a Naples, and *Ectocarpus* sp. lineage 4 New Zealand. Synchronous release of gametes from 3- to 4-week-old cultures was induced by transferring ten gametophytes to a humid chamber in the dark for approximately 14 h at 13 °C followed by the addition of fresh PES-supplemented NSW medium under strong light irradiation. Gametes were concentrated by phototaxis using unidirectional light, and collected in Eppendorf tubes. Gamete size was measure by impedance-based flow cytometry (Cell Lab QuantaTM SC MPL, Beckman Coulter). Values of gamete size shown represent the mean ± SE of each gamete and measurements were taken for at least three biological replicates. A *t*-test ($\alpha = 5\%$) was performed using GraphPad Prism software to compare female and male gamete size.

Measurement of Gametophyte Size and Fertility

For the analysis of gametophyte habit and fertility, male and female near-isogenic strains (Ec602 and Ec603; supplementary table S3, Supplementary Material online) were placed in culture conditions as described above at constant density (ten individuals per 140-mm Petri dish). In each Petri dish, all ten gametophytes grew synchronously and attained approximately the same size. The gametophytes attained sexual maturity (production of plurilocular gametangia) after 3–4 weeks in culture. The number of plurilocular gametangia, each containing approximately 300 gametes, was counted under an inverted microscope for one individual randomly taken from each Petri dish. It was not possible to accurately weigh a single gametophyte, so ten gametophytes were pooled, weighed and the individual weight estimated by dividing by 10. Results shown correspond to the mean ± SE for six biological replicates for Ec602 and eight biological replicates for Ec603. Significant differences were tested using a corrected *t*-test with R software ($\alpha = 5\%$).

Generation of RNA-seq Data

RNA-seq analysis was carried out to compare the relative abundances of gene transcripts at different developmental stages of the life cycle (fig. 1). For the gametophyte generation, synchronous cultures of gametophytes of the near-isogenic male and female lines Ec603 and Ec602 were grown under standard conditions and frozen at early stages of development (about 10 days after meiospore release) and at fertility (presence of plurilocular gametangia). For each stage, total RNA was extracted from 2 bulks of 400 male individuals and 2 bulks of 400 female individuals (two biological replicates for each sex) using the Qiagen Mini kit (http://www.qiagen.com) as previously described (Coelho et al. 2012). Two biological replicates of basal parthenosporophyte filaments from strain Ec32 (which carries the V chromosome) were frozen in liquid nitrogen 10 days after settlement of gametes. Similarly, two biological replicates of upright filament tissue were isolated 15 days after settlement of gametes.

Two biological replicates for each sex of *S. lomentaria* were prepared by pooling between 8 and 12 individuals per sample. RNA from male and female pools was extracted using the protocol described by Apt et al. (1995). RNA quality and quantity was assessed using an Agilent 2100 bioanalyzer, associated with an RNA 6000 Nano kit.

RNA Sequencing

For each replicate, the RNA was quantified and cDNA was synthesized using an oligo-dT primer. The cDNA was fragmented, cloned, and sequenced by Fasteris (CH-1228 Plan-les-Ouates, Switzerland) using an Illumina Hi-seq 2000 set to generate 100-bp single-end reads. Supplementary table S5, Supplementary Material online, shows the statistics for the sequencing and mapping. Data quality was assessed using the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html) and the reads were trimmed and filtered using a quality threshold of 25 (base calling) and a minimal size of 60 bp. Only reads in which more than 75% of the nucleotides had a minimal quality threshold of 20 were retained.

Filtered reads were mapped to the *Ectocarpus* sp. genome (Cock, Coelho, et al. 2010) (available at ORCAE; Sterck et al. 2012) using TopHat2 with the Bowtie2 aligner (Kim et al. 2013). More than 90% of the sequencing reads for each library could be mapped to the genome. The mapped sequencing data were then processed with HTSeq (Anders et al. 2014) to obtain counts for sequencing reads mapped to exons. Expression values were represented as RPKM and a filter of RPKM > 1 was applied to remove noise and genes with very low expression levels. This resulted in a total of 14,302 genes with expression values above the threshold (supplementary fig. S1, Supplementary Material online). The SRR accession

numbers for the raw sequence data are SRR1660827, SRR1660828, SRR1660829, and SRR1660830.

Differential expression analysis was performed with the DESeq package (Bioconductor) (Anders and Huber 2010) using an adjusted *P*-value cut-off of 0.1 and a minimal fold-change of 2 (supplementary fig. S2, Supplementary Material online). Full lists of sex-biased genes can be found in supplementary table S1, Supplementary Material online.

The sex-biased genes were also analyzed for the presence of duplicated genes to determine whether duplications might have arisen to resolve sexual conflict. Duplicated gene pairs were detected as described in Cock, Sterck, et al. (2010). Briefly, each *Ectocarpus* protein was compared with the entire set of *Ectocarpus* proteins using BLASTp and duplicate genes were defined as two sequences from different loci with a maximal *E* value of $e.10^{-4}$. The clustering analysis was performed using the MCL algorithm (Markov Cluster Algorithm; Li et al. 2003) with the inflation value fixed at 3.0.

Measurement of Synonymous and Nonsynonymous Mutation Rates

To estimate rates of evolution of sex-biased gene sequences, we searched *E. fasciculatus* transcriptome data (Gachon CM, unpublished data) for orthologs of sex-biased and unbiased control genes (the latter was a random subset of 47 genes without differences in expression levels between males and females) by retaining best reciprocal BLASTn matches with a minimum *e* value of $10\,e^{-10}$. The orthology of genes derived from duplications in *Ectocarpus* sp. was further evaluated by calculation of phylogenetic trees using *E. siliculosus* and *E. fasciculatus* sequences, along with *S. lomentaria* sequences as outgroups. MEGA6 (Larkin et al. 2007; Tamura et al. 2013) was used for maximum-likelihood analyses and branch support was assessed with by bootstrapping (1,000 replicates).

Putative orthologs were aligned using ClustalW implemented in MEGA6 (Larkin et al. 2007; Tamura et al. 2013) and manually curated. Sequences that produced a gapless alignment that exceeded 100 bp were retained for pairwise dN/dS (ω) analysis using Phylogenetic Analysis by Maximum Likelihood (PAML, CODEML, F3x4 model, runmode = -2) implemented in the PAL2NAL suit (Suyama et al. 2006; Yang 2007) Genes with saturated synonymous substitution values ($dS > 1$) and genes located in the SDR were excluded from the analysis.

The ENC and the CAI were calculated for all sex-biased and unbiased genes in this study using CAIcal server (http://genomes.urv.es/CAIcal/) (Puigbo et al. 2008).

Positive Selection Analysis

We used transcriptomic and genomic data from four different *Ectocarpus* species and another Ectocarpales species, *S. lomentaria* to detect positive selection (supplementary table S3, Supplementary Material online). *Ectocarpus* sp. lineage 1c Greenland, *E. fasciculatus*, and *S. lomentaria* transcriptome data were generated using Illumina HiSeq v3 paired-end technology (2×100 bp) and quality filtered using either the FASTX toolkit or Trimmomatic (http://www.ncbi.nlm.nih.gov/pubmed/24695404) (Gachon CM, unpublished data). Transcriptome assemblies were generated using the Trinity de novo assembler (Grabherr et al. 2011) with default parameters and using normalized mode. Transcripts were filtered for isoform percentage (>1) and RPKM (>1). *Ectocarpus siliculosus* lineage 1a genomic data were aligned to the reference genome and consensus sequences of coding regions with at least 10× coverage were recovered using the CLC Assembly Cell (www.clcbio.com).

Orthologs of *Ectocarpus* sp. lineage 1c Peru sex-biased and unbiased genes were identified in *E. siliculosus* lineage 1a, *Ectocarpus* sp. lineage 1c Greenland, *E. fasciculatus*, and *S. lomentaria* by selecting transcripts that could be aligned over at least 100 bp using a best reciprocal BLASTn approach (*E* value cutoff of 10^{-10}). Nucleotide alignments for genes identified from at least four of the five species were made using ClustalW implemented in MEGA6 (Larkin et al. 2007; Tamura et al. 2013) curated manually when necessary and transformed to PAML4 format using perl fasta manipulation scripts (provided by Naoki Takebayashi, University Alaska Fairbanks).

Levels of nonsynonymous (dN) and synonymous (dS) substitution were estimated by the maximum-likelihood method available in the CODEML program (PAML4 package) using the F3x4 model of codon frequencies and a user tree specified according to the phylogeny (Stache-Crain et al. 1997). CODEML paired nested site models (M0, M3; M1a, M2a; M7, M8) (Yang 2000, 2007) of sequence evolution were used and the outputs compared using the likelihood ratio test. Empirical Bayes methods allowed for identification of positively selected sites a posteriori (Yang 2000, 2007).

Breadth of Gene Expression

RNA-seq data corresponding to complete organisms from seven different stages of the life cycle (male and female gametes, parthenosporophytes, immature and fertile male and female gametophytes) and to two different tissue types (basal structures and upright filaments) were used to estimate breadth of gene expression. The gamete transcriptomic data (Lipinska et al. 2013) were converted to RPKM in order to make them comparable with the other libraries. The specificity index (τ) (Yanai et al. 2005) was used as a measure of breadth of expression for each gene, using the following formula

$$\tau = \frac{\sum_{i=1}^{N}(1 - x_i)}{N - 1}.$$

For each gene, we calculated x_i as the expression profile in the given library *i* normalized by the maximal expression value across all analyzed tissues and life cycle stages (*N*). τ index values range from 0 to 1, where 1 corresponds to strong tissue/life cycle stage specificity (low expression breadth).

Analysis of Predicted Gene Functions

InterProScan (Zdobnov and Apweiler 2001) and BLAST2GO (Conesa and Gotz 2008) were used to recover functional annotations for *Ectocarpus* proteins. For BLAST2GO, a Fisher exact test with an FDR corrected *P* value cutoff of 0.05 was used to detect enrichment of specific GO-terms in various groups of sex-biased genes.

Genomic Location of Sex-Biased Genes

A Chi-squared test of observed and expected distribution of sex-biased genes across the *Ectocarpus* linkage groups (Heesch et al. 2010) was used to test whether sex-biased genes were randomly distributed throughout the genome. The expected distribution was calculated with the assumption that the sex-biased genes were randomly distributed and therefore that representation on a particular chromosome should have been proportional to the number of genes on that chromosome. The Chi-squared test was performed in Excel 2010 (Microsoft, Redmond, WA). All other statistical analyses were performed in RStudio (R version 3.0.2).

Supplementary Material

Supplementary figures S1–S5 and tables S1–S5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the Centre National de la Recherche Scientifique, the Agence Nationale de la Recherche (Project SEXSEAWEED and project IDEALG), the University Pierre and Marie Curie Emergence program, the Interreg program France (Channel)-England (project Marinexus), the NERC National Biomolecular Analysis Facility (NBAF-E), and the NERC grant NE/J00460X/1 and NE/L013223/1. The authors thank Lieven Sterck for help with the analysis of duplicated genes, Julie Jaquiery for helpful discussions, D. Marie and D. Voulot for help with the cytometry mesurements, and K. Kogame for the sampling and culture of *Scytosiphon lomentaria*.

References

Ahmed S, Cock JM, Pessia E, Luthringer R, Cormier A, Robuchon M, Sterck L, Peters AF, Dittami SM, Corre E, et al. 2014. A haploid system of sex determination in the brown alga *Ectocarpus* sp. *Curr Biol.* 24:1945–1957.

Albritton SE, Kranz AL, Rao P, Kramer M, Dieterich C, Ercan S. 2014. Sex-biased gene expression and evolution of the X chromosome in nematodes. *Genetics* 197:865–883.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol.* 11:R106.

Anders S, Pyl PT, Huber W. 2014. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2):166–169.

Apt KE, Clendennen SK, Powers DA, Grossman AR. 1995. The gene family encoding the fucoxanthin chlorophyll proteins from the brown alga *Macrocystis pyrifera*. *Mol Gen Genet.* 246:455–464.

Arunkumar KP, Mita K, Nagaraju J. 2009. The silkworm Z chromosome is enriched in testis-specific genes. *Genetics* 182:493–501.

Arunkumar R, Josephs EB, Williamson RJ, Wright SI. 2013. Pollen-specific, but not sperm-specific, genes show stronger purifying selection and higher rates of positive selection than sporophytic genes in *Capsella grandiflora*. *Mol Biol Evol.* 30:2475–2486.

Assis R, Zhou Q, Bachtrog D. 2012. Sex-biased transcriptome evolution in *Drosophila*. *Genome Biol Evol.* 4:1189–1200.

Avila V, Marion de Proce S, Campos JL, Borthwick H, Charlesworth B, Betancourt AJ. 2014. Faster-X effects in two *Drosophila* lineages. *Genome Biol Evol.* 6:2968–2982.

Bachtrog D, Kirkpatrick M, Mank JE, McDaniel SF, Pires JC, Rice W, Valenzuela N. 2011. Are all sex chromosomes created equal? *Trends Genet.* 27:350–357.

Bachtrog D, Toda NR, Lockton S. 2010. Dosage compensation and demasculinization of X chromosomes in *Drosophila*. *Curr Biol.* 20: 1476–1481.

Barker MS, Demuth JP, Wade MJ. 2005. Maternal expression relaxes constraint on innovation of the anterior determinant, bicoid. *PLoS Genet.* 1:e57.

Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh YP, Hahn MW, Nista PM, Jones CD, Kern AD, Dewey CN, et al. 2007. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol.* 5:e310.

Bellott DW, Skaletsky H, Pyntikova T, Mardis ER, Graves T, Kremitzki C, Brown LG, Rozen S, Warren WC, Wilson RK, et al. 2010. Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature* 466:612–616.

Bohne A, Sengstag T, Salzburger W. 2014. Comparative transcriptomics in East African cichlids reveals sex- and species-specific expression and new candidates for sex differentiation in fishes. *Genome Biol Evol.* 6:2567–2585.

Bull J. 1978. Sex chromosomes in haploid dioecy: a unique contrast to Muller's theory for diploid dioecy. *Am Nat.* 112:245–250.

Campos JL, Halligan DL, Haddrill PR, Charlesworth B. 2014. The relation between recombination rate and patterns of molecular evolution and variation in *Drosophila melanogaster*. *Mol Biol Evol.* 31: 1010–1028.

Charlesworth B, Coyne JA, Barton NH.. 1987. The relative rates of evolution of sex chromosomes and autosomes. *American Naturalist* 130: 113–146.

Charlesworth B, Jordan CY, Charlesworth D. 2014. The evolutionary dynamics of sexually antagonistic mutations in pseudoautosomal regions of sex chromosomes. *Evolution* 68:1339–1350.

Cock JM, Coelho SM, Brownlee C, Taylor AR. 2010. The *Ectocarpus* genome sequence: insights into brown algal biology and the evolutionary diversity of the eukaryotes. *New Phytol.* 188: 1–4.

Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, Amoutzias G, Anthouard V, Artiguenave F, Aury J, Badger J, et al. 2010. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature* 465:617–621.

Coelho SM, Scornet D, Rousvoal S, Peters NT, Dartevelle L, Peters AF, Cock JM. 2012. How to cultivate *Ectocarpus*. *Cold Spring Harb Protoc.* 2012:258–261.

Conesa A, Gotz S. 2008. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics.* 2008: 619832.

Connallon T, Clark AG. 2011. The resolution of sexual antagonism by gene duplication. *Genetics* 187:919–937.

Duret L. 2000. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet.* 16:287–289.

Duret L, Mouchiroud D. 2000. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. *Mol Biol Evol.* 17:68–74.

Ellegren H, Parsch J. 2007. The evolution of sex-biased genes and sex-biased gene expression. *Nat Rev Genet.* 8:689–698.

Gallach M, Betran E. 2011. Intralocus sexual conflict resolved through gene duplication. *Trends Ecol Evol.* 26:222–228.

Gerstein AC, Cleathero LA, Mandegar MA, Otto SP. 2011. Haploids adapt faster than diploids across a range of environments. *J Evol Biol.* 24:531–540.

Gossmann TI, Schmid MW, Grossniklaus U, Schmid KJ. 2014. Selection-driven evolution of sex-biased genes is consistent with sexual selection in *Arabidopsis thaliana*. *Mol Biol Evol.* 31:574–583.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 29:644–652.

Grath S, Parsch J. 2012. Rate of amino acid substitution is influenced by the degree and conservation of male-biased transcription over 50 myr of *Drosophila* evolution. *Genome Biol Evol.* 4:346–359.

Haerty W, Jagadeeshan S, Kulathinal RJ, Wong A, Ravi Ram K, Sirot LK, Levesque L, Artieri CG, Wolfner MF, Civetta A, et al. 2007. Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. *Genetics* 177:1321–1335.

Hambuch TM, Parsch J. 2005. Patterns of synonymous codon usage in *Drosophila melanogaster* genes with sex-biased expression. *Genetics* 170:1691–1700.

Hastings KE. 1996. Strong evolutionary conservation of broadly expressed protein isoforms in the troponin I gene family and other vertebrate gene families. *J Mol Evol.* 42:631–640.

Heesch S, Cho GY, Peters AF, Le Corguille G, Falentin C, Boutet G, Coedel S, Jubin C, Samson G, Corre E, et al. 2010. A sequence-tagged genetic map for the brown alga *Ectocarpus siliculosus* provides large-scale assembly of the genome sequence. *New Phytol.* 188: 42–51.

Innocenti P, Morrow EH. 2010. The sexually antagonistic genes of *Drosophila melanogaster*. *PLoS Biol.* 8:e1000335.

Jaquiery J, Rispe C, Roze D, Legeai F, Le Trionnaire G, Stoeckel S, Mieuzet L, Da Silva C, Poulain J, Prunier-Leterme N, et al. 2013. Masculinization of the x chromosome in the pea aphid. *PLoS Genet.* 9:e1003690.

Jiang ZF, Machado CA. 2009. Evolution of sex-dependent gene expression in three recently diverged species of *Drosophila*. *Genetics* 183: 1175–1185.

Jordan CY, Charlesworth D. 2012. The potential for sexually antagonistic polymorphism in different genome regions. *Evolution* 66:505–516.

Kayserili MA, Gerrard DT, Tomancak P, Kalinka AT. 2012. An excess of gene expression divergence on the X chromosome in *Drosophila embryos*: implications for the faster-X hypothesis. *PLoS Genet.* 8: e1003200.

Khil PP, Smirnova NA, Romanienko PJ, Camerini-Otero RD. 2004. The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. *Nat Genet.* 36: 642–646.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14:R36.

Kirkpatrick M, Guerrero RF. 2014. Signatures of sex-antagonistic selection on recombining sex chromosomes. *Genetics* 197:531–541.

Kondrashov AS, Crow JF. 1991. Haploidy or diploidy: which is better? *Nature* 351:314–315.

Kousathanas A, Halligan DL, Keightley PD. 2014. Faster-X adaptive protein evolution in house mice. *Genetics* 196:1131–1143.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.

Leder EH, Cano JM, Leinonen T, O'Hara RB, Nikinmaa M, Primmer CR, Merila J. 2010. Female-biased expression on the X chromosome as a key step in sex chromosome evolution in threespine sticklebacks. *Mol Biol Evol.* 27:1495–1503.

Li L, Stoeckert CJ Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13:2178–2189.

Lipinska AP, D'Hondt S, Van Damme EJ, De Clerck O. 2013. Uncovering the genetic basis for early isogamete differentiation: a case study of *Ectocarpus siliculosus*. *BMC Genomics* 14:909.

Luthringer R, Cormier A, Ahmed S, Peters AF, Cock JM, Coelho SM.. 2015. Sexual dimorphism in the brown algae. *Perspect Phycol.* 1: 11–25.

Mank JE. 2013. Sex chromosome dosage compensation: definitely not for everyone. *Trends Genet.* 29:677–683.

Mank JE, Ellegren H. 2009. Are sex-biased genes more dispensable? *Biol Lett.* 5:409–412.

Mank JE, Hultin-Rosenberg L, Axelsson E, Ellegren H. 2007. Rapid evolution of female-biased, but not male-biased, genes expressed in the avian brain. *Mol Biol Evol.* 24:2698–2706.

Mank JE, Nam K, Brunstrom B, Ellegren H. 2010. Ontogenetic complexity of sexual dimorphism and sex-specific selection. *Mol Biol Evol.* 27: 1570–1578.

Martins MJ, Mota CF, Pearson GA. 2013. Sex-biased gene expression in the brown alga *Fucus vesiculosus*. *BMC Genomics* 14:294.

Meisel RP. 2011. Towards a more nuanced understanding of the relationship between sex-biased gene expression and rates of protein-coding sequence evolution. *Mol Biol Evol.* 28:1893–1900.

Meisel RP, Malone JH, Clark AG. 2012. Faster-X evolution of gene expression in *Drosophila*. *PLoS Genet.* 8:e1003013.

Müller DG. 1975. Sex expression in aneuploid gametophytes of the brown alga *Ectocarpus siliculosus* (Dillw.) Lyngb. *Arch Protistenk.* 117:297–302.

Nagasato C, Motomura T. 2002. Influence of the centrosome in cytokinesis of brown algae: polyspermic zygotes of *Scytosiphon lomentaria* (Scytosiphonales, Phaeophyceae). *J Cell Sci.* 115:2541–2548.

Orr HA, Otto SP. 1994. Does diploidy increase the rate of adaptation? *Genetics* 136:1475–1480.

Otto SP, Pannell JR, Peichel CL, Ashman TL, Charlesworth D, Chippindale AK, Delph LF, Guerrero RF, Scarpino SV, McAllister BF. 2011. About PAR: the distinct evolutionary dynamics of the pseudoautosomal region. *Trends Genet.* 27:358–367.

Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression. *Nat Rev Genet.* 14:83–87.

Perry JC, Harrison PW, Mank JE. 2014. The ontogeny and evolution of sex-biased gene expression in *Drosophila melanogaster*. *Mol Biol Evol.* 31(5):1206–1219.

Peters AF, van Wijk SJ, Cho GY, Scornet D, Hanyuda T, Kawai H, Schroeder DC, Cock JM, Boo SM. 2010. Reinstatement of E. *crouaniorum* Thuret in Le Jolis as a third common species of *Ectocarpus* (Ectocarpales, Phaeophyceae) in western Europe, and its phenology at Roscoff, Brittany. *Phycol Res.* 58:157–170.

Pointer MA, Harrison PW, Wright AE, Mank JE. 2013. Masculinization of gene expression is associated with exaggeration of male sexual dimorphism. *PLoS Genet.* 9:e1003697.

Presgraves DC. 2008. Sex chromosomes and speciation in *Drosophila*. *Trends Genet.* 24:336–343.

Puigbo P, Bravo IG, Garcia-Vallve S. 2008. CAIcal: a combined set of tools to assess codon usage adaptation. *Biol Direct.* 3:38.

Rice WR. 1984. Sex chromosomes and the evolution of sexual dimorphism. *Evolution* 38:735–742.

Schmid CE, Schroer N, Müller DG. 1994. Female gamete membrane glycoproteins potentially involved in gamete recognition in *Ectocarpus siliculosus* (Phaeophyceae). *Plant Sci.* 102:61–67.

Sharma E, Kunstner A, Fraser BA, Zipprich G, Kottler VA, Henz SR, Weigel D, Dreyer C. 2014. Transcriptome assemblies for studying sex-biased gene expression in the guppy, *Poecilia reticulata*. *BMC Genomics* 15:400.

Smith G, Chen YR, Blissard GW, Briscoe AD. 2014. Complete dosage compensation and sex-biased gene expression in the moth *Manduca sexta*. *Genome Biol Evol.* 6:526–537.

Stache-Crain B, Müller DG, Goff LJ. 1997. Molecular systematics of *Ectocarpus* and *Kuckuckia* (Ectocarpales, Phaeophyceae) inferred from phylogenetic analysis of nuclear and plastid-encoded DNA sequences. *J Phycol.* 33:152–168.

Starr RC, Zeikus JA. 1993. UTEX-The culture collection of algae at the University of Texas at Austin. *J Phycol.* 29(Suppl.): 1–106.

Sterck L, Billiau K, Abeel T, Rouze P, Van de Peer Y. 2012. ORCAE: online resource for community annotation of eukaryotes. *Nat Methods.* 9: 1041.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34:W609–W612.

Szovenyi P, Ricca M, Hock Z, Shaw JA, Shimizu KK, Wagner A. 2013. Selection is no more efficient in haploid than in diploid life stages of an angiosperm and a moss. *Mol Biol Evol.* 30:1929–1939.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* 30:2725–2729.

Uebbing S, Kunstner A, Makinen H, Ellegren H. 2013. Transcriptome sequencing reveals the character of incomplete dosage compensation across multiple tissues in flycatchers. *Genome Biol Evol.* 5:1555–1566.

Vicoso B, Charlesworth B.. 2009. Effective population size and the faster-X effect: an extended model. *Evolution* 63(9):2413–2426.

Vicoso B, Kaiser VB, Bachtrog D. 2013. Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. *Proc Natl Acad Sci U S A.* 110:6453–6458.

Whittle CA, Johannesson H. 2013. Evolutionary dynamics of sex-biased genes in a hermaphrodite fungus. *Mol Biol Evol.* 30: 2435–2446.

Wyman MJ, Cutter AD, Rowe L. 2012. Gene duplication in the evolution of sexual dimorphism. *Evolution* 66:1556–1566.

Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, Bar-Even A, Horn-Saban S, Safran M, Domany E, et al. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* 21: 650–659.

Yang Z. 2000. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. *J Mol Evol.* 51:423–432.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.

Zdobnov EM, Apweiler R. 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17: 847–848.

Zhang Z, Hambuch TM, Parsch J. 2004. Molecular evolution of sex-biased genes in *Drosophila*. *Mol Biol Evol.* 21:2130–2139.

Discussion et perspectives

Les résultats de l'analyse de l'expression différentielle ont mis en évidence une quantité relativement limitée de gènes différentiellement exprimés entre mâle et femelle chez *Ectocarpus* (~12%), aussi bien au stade immature que mature. Cette valeur faible serait cohérente avec le niveau limité de dimorphisme sexuel chez *Ectocarpus*. Cependant, cette hypothèse reste à tester et des analyses sont déjà prévues à cette fin, en comparant le nombre de gènes différentiellement exprimés entre mâles et femelles avec le niveau de dimorphisme sexuel pour plusieurs algues brunes. En effet, le groupe des algues brunes présente des espèces avec différents niveaux de dimorphisme sexuel, comme au niveau des gamètes, avec des espèces isogames, anisogames ou bien oogames (Luthringer et al. 2014) et permettrait donc de tester cette hypothèse.

L'analyse de l'expression différentielle chez les gamétophytes immatures a suggéré que les mécanismes de différenciation au niveau métabolique entre mâles et femelles sont mis en place de manière précoce. Une partie des gènes identifiés comme différentiellement exprimés l'étaient aussi bien au stade immature que mature, tandis qu'une autre partie des gènes étaient différentiellement exprimés soit au stade immature ou au stade mature. Cela semble indiquer dans, le second cas de figure, que certains gènes interviennent à des moments précis dans le cycle développement de l'organisme. L'utilisation d'autres espèces d'algues brunes et l'analyse de leurs transcriptomes à différents stades du cycle de vie, permettraient de vérifier l'établissement précoce des mécanismes de la différenciation sexuelle. De plus, l'étude approfondit des gènes biaisés par le sexe, à des stades spécifiques du développement de l'organisme, pourrait aider à identifier les voies métaboliques impliquées dans la croissance et la maturation des individus. En complément, pour *Ectocarpus*, la disponibilité de librairies RNA-seq des gamètes mâles et femelles permettrait d'affiner l'identification des gènes impliqués dans le processus de développement des gamétophytes.

Nous avons montré, chez *Ectocarpus*, que l'évolution moléculaire des gènes biaisés par le sexe était similaire entre les gènes mâles et femelles. Le faible dimorphisme sexuel entre mâles et femelles pourrait expliquer que la pression évolutive soit relativement faible entre mâles et femelles, et expliquerait en partie cette évolution symétrique des gènes biaisés. L'identification et l'analyse des gènes biaisés par le sexe dans d'autres algues brunes, présentant différents niveaux de dimorphisme sexuel, permettraient de vérifier cette hypothèse. L'identification des gènes dupliqués dans les autres

génomes d'algues brunes permettrait d'analyser la vitesse d'évolution moléculaire de ces gènes par rapport aux gènes non dupliqués.

Au niveau des outils bioinformatiques utilisés, les résultats de DESeq et DESeq2 ont été comparés afin de valider les résultats publiés et analysés en utilisant DESeq. Au stade immature, 1662 gènes ont été identifiés comme différentiellement exprimés par DESeq et 1532 par DESeq2. La comparaison des résultats a montré que 1418 gènes (80%) sont retrouvés en commun par les deux logiciels, 247 (13%) étant spécifiques à DESeq et 114 (6%) spécifiques à DEseq2. Au stade mature, 482 gènes ont été identifiés comme différentiellement exprimés par DESeq et 635 par DESeq2. La comparaison des résultats a montré que 425 gènes (61%) sont retrouvés en commun par les deux logiciels, 57 (8%) étant spécifiques à DESeq et 210 spécifiques à DEseq2 (30%). Après analyse des résultats, aucune différence significative n'a été constatée sur l'interprétation des données d'expression différentielle et donc sur les résultats et les conclusions de l'article publié.

Conclusions générales et perspectives

Les travaux présentés dans ce Chapitre ont permis d'apporter une importante contribution sur les connaissances dans le domaine de l'analyse fonctionnelle et évolutive du déterminisme sexuel, plus particulièrement pour les chromosomes sexuels de type UV. *Ectocarpus* représente un modèle particulièrement intéressant dans l'étude de mécanismes évolutifs et fonctionnels des chromosomes sexuels. D'une part, en raison de l'importante distance phylogénétique la séparant des autres espèces communément étudiées et d'autre part, car ce type d'étude permet d'accroitre la quantité de données disponibles pour le système sexuel UV, afin de tester les différentes hypothèses sur la dynamique évolutive des systèmes de détermination du sexe.

Les résultats présentés ont montré que les chromosomes UV chez *Ectocarpus* ont une trajectoire évolutive différente comparée aux systèmes XY et ZW. Cependant, certaines caractéristiques qui ont été révélées montrent des similarités très fortes avec ces deux systèmes, tels que l'accumulation d'éléments répétés dans le chromosome sexuel ou bien encore une évolution moléculaire plus rapide des gènes de la SDR. Ces similarités montrent une certaine universalité des mécanismes impliqués dans l'évolution des chromosomes sexuels à travers des lignées très éloignées.

Le chromosome sexuel d'*Ectocarpus*, d'une taille de 5 Mpb, est caractérisé par la présence de deux régions non recombinantes (PAR), respectivement d'une taille de 2,5 Mpb et de 1,5 Mpb, bordant la région non recombinante, d'environ 1 Mpb, aussi bien chez le mâle que chez la femelle. La région non recombinante aurait évolué entre 70 et 100 millions d'années, mais présente toujours une taille et une dégénérescence génétique relativement faible. La SDR d'*Ectocarpus* présente un enrichissement en gènes surexprimés au stade gamétophytes matures, phase durant laquelle le sexe est exprimé, avec des caractéristiques particulières au niveau de se structure génomique. La SDR est aussi enrichie en éléments transposables et présente une diminution de la densité de gènes. Ces gènes sont caractérisés aussi par un niveau moyen d'expression plus faible que dans le reste du génome, tandis que les gènes de la PAR, aussi plus faiblement exprimés, sont plus spécifiques du stade sporophytique, avec la présence de clusters de gènes surexprimés durant cette phase.

Au niveau du génome, le nombre de gènes différentiellement exprimés entre mâles et femelles est d'environ 12% lors de l'ensemble de la phase gamétophytique. Cette valeur, relativement faible par rapport à d'autres espèces, pourrait s'expliquer par le faible dimorphisme sexuel observé entre les

gamétophytes, pouvant impliquer la modification d'un nombre limité de voies métaboliques. Des études comparatives entre plusieurs algues brunes présentant des différences au niveau du dimorphisme sexuel seront menées afin de confirmer ou d'invalider cette hypothèse. De plus, un suivi temporel plus fin lors du développement des gamétophytes permettrait de mieux comprendre les mécanismes sous-jacents et les différentes voies métaboliques impliquées dans la mise en place du dimorphisme sexuel. Ces analyses seront complétées au niveau de l'étude de l'évolution moléculaire des gènes biaisés par le sexe par l'obtention de nouvelles données transcriptomiques, mais aussi génomique, sur un nombre important d'autres algues brunes. Elles permettront d'étudier plus en détail le lien entre la présence de gènes dupliqués et le niveau d'expression différentielle entre les copies, potentielle source de résolution de l'antagonisme sexuel.

Bien que l'ensemble de ces informations permette d'apporter un éclairage supplémentaire sur les mécanismes pour le système UV, beaucoup de travail reste à réaliser avant de pouvoir généraliser les conclusions qui ont été obtenues avec *Ectocarpus*. La disponibilité récente du génome d'une autre algue brune, *Saccharina japonica* (Ye et al. 2015), et les différents projets amorcés afin de séquencer différents génomes d'algues brunes devraient permettre d'étudier l'évolution de la structure des chromosomes sexuels et de leurs gènes au sein de ce clade, et ainsi aider à comprendre les mécanismes évolutifs du déterminisme sexuel.

Chapitre 2 : Annotation structurale et fonctionnelle chez l'algue brune modèle *Ectocarpus* sp.

Introduction

L'avènement du séquençage NGS a permis de grandement faciliter l'accès à l'information des gènes et des génomes. Cependant, l'effort d'annotation et de mise à jour de cette dernière sont focalisés sur un nombre restreint d'espèces, comme l'homme (Harrow et al. 2012) ou bien *Arabidopsis thaliana* (Lamesch et al. 2012), espèces qui présentes la particularité d'être des modèles en biologie. Ces modèles sont particulièrement importants pour diverses raisons, présentées plus en détail dans les parties suivantes. Comme développé dans le Chapitre 1, *Ectocarpus* est le modèle biologique pour les algues brunes. De l'importance de ce modèle pour ce groupe et dans le contexte de cette thèse, il a été possible de fournir une nouvelle version du génome ainsi que de l'annotation des gènes, travail présenté dans ce Chapitre.

Définition et caractéristiques d'un modèle biologique

Un modèle au sens large du terme est une généralisation, une représentation ou un point de référence de quelque chose qui peut prendre différentes formes, allant d'un objet, une équation ou bien un organisme (Hesse 1963; Ransom 1981; Hestenes 1987; Bolker 2009). Il sert de représentation par l'exemple ou par substitution. L'origine de la notion d'organisme modèle en biologie est difficile à retracer, mais l'établissement de ce concept s'est ancré dans les années 1960 et 1970 avec le développement des techniques de biologie moléculaire (Ankeny and Leonelli 2011; Dietrich et al. 2014). Aujourd'hui, l'une des définitions d'un organisme modèle en biologie est « *des espèces non humaines qui sont largement étudiées afin de comprendre un ensemble de phénomènes biologiques, avec l'espoir que les données et les théories générées par l'utilisation du modèle seront applicables à d'autres organismes, en particulier ceux qui sont en quelque sorte plus complexes que le modèle original* » (Ankeny and Leonelli 2011).

Le concept d'organisme modèle en biologie peut, selon les auteurs, se diviser en deux groupes : les organismes modèles expérimentaux et les organismes modèles génétiques (Bier and Mcginnis 2004; Ankeny and Leonelli 2011). On distingue d'abord les organismes modèles expérimentaux utilisés pour répondre à certains types de questions, pour lesquelles l'information de la séquence nucléotidique n'est pas nécessaire. Parmi ces modèles, on retrouve principalement les espèces utilisées pour l'étude du développement embryologique chez les vertébrés, comme le xénope (Bier and Mcginnis 2004). L'autre type de modèle est apparu plus récemment, avec la possibilité d'accéder à la séquence génomique, et correspond aux organismes modèles génétiques. Il qualifie les espèces qui sont utilisées pour des analyses génétiques et génomiques et regroupe un très grand nombre d'espèces, comme la souris, la drosophile, *Arabidopsis*, *C. elegans* ou encore le zebrafish. Une espèce n'est pas forcément exclusive à un groupe, par exemple, le chien appartient aux deux catégories. Il est utilisé à la fois comme ressource génétique pour, entre autres, l'étude des maladies génétique (Galibert et al. 2004), mais aussi comme ressource pour des études comportementales (Hare et al. 2002).

Que le modèle soit expérimental ou génétique, on distingue encore deux groupes de modèles en fonction des questions que l'on cherche à résoudre. Un premier groupe est constitué par les modèles « exemples » qui correspondent aux espèces, chacune représentative de son taxon. Le but est d'acquérir des connaissances sur ces espèces modèles afin de comprendre les processus et mécanismes de la biologie, et utiliser les connaissances acquises pour les généraliser et les étendre à l'ensemble du taxon représenté (Bolker 2009). Cependant, la généralisation de ces connaissances reste soumise au fait que l'on considère que l'espèce modèle utilisée est bien représentative du taxon. Le deuxième groupe contient les modèles de « substitution » dont l'utilisation principale est centrée à des fins de recherche biomédicale. L'objectif est d'utiliser des espèces proches de l'homme pour comprendre les mécanismes des maladies humaines, dans le but de développer des traitements (Bier and Mcginnis 2004), mais aussi de tester la nocivité de substances ou pathogènes (Bolker 2009). Le NIH fournit une liste des organismes pouvant être utilisés à des fins de recherche biomédicale (http://www.nih.gov/science/models/).

Méthode de sélection d'un modèle biologique

La sélection d'un modèle est une tâche relativement complexe qui dépend à la fois des connaissances acquises ce potentiel modèle et ce que l'on cherche à connaître. Les questions dont on cherche à connaître les réponses sont donc importantes pour orienter le choix du modèle. Ainsi, plus la quantité des informations disponibles sera importante, plus théoriquement, il sera facile de

déterminer si le modèle sera adapté ou non pour répondre aux différentes questions. D'autres critères peuvent entrer en jeu pour la sélection du modèle, par exemple l'aspect historique pour les espèces qui sont étudiées de longue date et dont la littérature est assez étendue. Ou bien la position phylogénétique, pour permettre l'utilisation du modèle dans le but obtenir des inférences pour les autres espèces du groupe ou pour les comparer avec les modèles d'autres taxons pour des analyses d'évolution. Les organismes modèles ont des caractéristiques expérimentales particulières, comme un temps de développement de génération court avec un haut taux de fertilité, une taille physique et de génome réduit, un organisme facile à maintenir et une disposition à l'utilisation de techniques de modification génétique.

Cependant, les progrès de ces dernières années dans le développement des technologies de séquençage haut débit (Illumina, Roche, PacBio, etc) et des techniques d'édition des génomes (CRISPR/cas9 ou TALEN) font que la limite qui existait jusqu'à présent entre les organismes modèles génétiques et non modèles génétiques commence à devenir floue (Müller and Grossniklaus 2010). Ces technologies permettent d'obtenir et d'étudier de manière précise une grande quantité d'informations sur des organismes qui ne correspondent pas aux critères de sélection d'un organisme modèle, mais qui présentent des intérêts particuliers à divers niveaux. Par exemple, des organismes avec des intérêts pour l'étude d'un groupe, de certaines maladies ou qui sont relativement proches du modèle de référence d'un groupe, mais avec de fortes différences phénotypiques ou génotypiques. Les cas les plus emblématiques correspondent aux projets génome nK qui consistent à faire du séquençage massif de populations, par exemple les projets 1000 génomes (Consortium 2010) et 100 000 génomes (http://www.genomicsengland.co.uk/the-100000-genomes-project/), ou bien d'individus isolés dans une multitude de taxons, comme dans le projet génome 1 kp (Matasci et al. 2014).

Séquençage du génome de l'organisme modèle

Dans le cas d'un organisme modèle génétique, la première étape suivant la sélection de l'espèce est le lancement d'un projet génome consistant à obtenir la séquence nucléotidique du modèle retenu. Cette étape pose de nombreux défis, aussi bien au niveau du séquençage que de l'assemblage du génome. Les premiers projets de génomes Eucaryotes dans les années 1990 et début 2000 étaient basés sur l'utilisation de librairies YAC puis BAC et sur la technologie de séquençage Sanger (Lander et al. 2001). Le principal problème de premiers projets de séquençage et assemblage de génome était l'effort considérable qu'ils réclamaient, aussi bien au niveau financier, humain et en terme de durée, par exemple plus de 13 ans et 3 milliards de dollars pour le projet génome humain (Lander et al. 2001;

Sboner et al. 2011). L'avènement de la seconde génération de séquençage, avec les technologies développées principalement par des sociétés telles que Roche et Illumina, a permis de démocratiser les projets génomes. En effet, ces technologies ont induit une diminution drastique du coût et du temps de séquençage tout en augmentant le débit et en déplaçant ainsi les efforts en termes de ressources et temps de la partie séquençage vers les parties assemblage et post-analyse (Sboner et al. 2011). La réduction de la taille des reads, principalement dans le cas de la technologie développée par la société Illumina, a demandé le développement d'outils d'assemblage de génome adapté à ces types de séquences et pose des problèmes, tels que l'assemblage des régions répétées. Les derniers développements des technologies de séquençage ont pour but de faciliter la partie d'assemblage des génomes en allongeant la taille des reads afin de résoudre les problèmes posés par l'assemblage des reads courts. Les développements en cours sont réalisés par les sociétés telles que Pacific Biosciences qui propose déjà la technologie PacBio, Oxford Nanopore Technologies Ltd, Genia Corporation, BioNano Genomics, LightSpeed Genomic.

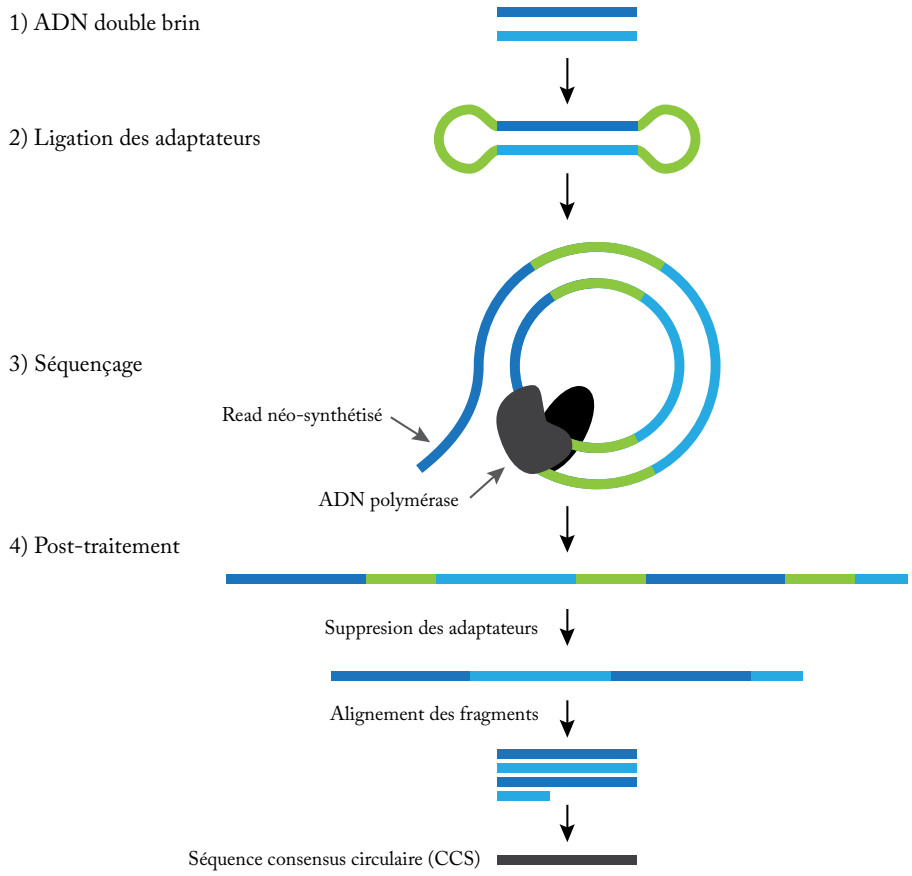


Figure 14 : Principe du séquençage PacBio CCS (circular consensus sequence) (Fichot and Norman 2013).

A titre d'exemple, pour *Ectocarpus*, deux génomes ont été assemblés à des périodes et avec des méthodologies différentes. Le premier génome, présenté dans le chapitre précédent, correspond au génome de référence, issu d'un individu mâle haploïde (Cock et al. 2010). Ce dernier a été séquencé en utilisant la technique Sanger à partir de librairies plasmidiques, produisant des reads d'environ 600 pb et avec une couverture supérieure à 10x. Les reads ont été assemblés avec le logiciel Arachne 2 (Jaffe et al. 2003), qui a produit 14 043 contigs assemblés en 1 902 scaffolds. Le deuxième génome, dont seulement une partie des séquences a été publié (Ahmed et al. 2014), provient d'un individu femelle haploïde. Pour ce génome, plusieurs technologies de séquençage ont été utilisées et combinées. Dans un premier temps, quatre librairies single-end basées sur la technologie de pyroséquençage Roche 454, ainsi qu'une librairie paired-end et deux librairies mate-paire utilisant la technologie Illumina, ont été séquencées, pour une couverture avoisinant au final les 100x. L'assemblage de ces données avec Velvet (Zerbino and Birney 2008) a produit 34 856 contigs assemblés en 19 264 scaffolds. Dans un second temps, afin de tenter de réduire la fragmentation de l'assemblage, un séquençage PacBio CCS a été réalisé, produisant une couverture de 1,5 x (**Figure 14**). L'amélioration de l'assemblage a été effectué en utilisant les outils du package SMRT-Analysis, plus particulièrement l'outil AHA (A Hybrid Assembler) (https://github.com/PacificBiosciences/SMRT-Analysis) (**Figure 15**). L'utilisation de cette technologie, permettant d'obtenir de long reads, peut partiellement aider à faire le lien entre plusieurs contigs, n'a cependant pas permis d'améliorer l'assemblage. D'une part, la longueur des reads obtenus a été plus courte (~3,5 Kb) que la moyenne normalement obtenue (~5Kb) lors du séquençage, résultat d'un processus d'extraction de l'ADN chez *Ectocarpus* qui n'autorise que difficilement l'obtention de fragments de grande taille. D'autre part, l'efficacité du séquençage n'était pas optimale et la couverture très limitée n'a pas permis de lier correctement les différents contigs entre eux.

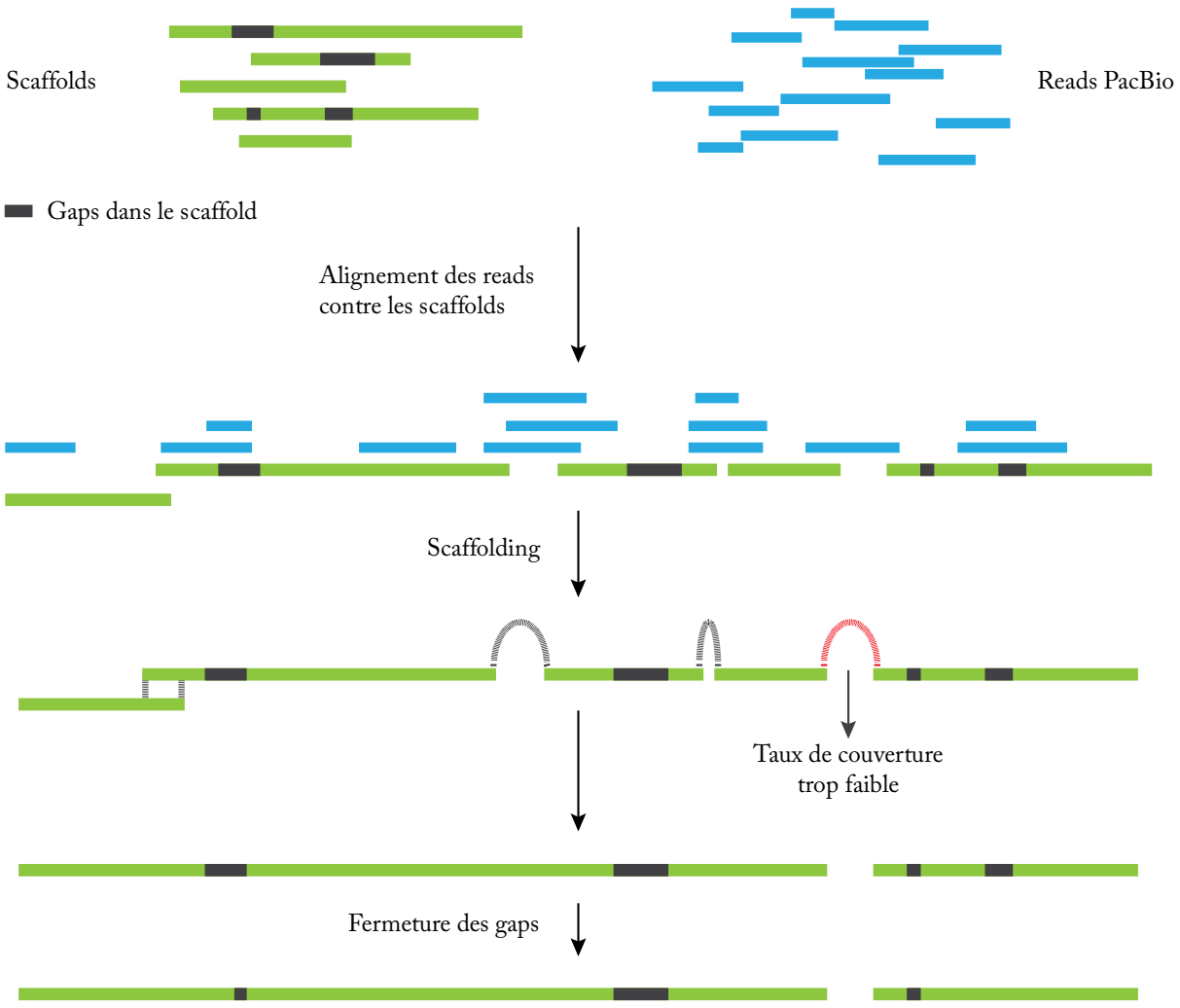


Figure 15 : Principe de fonctionnement de l'outil AHA (A Hybrid Assembler) pour le scaffolding de séquence génomique. Les reads PacBio sont dans un premier temps alignés contre les scaffolds avec la possibilité que certains reads chevauchent plusieurs scaffolds. Ensuite, les différents liens entre scaffolds sont répertoriés et seuls ceux ayant un taux de couverture suffisant sont conservés. En parallèle, les liens formés directement entre les scaffolds sont identifiés. La séquence des reads PacBio est utilisée afin de joindre les scaffolds entre eux au niveau des liens validés. Enfin, une étape supplémentaire peut-être réalisée, qui permet d'utiliser l'information des alignements pour fermer les gaps au sein des scaffolds.

Annotation d'un génome

Un génome est le résultat d'un ancien et long processus évolutif, et contient une grande diversité d'éléments aussi bien structuraux que régulateurs. L'accès à la séquence du génome ne représente qu'un premier niveau d'information finalement assez peu utilisable à l'état brut. Un processus d'annotation est donc indispensable pour permettre l'ajout des couches d'informations supplémentaires indispensables pour extraire l'information biologique afin de permettre l'analyse et l'interprétation des processus biologiques (**Figure 16**). Un prérequis essentiel avant de démarrer cette

étape d'annotation est de posséder un niveau de qualité de l'assemblage du génome suffisant, comme une faible fragmentation des contigs et assez peu de nucléotides inconnus, pour permettre une annotation correcte des différents éléments constitutifs. Dans le cas contraire, le processus d'annotation ne peut se dérouler de manière correcte (Yandell and Ence 2012). Le processus d'annotation est découpé en un ensemble de sous processus adaptés à l'identification et l'annotation de chaque type d'élément présent dans le génome. C'est un processus qui est lent, très consommateur en temps et en ressources computationnelles et humaines. Il est, de plus, différent entre génome Procaryote et Eucaryote, les structures et l'organisation des gènes étant très différentes entre les deux clades, mais aussi à l'intérieur de chaque clade. Dans la suite du manuscrit, la description de l'annotation sera focalisée sur le groupe des Eucaryotes.

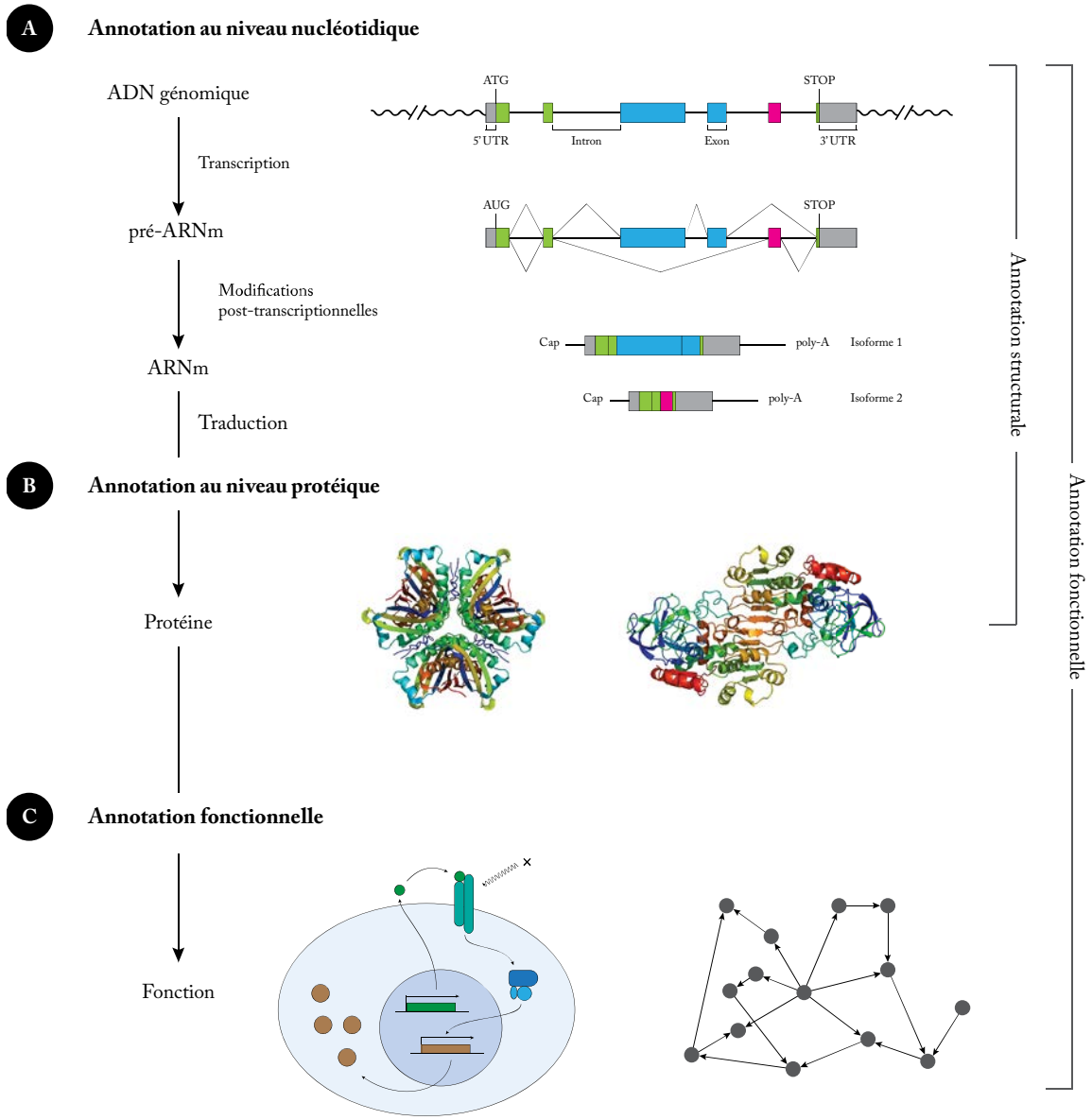


Figure 16 : Les différents niveaux de l'annotation. Adapté de (Stein 2001; Zhang 2002).

Annotation structurale, au niveau nucléotidique

Le début du processus d'annotation est une étape massivement automatique faisant appel à une grande quantité de ressources computationnelles.

L'identification des éléments répétés représente la toute première phase du processus d'annotation structurale. Il a pour but de détecter et d'identifier les régions de faible complexité et les éléments transposables, afin de les recenser et localiser. Cette identification est un point crucial à réaliser pour chaque génome, les éléments répétés étant souvent spécifiques à chaque génome (Yandell and Ence 2012). En plus de l'information apportée aussi bien quantitative que qualitative par cette étape de détection et d'identification, elle est utilisée dans la suite du processus d'annotation afin de « masquer » ces régions (A, T, G, C vers N). Cette étape de « masking » des éléments répétés est particulièrement importante lors de la phase de prédiction automatique des gènes codant pour minimiser le risque de produire des faux positifs. En effet, la présence d'une phase ouverte de lecture (ORF – Open Reading Frame) dans certains transposons peut induire en erreur les prédicteurs de gènes (Yandell and Ence 2012), d'autant plus que ces transposons peuvent représenter une grande fraction d'un génome Eucaryote. Chez l'homme, 47 % de la séquence génomique est constituée de répétitions (Lander et al. 2001). De plus, l'absence de cette étape peut poser des problèmes, par exemple pour les recherches de similitude BLAST (Konerding 2004).

L'alignement des données représente la seconde étape. Elle consiste le plus souvent à aligner les protéines de l'espèce ou d'espèces proches sur le génome, ainsi que les données EST partielles ou pleine longueur et enfin les données RNA-seq, afin de disposer d'informations sur la structure des gènes avant même le processus de prédiction. Les données RNA-seq représentent une source d'information particulièrement remarquable, quel que soit le type de séquençage utilisé (Illumina, 454, PacBio), permettant une bien meilleure délimitation des exons, des sites d'épissage alternatif et l'obtention de l'information sur les évènements d'épissage alternatif (Yandell and Ence 2012; Eckalbar et al. 2013). Plus précisément, l'assemblage des reads RNA-seq, *de novo* ou avec génome de référence, permet la reconstruction complète d'une grande partie des transcrits, aussi bien dans la partie codante (CDS) que non codante (UTR) (Lomsadze et al. 2014), ce qui fournit une excellente source d'information pour les prédicteurs de gènes.

Ensuite, vient l'étape à proprement parlé de l'identification des gènes codants qui consiste à prédire la localisation et la structure des gènes du génome. Étant donné les variations extrêmement

importantes entre les organismes au niveau de la structure des gènes, cela réclame des outils suffisamment souples pour pouvoir s'adapter à tout type de génome. En contrepartie, les logiciels de prédiction de gènes demandent un entraînement spécifique à chaque espèce ou groupe d'espèces afin d'optimiser leurs paramètres (Yandell and Ence 2012). Il existe deux méthodes pour réaliser la prédiction des gènes. La première consiste à utiliser des prédicteurs *ab initio* apparus dans les années 90 (Brent 2008). Ils utilisent des modèles mathématiques afin de réaliser la prédiction de gènes et sans avoir besoin d'une source externe d'information autre que le génome. La limitation principale à ce type de prédiction est qu'elle fournit des structures de gènes incomplètes. Elle ne permet la prédiction qu'exclusivement de la structure codante du gène (CDS) faisant l'impasse sur la structure des UTR (Brent 2008). Le deuxième type d'approche est, lui aussi, basé sur l'utilisation d'une prédiction *ab initio,* mais avec l'utilisation de sources externes d'informations. On parle alors de logiciel de prédiction de gènes fondée sur des preuves (evidence-driven gene predictions) (Yandell and Ence 2012). Pour cela, ils utilisent les résultats d'une grande variété de sources externes, par exemple les alignements des EST et/ou des transcrits, des résultats d'alignement de protéine, la position des éléments répétés ou encore la probabilité d'un site AG ou GT d'être respectivement accepteur ou donneur d'un exon. L'obtention de ces données externes au prédicteur demande un travail considérable, chacune de ces sources d'information réclamant un pipeline d'analyse dédié, mais l'amélioration de la qualité de la prédiction est particulièrement importante par rapport à une simple prédiction *ab initio*, pouvant permettre l'obtention de la structure complète des gènes et de ses isoformes.

Identification des ARN non codants

L'annotation des génomes est passée d'un stade où on se limitait à la recherche et l'annotation des gènes codant pour des protéines, à un stade où l'on cherche aussi à déterminer et annoter le plus grand nombre d'éléments possible, par exemple les pseudogènes, les régions régulatrices et les ncRNA (non-coding RNA) (Griffiths-Jones et al. 2003; Karro et al. 2007; Lagesen et al. 2007). Par exemple chez l'humain, des efforts ont été faits pour compléter et améliorer l'annotation des gènes non codants, notamment des lncRNA (Consortium et al. 2005; Derrien, Johnson, et al. 2012; Harrow et al. 2012). L'identification de ces ncRNA demande des méthodologies et des outils spécifiques et est réalisée soit en parallèle ou après l'annotation des gènes codants. Ces dernières années, deux grandes catégories de ncRNA sont particulièrement étudiées, les miRNA et les lncRNA, présentés dans les deux parties suivantes.

miRNA

Les miRNA sont des gènes non codants d'environ 22 nucléotides de long avec un rôle de régulation de l'expression des gènes dans le génome, par leur action de répression de l'expression (Lee et al. 1993; Krol et al. 2010). Ils sont situés dans les régions intergéniques, mais aussi dans des gènes codants ou non codants, aussi bien dans les introns que les exons (Kim et al. 2009; Ha and Kim 2014). On les retrouve chez les animaux et les plantes, mais entre les deux groupes, les mécanismes de biogenèse des miRNA sont différents. Les miRNA semblent absents dans le groupe des champignons. Cette absence chez les champignons et la présence de deux types de mécanismes de biogenèse suggèrent que le système de régulation par les miRNA aurait évolué au moins à deux reprises (Jones-Rhoades et al. 2006).

Comme dit précédemment, les mécanismes de biogenèse des miRNA sont différents entre les animaux et les plantes (**Figure 17**). La différence principale vient de l'étape de maturation du miRNA qui se déroule uniquement dans le noyau chez les plantes, et dans le noyau et le cytoplasme chez les animaux. Chez les animaux (**Figure 17a**) (Bartel 2004; He and Hannon 2004; Ha and Kim 2014), l'étape de transcription du miRNA par une polymérase II permet l'obtention d'un pri-miRNA. La première étape de maturation du miRNA est initiée par l'enzyme DROSHA qui va cliver le pri-miRNA afin d'obtenir le pre-miRNA. Ce dernier est ensuite exporté vers le cytoplasme à l'aide de la protéine exportin5, pour qu'il puisse terminer sa maturation. Une fois dans le cytoplasme, le pre-miRNA est clivé par l'enzyme Dicer pour libérer le duplex double brin du miRNA. Ce duplex produit deux miRNA matures, l'un qui correspond au brin 5' (suffixe -5p) et l'autre au brin 3' (suffixe -3p). Seulement l'un des deux brins est prévalent par rapport à l'autre, en terme d'expression et d'activité biologique. Le brin prévalent est simplement nommé miRNA mature, tandis que l'autre brin correspond au miRNA* (Ha and Kim 2014). Le duplex miRNA va se coupler à une protéine Argonaute (AGO) et former le complexe pre-RISC pour « RNA-induced silencing complex ». Le miRNA* est ensuite clivé pour ne conserver que le miRNA mature.

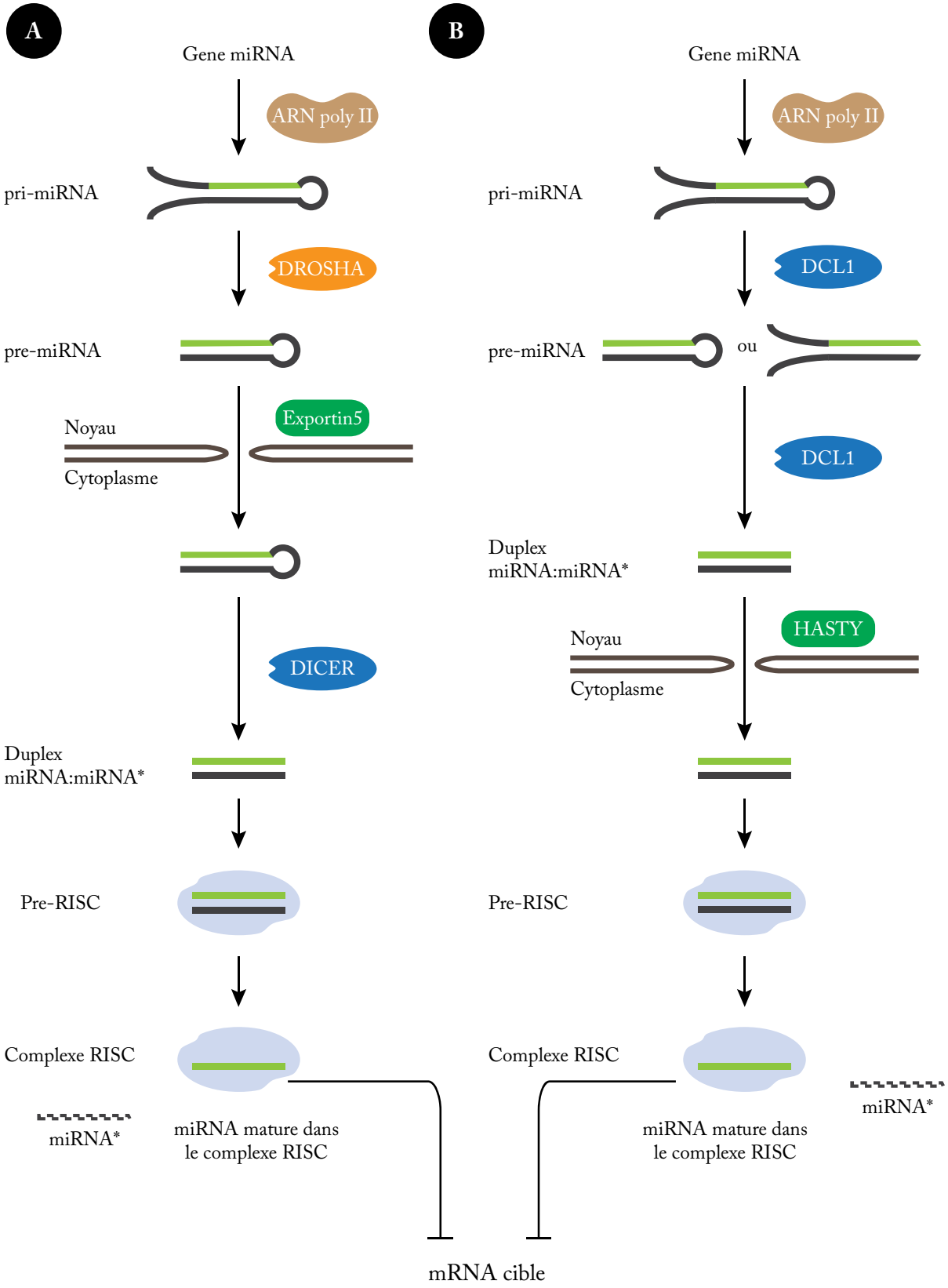


Figure 17 : Les voies de biogenèse de miRNA. Adapté de (Bartel 2004; Ha and Kim 2014). A) La voie de biogenèse des miRNA chez les animaux. B) La voie de biogenèse chez les plantes. Pour plus de détails, se référer au texte.

Chez les plantes (**Figure 17b**) (Bartel 2004; Axtell et al. 2011), contrairement aux animaux, la maturation du miRNA est complétée dans le noyau. L'étape de transcription du miRNA est réalisée par une polymérase II qui permet l'obtention d'un pri-miRNA. La première étape de maturation du miRNA est initiée par l'enzyme DCL1 pour « Dicer like ». Habituellement, le clivage est réalisé à la base du pri-miRNA, puis dans un deuxième temps, DCL1 clive la boucle du pre-miRNA précédemment obtenu pour former le miRNA. Cependant, il est possible que les deux étapes soient inversées, c'est-à-dire que dans un premier temps, il y a le clivage de la boucle pour former le pre-miRNA puis le clivage de la base qui donne le miRNA. Une fois la maturation du miRNA terminée, ce dernier est exporté vers le cytoplasme par la protéine HASTY, une protéine homologue à l'exportine5. Enfin, le miRNA se fixe à la protéine AGO pour former le complexe pre-RISC. Le miRNA* est ensuite clivé pour ne conserver que le miRNA mature.

Une fois le complexe RISC formé, l'action de régulation de l'expression des gènes se fait par ciblage du complexe par complémentarité de séquence entre le miRNA mature et le gène, dont le site de fixation est le plus souvent localisé au niveau du 3' UTR du gène cible (Bartel 2009).

Plusieurs méthodes bio-informatiques sont disponibles afin de permettre la détection et l'identification de miRNA. Une première méthodologie consiste à rechercher de manière *ab initio* les structures qui sont potentiellement des miRNA, avec des logiciels tels que MiPred (Jiang et al. 2007), miRPara (Wu et al. 2011) ou bien miRscan (Lim et al. 2003). La deuxième méthodologie est basée sur l'utilisation de données de séquençage afin de détecter la présence des miRNA dans un génome, avec des outils tels que mirDeep (Friedländer et al. 2008), DASP (Huang et al. 2010) ou bien miRAnalyzer (Hackenberg et al. 2009). Une dernière méthodologie se base sur la recherche de structures secondaires par la minimisation d'énergie afin d'identifier les motifs de repliement spécifiques aux miRNA, avec des outils tels que RNAfold (Hofacker 2002), UNAfold (Keith 2008) ou le ViennaRNA Package (Lorenz et al. 2011).

Malgré les performances correctes de ces logiciels, la structure complexe des miRNA fait qu'il est difficile de prédire de manière fiable leur structure, ce qui induit la prédiction d'un nombre significatif des faux positifs (Hu et al. 2012). Beaucoup reste à faire dans le domaine afin d'améliorer la sensibilité et la spécificité des logiciels.

MiRBase est aujourd'hui le principal catalogue de séquences et d'annotations de miRNA et contient des données pour plus de 200 espèces avec environ 25 000 loci de miRNA, soit environ

30 000 miRNA matures (Kozomara and Griffiths-Jones 2014). Cependant, une partie des annotations est incorrecte voire fausse (Meng et al. 2012). Par exemple, une analyse réalisée sur un sous ensemble des données a montré qu'un tiers des miRNA était potentiellement faux (Chiang et al. 2010). La grande quantité d'informations erronées présente dans miRBase s'explique en partie par l'insertion d'annotations obtenues à partir de données de NGS, données qui ne sont pas toujours assez sensibles pour la détection des miRNA (Ha and Kim 2014). Mais des efforts sont faits par les développeurs de miRBase afin de palier à ce problème, par exemple l'attribution de score de confiance pour chaque miRNA ou encore l'intervention de la communauté dans l'annotation experte des miRNA (Kozomara and Griffiths-Jones 2014).

lncRNA

Les lncRNA sont des gènes non codants de longueur variable, mais d'une taille minimum de 200 pb (Derrien et al. 2012; Rinn and Chang 2012). Les données de GENCODE montrent que 98 % des lncRNA sont épissés, mais qu'ils ne possèdent généralement que deux exons. Les exons des lncRNA ont une taille similaire à ceux des gènes codants, tandis que les introns des lncRNA ont une longueur significativement supérieure (Derrien et al. 2012). Au niveau de leur localisation dans le génome, il existe plusieurs catégories de lncRNA (**Figure 18**), avec une localisation intergénique ou génique. Dans le cas de la localisation génique, les lncRNA peuvent se situer au niveau d'un exon, dans un intron ou bien certains peuvent chevaucher un gène. Chez l'homme, la majorité des lncRNA est intergénique (Derrien et al. 2012). Les lncRNA sont majoritairement retrouvés dans le noyau de la cellule, mais certains sont présents dans le cytoplasme (Derrien et al. 2012; Fatica and Bozzoni 2014).

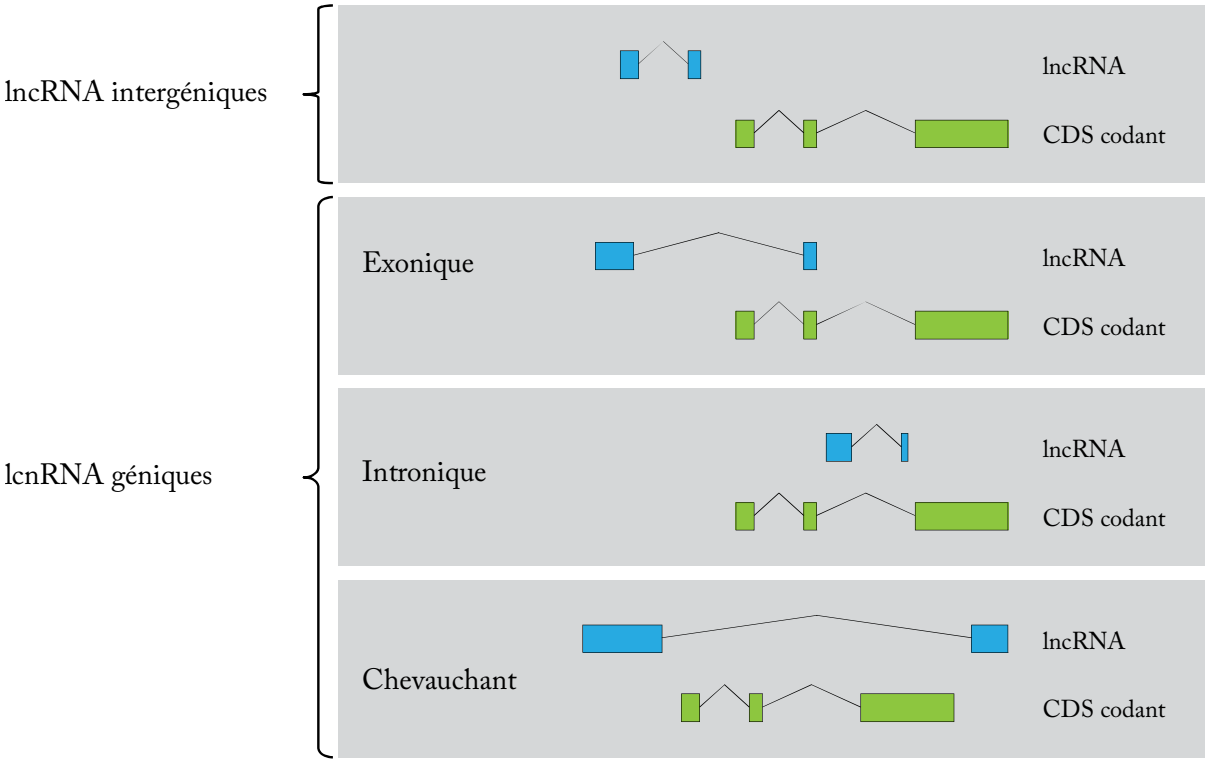


Figure 18 : Les différentes catégories de lncRNA selon leur localisation génomique (Derrien et al. 2012).

Tout comme les miRNA, les lncRNA régulent le niveau d'expression des gènes, mais aussi leur traduction, par répression ou par activation des gènes, via des mécanismes *cis* (Ørom et al. 2010; Derrien et al. 2012) ou *trans* régulateurs (**Figure 19**) (Cabili et al. 2011; Guttman et al. 2011; Derrien et al. 2012). Par exemple chez la drosophile, la transcription est réprimée par un complexe PRE/PcG qui va se fixer à la chromatine, tandis que le complexe PRE/trxG permet l'activation des gènes. La régulation entre les deux complexes est assurée par la présence de lncRNA qui va favoriser la formation du complexe PRE/trxG, et ainsi permettre l'expression des gènes (Schmitt et al. 2005). Au niveau de leur transcription, l'expression moyenne des lncRNA est plus faible que celle des gènes codant pour des protéines. De plus, l'expression des lncRNA semble être spécifique à certains tissus (Derrien et al. 2012; Djebali et al. 2012).

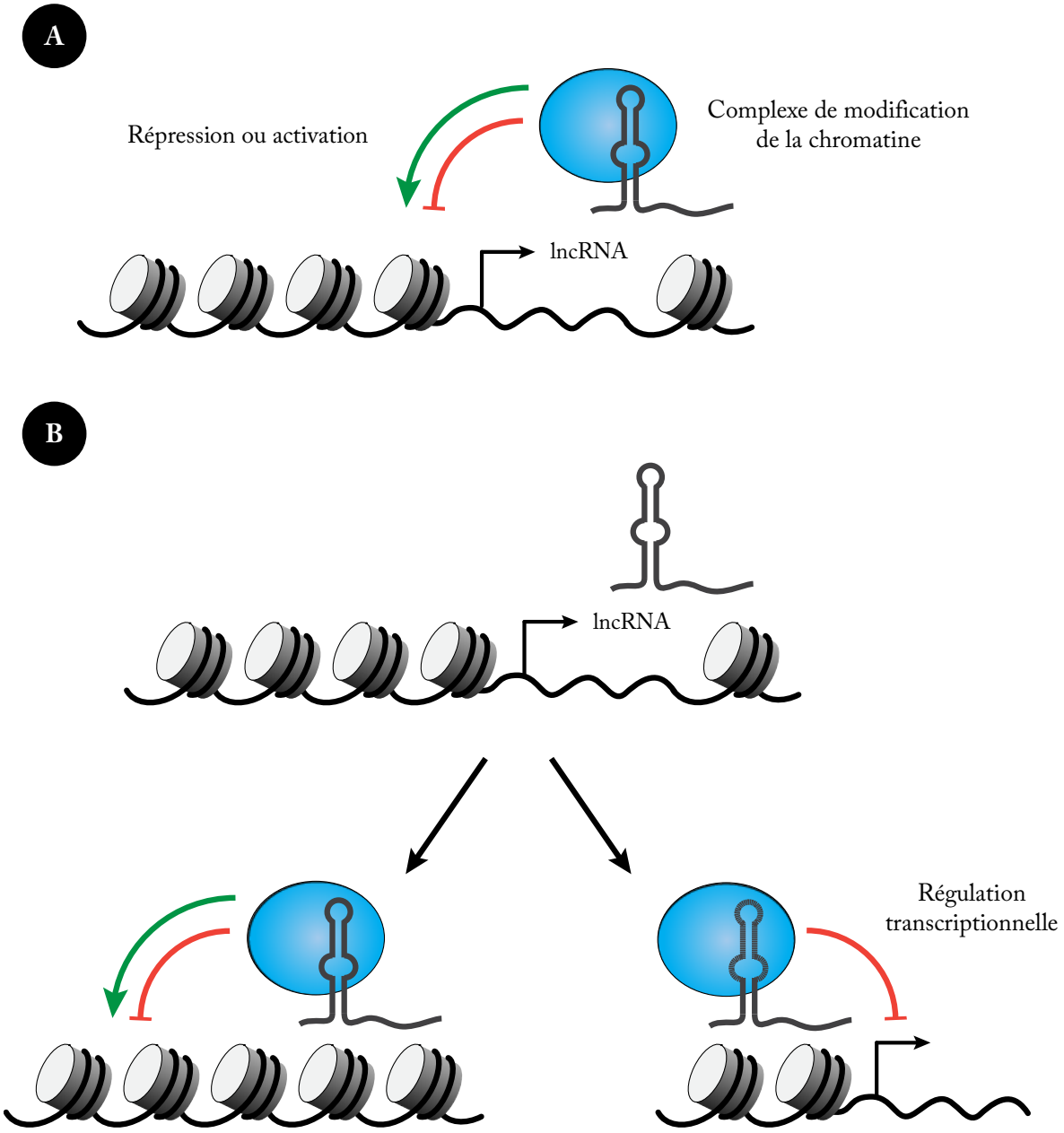


Figure 19 : Modèles d'actions des lncRNA nucléaires. Adapté de (Fatica and Bozzoni 2014). A) Exemple de régulation transcriptonnelle *cis* par les lncRNA. B) Exemples de régulation transcriptonnelle *trans* par les lncRNA.

De même que miRBase pour les miRNA, plusieurs bases de données ont été créées et centralisent les informations sur les lncRNA, avec des bases généralistes telles que lncRNAdb (Amaral et al. 2011; Quek et al. 2015) ou bien spécifiques à une espèce, comme LNCipedia pour l'homme (Volders et al. 2013; Volders et al. 2015). Elles ont pour but d'améliorer l'annotation des lncRNA afin d'en comprendre les différentes fonctions biologiques, en proposant aux utilisateurs des données manuellement annotées, ainsi que des outils pour réaliser et visualiser ces annotations.

La prédiction des lncRNA reste un domaine relativement récent. Actuellement, un nombre limité d'outils est disponible afin de réaliser cette tâche, tels que lncRNA-MFDL (Fan and Zhang 2015) ou LncRNA-ID (Achawanantakun et al. 2015), CPAT (Wang et al. 2013) ou bien PLEK (Li et al. 2014). Il reste donc encore beaucoup de développement de méthodologie et d'outils pour compléter la prédiction, la classification et la détermination de la fonction biologique des différents lncRNA (Derrien, Guigó, et al. 2012).

Identification des ARN non codants chez *Ectocarpus*.

Dans le cas d'*Ectocarpus*, une première prédiction des miRNA a été réalisée lors du projet génome, permettant l'identification de vingt-six miRNA. De plus, l'annotation structurale et fonctionnelle des mRNA a permis de mettre en évidence la présence d'un gène Dicer, Argonaute ainsi que de deux gènes codant pour des polymérases ARN-dépendant, indiquant la présence des gènes impliqués dans le processus de biogenèse des miRNA (Cock et al. 2010). La disponibilité de nouvelles données de séquençage spécifiques au sRNA (smallRNA) avec une profondeur de séquençage supérieure, et l'amélioration des logiciels d'analyses ont permis de vérifier l'exhaustivité et la qualité des premières annotations. De plus, la disponibilité de données RNA-seq, présentées dans la Chapitre 1, a permis de réaliser la détection de lncRNA dans le génome d'*Ectocarpus*. Ces deux analyses s'inscrivent dans un projet plus global de réannotation complète du génome d'*Ectocarpus*.

Annotation fonctionnelle, au niveau des protéines

Après l'annotation structurale, qui permet d'obtenir la position et la structure des gènes du génome, il est important de déterminer la fonction associée à ces gènes. Le processus d'annotation fonctionnelle est réalisé au niveau de la structure protéique des gènes dans l'objectif d'identifier la fonction potentielle des gènes, les domaines et les motifs protéiques connus, ainsi que la recherche des familles de gènes.

L'une des méthodes les plus couramment utilisées est la recherche d'une homologie de séquence avec d'autres protéines avec des outils tels que le blastp (Altschul et al. 1990) contre différentes bases de données. Il existe cependant des bases de données protéiques dites de « haute qualité » telles que SwissProt/UniProtKB (Consortium 2015) qui contient les séquences de protéines manuellement curées.

En plus de l'annotation par homologie de séquence, la recherche de domaines et de motifs protéiques est un complément d'information particulièrement informatif. De nombreuses bases

existent, telles que PFAM (Finn et al. 2014) ou PROSITE (Sigrist et al. 2013). Le développement de l'outil InterPro (Jones et al. 2014) a été motivé dans le but de pouvoir facilement faire de l'interrogation croisée de ces différentes bases de données de domaines et motifs pour obtenir une annotation unique au format InterPro.

Annotation fonctionnelle, au niveau des processus biologiques

La dernière étape de l'annotation consiste à faire le lien entre l'annotation des gènes et les différents processus biologiques associés, le but étant d'avoir une carte donnant les interactions entre l'ensemble des gènes du génome. Différents systèmes existent, comme Gene Ontology (Ashburner et al. 2000), KEGG (Kanehisa and Goto 2000) ou encore BioCyc (Caspi et al. 2014).

Annotation experte

L'information apportée par la prédiction automatique aussi bien structurale que fonctionnelle ne représente qu'une petite partie du travail d'annotation d'un génome, s'appuyant sur des ressources logicielles. Il est important d'aborder l'aspect humain dans ce genre de projet. Il représente la clef de voûte d'un projet génome, afin de fournir une annotation de qualité et durable dans le temps. Beaucoup de projets génomes intègrent une vérification manuelle de l'annotation qui consiste à identifier puis corriger les erreurs commises par les outils d'annotation automatique. Pour l'annotation structurale, cela consiste à modifier la structure des exons – introns. Pour l'annotation fonctionnelle, en plus de la correction possible de l'annotation, le but est aussi d'ajouter l'information contenue dans la littérature scientifique dans la base de données, afin de fournir l'annotation la plus précise possible et ajouter l'information qui n'est pas disponible dans les bases de données. Ces étapes de curation des données d'annotation automatique peuvent se dérouler de différentes façons. Elles peuvent faire appel à une communauté d'annotateurs dispersée et supervisée qui comprend plusieurs groupes indépendants travaillant sur un sous-ensemble des gènes (Mazumder et al. 2010) et qui sont supervisés par le responsable du projet d'annotation. Un rassemblement de la communauté d'annotation (Stein 2001; Mazumder et al. 2010; Yandell and Ence 2012) est une autre méthode utilisée avec succès pour les projets génomes entre autres de la drosophile (Hartl 2000; Pennisi 2000) et de la souris (Kawai et al. 2001). Elle consiste à réunir, dans un lieu et pour une période déterminée, un regroupement de biologistes et de bioinformaticiens afin de vérifier et d'améliorer l'annotation de manière intensive et rapide. Une autre solution est de faire appel à une communauté d'annotation composée d'étudiants ayant au préalable reçu une formation sur le fonctionnement de l'annotation

manuelle (Mazumder et al. 2010). Elle a par exemple été utilisée avec succès pour l'annotation de métagénomes (Hingamp et al. 2008).

Visualisation des données

La collecte et l'organisation des données telles que la séquence génomique, les annotations, les SNP, les données d'expression (microarray, RNA-seq) ou encore la carte génétique, sont particulièrement importantes afin de pouvoir fournir une ressource centralisée et un système de visualisation facile d'accès pour la communauté scientifique. Ces données sont intégrées dans des bases de données pour les organismes modèles qui ont pour rôle de mettre en relation les différentes données dans le contexte du génome afin de pouvoir les interpréter dans le contexte de la biologie de l'organisme (Cline and Kent 2009). La mise en place d'un système de bases de données pour un organisme modèle réclame un important travail de développement, autant au niveau de la mise en place du schéma de la base de données, que dans le développement des outils de visualisation et de l'interface (Stein et al. 2002). Afin de réduire ces coûts et d'éviter de refaire un développement pour chaque organisme modèle, les développements ont été centralisés au sein du projet GMOD (Generic Model Organism Database) (Stein et al. 2002) dans le but de proposer un ensemble d'outils pour utiliser pleinement les données associées à l'organisme modèle. Le projet GMOD propose par exemple des outils comme la base de données CHADO (Mungall and Emmert 2007), la visualisation des données avec JBrowse (Skinner et al. 2009) ou encore l'annotation structurale et fonctionnelle des gènes avec WebApollo (Lee et al. 2013). Les bases de données pour les organismes modèles ont la particularité d'être très dynamiques du fait de la possibilité d'apporter des corrections aux données et d'ajouter de l'information de manière manuelle ou automatique à travers une interface dédiée (Stein et al. 2002; Skinner et al. 2009).

Evolution de l'annotation

L'annotation d'un génome n'est pas une donnée statique, elle demande un travail continu après la mise à disposition de la première version. Elle est amenée à évoluer avec le temps par l'intégration aussi bien des nouvelles connaissances que de nouvelles données. L'évolution de l'annotation se fait généralement de manière constante et progressive dans le temps, avec la mise à jour régulière de l'annotation des gènes. Par exemple, ces dernières années chez l'homme, cette amélioration s'est entre autres focalisée sur la détection des différentes isoformes des transcrits (Trapnell et al. 2010; Grabherr et al. 2011; Au et al. 2013).

Elle peut cependant recevoir une nouvelle version pour laquelle l'ensemble de l'information de l'annotation est mise à jour. En effet, l'un des problèmes récurrents rencontré par les génomes annotés il y a plusieurs années est la faible quantité de gènes dont la prédiction inclut les UTR (Haas et al. 2005; Lorenzi et al. 2010; Eckalbar et al. 2013) et le manque de couverture fourni par les EST, limité aux gènes les plus fortement exprimés (Morozova et al. 2009). Dans un projet standard de séquençage des ADNc, approximativement 20-40 % des transcrits sont partiellement ou pas du tout séquencés (Brent 2008). Les avancées dans les technologies de séquençage et la diminution des coûts offrent une grande opportunité pour améliorer l'assemblage et l'annotation de ces génomes, par, entre autres, l'inclusion de données RNA-seq durant le processus de prédiction des gènes (Eckalbar et al. 2013). Ces données, dans le cas de la seconde génération de séquençage, permettent théoriquement de reconstruire les transcrits dans leur totalité (Lomsadze et al. 2014) (CDS + UTR) et autorisent la détection des évènements d'épissage alternatif (Ozsolak and Milos 2011). Les générations de séquençage ultérieures, grâce à leur taille de reads beaucoup plus longue, permettent d'obtenir directement les transcrits et potentiellement une partie ou la totalité des isoformes sans processus d'assemblage (Au et al. 2013). Ces données autorisent la détection aussi biens des transcrits faiblement exprimés que fortement exprimés et permet donc de lever la limitation liée aux EST. La diminution des coûts de séquençage offre aussi la possibilité de séquencer différents tissus et/ou différents stades du cycle de vie pour obtenir un transcriptome le plus complet possible. Par exemple, le processus de réannotation chez *Anolis carolinensis* utilise 14 librairies RNA-seq, de différents tissus et à différents stades du cycle de vie (Eckalbar et al. 2013).

La principale difficulté du processus de réannotation est de définir quand un gène doit être mis à jour et dans quelle mesure. La comparaison de l'annotation de référence avec la nouvelle prédiction et la consolidation des résultats est une tâche spécifique qui peut être réalisée par un nombre limité d'outils comme AEGeAN (Standage and Brendel 2012), PASA (Haas et al. 2003) ou encore MAKER (Cantarel et al. 2008). Cette tâche demande en effet de pouvoir obtenir une annotation consensus entre la précédente version de l'annotation et les informations apportées par les nouvelles données disponibles.

La base de données GenBank permet la mise à disposition de cette nouvelle annotation afin de remplacer l'ancienne, sous contrainte que la nouvelle annotation soit soumise par l'un des auteurs de la publication originale. Cependant, si la mise à jour apporte une amélioration significative, son intégration dans les bases de données publiques est possible pour des personnes ne faisant pas partie

de la publication originale (Yandell and Ence 2012). La démarche pour la mise à jour est relativement simple et demande de faire une nouvelle soumission des annotations (www.ncbi.nlm.nih.gov/genbank/wgs_update ; www.ncbi.nlm.nih.gov/genbank/eukaryotic_genome_submission).

Impact de l'annotation sur les analyses ultérieures

L'accessibilité des données NGS a permis le développement de nombreuses analyses, par exemple la détection de variants ou bien les analyses d'expression différentielle. Dans le cas d'études sur des espèces dont le génome et les annotations sont disponibles, le plus souvent les données de séquençage sont mappées contre le génome en se basant sur les prédictions des structures de gènes obtenus pendant le processus d'annotation. Seulement, il est rare de disposer d'un génome dont l'annotation structurale est parfaite. Récemment, plusieurs équipes se sont intéressées à l'évaluation de l'influence des données d'annotation structurale sur les résultats obtenus avec les données NGS. L'utilisation de l'annotation structurale lors du mapping des reads contre le génome de référence permet par exemple d'augmenter le taux de reads correctement mappés de manière globale. L'annotation fournit surtout une aide pour permettre un mapping correct des reads au niveau des sites de jonction exon-exon. Sans cette source d'information, une partie des reads ne s'alignent pas au niveau des sites de jonction ou alors de manière incorrecte (Zhao 2014; Zhao and Zhang 2015). Il a été aussi montré que selon l'origine de l'annotation structurale utilisée (RefSeq, ENSEMBLE, GENCODE, UCSC), des variations sur les résultats du mapping sont constatés, ce qui peut changer de manière importante le comptage des reads associés à chaque gène et les analyses d'expression différentielle (Zhao and Zhang 2015). De même, dans les analyses de détection de variants, le choix de l'annotation est particulièrement critique pour la suite des analyses et influence de manière significative les résultats (McCarthy et al. 2014; Frankish et al. 2015). Il est cependant important de garder à l'esprit que les variations des résultats peuvent être dues à la source d'information utilisée, mais aussi au choix des logiciels et des méthodologies employées (McCarthy et al. 2014).

Dans le cadre du projet *Ectocarpus*, la première version de l'annotation a été réalisée en 2007 et se basait sur l'utilisation du prédicteur Eugène (Schiex et al. 2001) et des données EST obtenues par la méthode Sanger (Cock et al. 2010). Comme montré précédemment, le développement des techniques de séquençage NGS et l'amélioration des outils de prédiction permettent d'apporter de nettes

améliorations des annotations structurales actuelles. Lors de cette thèse, de nombreuses données RNA-seq ont été générées, donnant la possibilité d'initier un processus de réannotation structurale et fonctionnelle du génome d'*Ectocarpus*. Ce projet a inclus aussi bien la mise à jour des gènes existants, que la prédiction de nouveaux gènes, codants (ARNm) que non codants (snoRNA (small nucleolar RNA), miRNA et lncRNA). Les résultats de ces travaux sont présentés dans les articles suivants.

Article 1 - Re-annotation and improved large-scale assembly of the genome of the brown algal model *Ectocarpus*

Introduction

Ectocarpus représente le modèle biologique chez les algues brunes, groupe phylogénétiquement éloigné des autres espèces modèles, dont le génome a été publié en 2010 (Cock et al. 2010), mais dont les annotations structurales ont été obtenues en 2007. Le processus de réannotation manuelle que j'ai réalisé sur le chromosome sexuel ainsi que sur le groupe de liaison 4, travaux présentés dans l'article 2 du chapitre 1, a montré que la disponibilité de données RNA-seq pouvait grandement aider dans l'amélioration globale de l'annotation structurale du génome. Afin de pouvoir intégrer l'information de ces données et fournir une réannotation complète de manière automatique, le pipeline utilisé lors de la première version de l'annotation a été mis à jour et relancé. A cette fin, un assemblage *de novo* avec Trinity, couplant dix librairies RNA-seq (quatre librairies provenant de gamétophytes immatures mâles et femelles, deux librairies par sexe ; quatre librairies provenant de gamétophytes matures mâles et femelles, deux librairies par sexe ; et deux librairies provenant d'un parthéno-sporophyte mature), a été réalisé afin d'obtenir le transcriptome le plus complet possible. Les transcrits assemblés ont été alignés contre le génome d'*Ectocarpus* à l'aide de GenomeThreader. Le résultat des alignements a été fourni à Eugène, en plus des données déjà générées lors de la première annotation. Le consensus entre la première version de l'annotation et cette nouvelle version a été réalisé en utilisant AEGeAn.

De plus, des données RAD-seq ont été générées permettant ainsi de réaliser une amélioration de l'assemblage du génome, en autorisant l'organisation et l'orientation des super-contig afin de les organiser en pseudo-chromosomes. Les annotations nouvellement obtenues ont ensuite été transférées sur la nouvelle structure génomique en utilisant ALLMAPS.

Article

Re-annotation and improved large-scale assembly of the genome of the brown algal model *Ectocarpus*

Auteurs : Cormier A[1], Avia K[1], Sterck L[2], Hitte C[3], Andres G[4], Godfroy O[1], Derrien T[3], Van De Peer Y[2], Corre E[4], Coelho SM[1], Cock JM[1*]

Affiliations :

[1] Sorbonne Université, UPMC Univ Paris 06, CNRS, Algal Genetics Group, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688, Roscoff, France;

[2] Department of Plant Systems Biology, VIB, Technologiepark 927, B-9052 Gent, Belgium;

[3] Institut de Génétique et Développement de Rennes, CNRS-URM6290, Université Rennes1, Rennes, France

[4] Abims Platform, CNRS-UPMC, FR2424, Station Biologique de Roscoff, CS 90074, 29688 Roscoff, France

Correspondance : cock@sb-roscoff.fr

Article en cours de rédaction

Re-annotation and improved large-scale assembly of the genome of the brown algal model *Ectocarpus*.

Authors: Cormier A[1], Avia K[1], Sterck L[2], Hitte C[3], Andres G[4], Godfroy O[1], Derrien T[3], Van De Peer Y[2], Corre E[4], Coelho SM[1], Cock JM[1]

Affiliations: [1]Sorbonne Université, UPMC Univ Paris 06, CNRS, Algal Genetics Group, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688, Roscoff, France, [2]Department of Plant Systems Biology, VIB, Technologiepark 927, B-9052 Gent, Belgium, [3]Institut de Génétique et Développement de Rennes, CNRS-URM6290, Université Rennes1, Rennes, France [4]Abims Platform, CNRS-UPMC, FR2424, Station Biologique de Roscoff, CS 90074, 29688 Roscoff, France

Introduction

Ectocarpus has been studied since the nineteenth century and work on this organism has provided many insights into novel aspects of brown algal biology (Müller, 1967; Charrier *et al.*, 2008). This long research history, together with several features of the organism that make it well adapted for genetic and genomic approaches (Coelho *et al.*, 2012), led to it being proposed as a general model organism for the brown algae in 2004 (Peters *et al.*, 2004) and to the initiation of a genome sequencing project that produced a first complete genome assembly in 2010 (Cock *et al.*, 2010). The publication of the genomic sequence was followed up with the development of many additional tools and resources including a genetic map (Heesch *et al.*, 2010), gene mapping techniques, microarrays (Dittami *et al.*, 2009; Coelho *et al.*, 2011), transcriptomic data (Ahmed *et al.*, 2014; Lipinska *et al.*, 2015), proteomic techniques (Ritter *et al.*, 2010) and bioinformatics tools (Gschloessl *et al.*, 2008; Prigent *et al.*, 2014). The genome information, together with these complementary resources, is currently being exploited to further our understanding of a broad range of processes, including life cycle regulation (Coelho *et al.*, 2011), sex determination (Ahmed *et al.*, 2014; Lipinska *et al.*, 2014; Lipinska *et al.*, 2015), development and morphology (Le Bail *et al.*, 2011), interaction with pathogens (Zambounis *et al.*, 2012) and metabolism (Meslet-Cladière *et al.*, 2013; Prigent *et al.*, 2014).

The brown algae are an important taxonomic group for several reasons; they are key, primary producers in many coastal ecosystems and, as such, have a major influence on marine biodiversity and ecology (Dayton, 1985; Steneck *et al.*, 2002; Bartsch *et al.*, 2008; Klinger, 2015; Wahl *et al.*, 2015). Brown algae also represent an important resource of considerable commercial value (Kijjoa & Sawangwong, 2004; Smit, 2004; Hughes *et al.*, 2012) and industrial exploitation of these organisms has increased markedly in recent years with the expansion of aquaculture activities, particularly in Asia (Tseng, 2001). Finally, brown algae are of considerable phylogenetic interest because they are very distantly related to well studied groups such as the animals, fungi and land plants and, moreover, have evolved complex multicellularity independently of these other lineages (Cock *et al.*, 2010; Cock & Collén, 2015).

A high-quality genome resource is essential if these important features of the brown algae are to be investigated effectively. The version of the *Ectocarpus* genome that was published in 2010 included detailed manual annotations of a large proportion of the genes but gene structure predictions were based on a limited amount of transcriptomic data (Sanger expressed sequence tags) and the large-scale assembly of the sequence contigs only associated about 70% of the genome sequence with linkage groups. Moreover, annotation efforts had focused almost exclusively on protein-coding genes, largely ignoring the non-coding component of the genome. The study described here set out to address these shortfalls exploiting the large amount of transcriptomic data that has been generated since the publication of the first version of the genome and using recently developed genetic and bioinformatic approaches to improved large-scale assembly and genome annotation. Here we report a complete reannotation of the genome based on extensive RNA-seq data. This updated version of the genome annotation includes information about transcript isoforms and integrates non-coding loci such as microRNAs (miRNAs) and long non-coding RNAs (lncRNAs). The large-scale assembly of the genome has also been considerably improved using a high density, RAD-seq-based genetic map to anchor sequence scaffolds onto the chromosomes. Finally, we also report additional resources including a genome-wide set of single nucleotide polymorphisms for genetic mapping and improvements to the genome database such as the addition of a JBrowse-based genome browser that allows multiple types of genome-wide data to be visualised simultaneously.

Results

Reannotation of gene structure based on RNA-seq data

The initial set of *Ectocarpus* gene models (referred to hereafter as the v1 annotation) was generated using EuGène (Schiex *et al.*, 2001) based on a limited amount of transcriptomic information (91,041 Sanger expressed sequence tags, ESTs) together with information such as genomic sequence composition, splice site predictions and sequence similarity to coding regions from other species (Cock *et al.*, 2010). The available EST data did not cover all the predicted genes and many genes were only covered partially. Consequently, the v1 annotation was based to a large extent on *de novo* predictions and only included limited information about the untranslated regions (UTRs) of the genes. The v1 annotation has been gradually improved since 2010 by the addition of 325 and 410 new functional and structural annotations, respectively, for individual genes through the Orcae database (http://bioinformatics.psb.ugent.be/orcae/overview/Ectsi; Sterck *et al.*, 2012). Most of the new structural annotations were based on more complete transcriptome information obtained through RNA-seq analyses that had been carried out since the publication of the genome sequence. This gene-by-gene approach improved the quality for a number of selected genes but it was necessary to extend the approach to improve annotation quality across the whole genome.

A genome-wide reannotation was therefore carried out based on the analysis of 507 million base pairs of RNA-seq data from ten libraries corresponding to three stages of the *Ectocarpus* life cycle: partheno-sporophyte (this study) and young and mature gametophytes (Ahmed *et al.*, 2014; Lipinska *et al.*, 2015; Table S1). Note that this genome-wide reannotation integrated the results of the manual gene-by-gene annotation carried out since publication of the v1 annotation by preferentially retaining high quality, expert functional and structural annotations during the integration of the genome-wide analysis into the database. The final result of the consolidation of these analyses will be referred to hereafter as the v2 annotation.

To improve the prediction of gene structure genome-wide, the 507 million base pairs of RNA-seq data was assembled into 34,551 *de novo* transcripts using Trinity (Grabherr *et al.*, 2011). GenomeThreader (Gremme *et al.*, 2005) was able to align 91% of these transcripts to the genome, often with multiple, alternative transcripts mapping to a single gene locus. Gene prediction for the v2 annotation was then carried out using EuGène and the 34,551 *de novo* transcripts, along with 83,502 Sanger ESTs (after cleaning) and SpliceMachine (Degroeve *et al.*, 2005) splice site predictions. The 21,958 gene models generated by this prediction were then compared with the 16,256 genes of the v1 annotation (Cock *et al.*, 2010) using AEGeAn (Standage & Brendel, 2012) and a combination of automatic and manual criteria were used to select the optimal gene model for each locus.

The gene predictions produced by the RNA-seq-based analysis fell into three groups: 1) loci in which the exon structure of the coding sequence was identical with that predicted by the v1 annotation (10,426 genes), 2) loci where the predicted coding sequence exon structure was different to that of the v1 annotation (5,336 genes) and 3) novel loci that were not predicted by the v1 annotation (5,237 genes). For the first set of genes, the v1 gene models could be replaced directly with the RNA-seq-based models, providing considerable additional information about the UTR structure of the genes (addition or extension of UTRs for 5,661 of the 10,426 genes, e.g. Figure 1A). For the second set of genes, if the RNA-seq-based prediction had a similar structure to the v1 annotation but contained modified or additional exons (e.g. Figure 1B), it was retained. RNA-seq-based models with a similar structure but which predicted fewer exons than the v1 annotation were only retained if the predicted protein shared more than 65% with the v1 protein for loci with four or more exons or more than 30% identity for loci with less than four exons. This second set of genes also included predictions which indicated that v1 annotation genes needed to be fused (e.g. Figure 1C) or split (e.g. Figure 1D). For these loci, the v1 and RNA-seq-based predictions were inspected manually to select the optimal model for each locus. Finally, the 5,237 RNA-seq-based predictions that corresponded to loci that had not been identified by the v1 annotation were filtered to remove

probable false positives. Predictions were retained only if 1) their transcripts had an abundance of >1 RPKM across the entire set of RNA-seq data, 2) the start codon of the gene was not located in a repeated region (to exclude transposon-derived ORFs; Yandell & Ence, 2012) and 3) their coding region was >100 bp. After applying these filters, 2,030 of the 5,237 new predictions were retained and integrated in the v2 annotation.

Overall, the addition of these new genes and updates to the existing genes (fusing or splitting existing gene models) brought the total number of genes in the v2 annotated genome to 17,407. Compared with the v1 annotation, the final v2 annotation involved modifications to 11,035 gene loci, including 5,383 modifications that affected the exon structure of the coding sequence and 5,652 that only involved adding information about the UTRs. Of the former, 831 involved gene fusions (to produce 402 genes in the v2 annotation), 18 involved splitting v1 annotation gene predictions to create 37 genes in the v2 annotation and 120 genes were deleted. The v2 annotation now includes coordinates for at least one of the UTR regions for 78.7% of the 17,407 genes (compared to 52.6% for the v1 annotation; Figure 2). The v2 annotation is publically available through the ORCAE database (http://bioinformatics.psb.ugent.be/orcae/overview/Ectsi; Sterck *et al.*, 2012).

Prediction of gene function

The final 17,407 genes of the v2 annotation were further analysed to improve the prediction of gene function. This was carried out by comparing the predicted proteins with the InterPro database using InterProScan. For the 2030 new genes we obtained 212 matches with the database that allowed 135 new annotations including 79 associated GO terms. Overall, whilst only 5,583 of the 16,256 genes of the v1 annotation had been annotated (5,989 with associated GO terms), a total of 10,688 genes now have functional annotations in the v2 annotation (7,383 with associated GO terms).

Integration of non-protein-coding genes

With the exception of tRNA loci (Cock *et al.*, 2010), the v1 annotation provided very little information about non-protein-coding genes. The v2 annotation includes considerably more information about this type of locus, in particular integrating 64 microRNA loci, eight ribosomal RNA loci and 656 small nucleolar RNA loci recently predicted by Tarver et al. (2015).

Improved chromosome-scale assembly of the genome scaffolds using a high-density, RAD-seq-based genetic map

A microsatellite-based genetic map (Heesch *et al.*, 2010) was originally used to produce a large-scale assembly of the *Ectocarpus* genome consisting of 34 pseudo-chromosomes (Cock *et al.*, 2010) corresponding to the 34 linkage groups of the genetic map. The pseudo-chromosomes were constructed by concatenating sequence scaffolds based on the genetic order of sequence-anchored microsatellite markers on the genetic map (Cock *et al.*, 2010). However, due to the low density of the markers, the large-scale assembly included only 325 of the 1,561 sequence scaffolds (70.1% of the total sequence length) and, moreover, only 40 (12%) of the mapped scaffolds could be orientated relative to the chromosome (i.e. only 12% of the scaffolds contained at least two microsatellite markers which recombined relative to each other).

To improve the large-scale assembly of the *Ectocarpus* genome, we took advantage of a high-density, single nucleotide polymorphism (SNP)-based genetic map that has recently been generated using a Restriction site associated DNA (RAD)-seq method (unpublished data). The 4,207 SNP markers used to construct the genetic map were mapped to sequence scaffolds and the recombination information for these markers used to construct a new set of pseudo-chromosomes.

The new large-scale assembly represents a significant improvement because it includes 530 of the 1,561 sequence scaffolds (90.5% of the total sequence length) and 49% of the scaffolds have been orientated with respect to their chromosome. Moreover, the high-density map has allowed several fragmented linkage groups / pseudo-chromosomes to be assembled, reducing the total number from 34 to 28. The exact number of chromosomes in *Ectocarpus* sp. strain Ec32 is not known but cytogenetic analysis of European strains of another *Ectocarpus* species, *E. siliculosus* indicated the presence of approximately 25 chromosomes (Müller, 1966; Müller, 1967).

The genome of *Ectocarpus* strain Ec32 contains an integrated copy of a large DNA virus, closely related to the *Ectocarpus* phaeovirus EsV-1 (Cock *et al.*, 2010). Microarray analysis had shown that all the viral genes were silent and the RNA-seq data analysed here confirmed this observation, indicating complete silencing of this region of the chromosome under all conditions analysed (Figure S1).

The *Ectocarpus* genome database has been modified to take into account the large-scale assembly of the sequence scaffolds. For example, in the v2 annotation, sequentially numbered locusID have been assigned to genes to indicate their position along each pseudochromosome.

Structure of the sex chromosome

Linkage group 30 of the v1 assembly was recently shown to carry the sex-determining region and therefore to correspond to the sex chromosome in *Ectocarpus* (Ahmed *et al.*, 2014). This linkage group consisted of 20 scaffolds in the v1 assembly but has been considerably extended in the v2 assembly with the addition of a further 16 scaffolds, increasing the estimated physical length of the chromosome (cumulative scaffold length) from 4,994 to 6,933 kbp. The non-recombining sex-determining region was not affected by this update, as all the additional scaffolds are located in the pseudoautosomal regions of the chromosome. However, as we have recently described a number of unusual features of the pseudoautosomal regions (Luthringer *et al.*, 2015), we verified that these observations were still valid for the updated version of the chromosome. This analysis confirmed that the updated pseudoautosomal regions continue to exhibit a number of structural features that are intermediate between those of the autosomes and the sex-determining region. In particular, compared with the autosomes, the updated pseudoautosomal regions still exhibit significantly reduced gene density, increased content of transposable element sequences, lower %GC content and the genes had significantly smaller and fewer exons (supplementary Figure S2). The conclusions of the Luthringer *et al*. study therefore remain valid for the updated version of the sex chromosome.

A genome-wide sequence variant resource for genetic analysis of brown algal gene function

One of the major objectives of the *Ectocarpus* genome project was to facilitate the investigation of gene function in the brown algae. To create an additional genetic resource for gene mapping in *Ectocarpus*, a genome re-sequencing approach was used to identify sequence variants (single nucleotide polymorphisms, SNPs, and indels) across the entire genome. DNA-seq sequence data was generated for the genome of the female outcrossing line Ec568 (Heesch *et al.*, 2010; Peters *et al.*, 2010) and this data was compared with the reference genome of the male strain Ec32 (Cock *et al.*, 2010) plus the female sex-determining region from the Ec32-related strain Ec597 (Ahmed *et al.*, 2014). Hi-seq2500 Illumina technology was used to generate 25,976,388,600 bp of 2x100 bp paired-end, sequence reads for the female outcrossing line Ec568, corresponding to 121x genome coverage. The sequence reads were mapped onto the reference genome scaffolds using Bowtie2 and sequence variants were identified by combining the output of three different variant predictors: Samtools mpileup and bcftools, SHORE qVar and the GATK UnifiedGenotyper. The 340,665 high quality sequence variants identified using this approach are listed in Table S2.

To further validate the sequence variants as potential genetic markers, we used a bulked segregant approach to determine whether they behaved as Mendelian loci. A population of segregating progeny was generated by crossing a UV-mutagenised derivative of the reference genome strain Ec32 (strain Ec722) with the female outcrossing line Ec568. One hundred and eighty haploid gametophyte progeny, each corresponding to an independent meiotic event, were obtained from the resulting diploid sporophyte and sorted according to phenotype. Genomic DNA from two pools corresponding to 96 wild type and 84 mutant individuals were sequenced on an Illumina platform to generate 23,429,143,400 bp and 20,785,058,400 bp of 2x100 bp paired-end sequence, respectively. Each dataset was then independently mapped against the genome sequence scaffolds of the Ec32 reference strain (Cock *et al.*, 2010) plus the scaffolds for the female sex-determining region from strain Ec597 (Ahmed *et al.*, 2014) and two lists of variants were generated using SHORE consensus. We retained only variants that were shared by these two lists (i.e. identified in both the wild type and mutant pools) and applied an additional filter, retaining only variants for which the sum of the two observed frequencies was 1 ±0.2. This strategy allowed the identification of 390,804 sequence variants that exhibited a 1:1 segregation pattern in the progeny population and were therefore behaving as Mendelian loci. Using this list, 237,839 of the 340,665 sequence variants obtained by mapping the Ec568 DNA-seq data against the reference scaffolds (see above) were validated as Mendelian genetic markers (Table S2). The average distance between adjacent pairs of the 237,839 potential genetic markers is 823 bp, providing a high-density resource for genetic analysis in this species.

Extension and improvement of the *Ectocarpus* genome database

The v1 annotation of the *Ectocarpus* genome has been publically available on the Orcae database (http://bioinformatics.psb.ugent.be/orcae/overview/EctsiV2) (Sterck *et al.*, 2012) since its publication in 2010. We have replaced the v1 annotation with the v2 annotation described in this study to make the latter broadly available. In addition, a Jbrowse genome browser has been created (http://mmodev.sb-roscoff.fr/jbrowse/index.html?data=data/public/ectocarpus/) to allow simultaneous visualisation of multiple types of data in a genome context. The Jbrowse genome browser provides access to both the v1 and v2 annotations, raw Eugène and Cufflinks gene predictions, EST and RNA-seq transcript data, raw RNA-seq data for messenger RNAs and small RNAs, genetic makers including microsatellites and SNP markers, micro-array data and tiling array data. The Jbrowse genome browser is intended to be complementary to the Orcae database, providing an environment where users can compile and compare diverse datasets. It is also possible for registered users to create private versions of the browser in order to upload unpublished and working datasets.

Discussion

The *Ectocarpus* genome has become an important resource, both for scientists working on this filamentous brown alga as a model organism and as a keystone genome for the relatively poorly characterised stramenopile lineage within the eukaryotic tree. The work described here has significantly increased the quality of this resource in several respects. Extensive RNA-seq data has been used to improve 12,160 existing gene models, to identify 2,030 new protein coding genes and to determine the abundance and nature of alternative transcripts of these genes. The non-protein-coding part of the genome has also been characterised, notably with the inclusion of a genome-wide catalogue of lncRNA loci. In addition, a high-density, RAD-seq-based genetic map was used to significantly improve the large-scale assembly of the genome and a genome-wide SNP resource has been developed for future genetic analyses. These updated and new resources have been integrated

into the *Ectocarpus* genome database, which has also been improved to facilitate exploitation of the genome data and associated information.

With the integration of the new information and resources described here, the *Ectocarpus* genome represents one of the most extensively annotated genomes within the stramenopile group and, as such, will serve as an important reference genome for future genome analysis projects. Recently, the *Ectocarpus* genome provided a reference for the analysis of the larger and more complex genome of the kelp *Saccharina japonica* (Ye *et al.*, 2015) and similar comparisons are expected in the future as part of the many ongoing brown algal and stramenopile genome projects.

Methods

Biological material

Ectocarpus strains were cultured at 13°C in autoclaved natural sea water (NSW) supplemented with half-strength Provasoli solution (Starr & Zeikus, 1993) with a light:dark cycle of 12h:12h (20 µmol photons m^{-2} s^{-1}) using daylight-type fluorescent tubes. All manipulations were performed under a laminar flow hood under sterile conditions. The male genome sequenced strain Ec32 is a meiotic offspring of a field sporophyte, Ec17, collected in 1988 in San Juan de Marcona, Peru (Peters *et al.*, 2008). Ec722 is a UV-mutagenised descendant of Ec32. The female outcrossing line Ec568 is derived from a sporophyte collected in Arica in northern Chile (Heesch *et al.*, 2010).

RNA-seq

The analyses carried out in this study used RNA-seq data generated for biological replicate (duplicate) samples of partheno-sporophytes and of both young and mature samples for both male and female gametophytes (ten libraries in all). The production of the young (Lipinska *et al.*, 2015) and mature (Ahmed *et al.*, 2014) gametophyte RNA-seq data (100 bp Illumina HiSeq 2000 single-end reads) has been described previously. For each of the replicate partheno-sporophyte samples, total RNA was extracted and used as a template by Fasteris (CH-1228 Plan-les-Ouates, Switzerland) to synthesise cDNA using an oligo-dT primer. The cDNA libraries were sequenced with Illumina HiSeq 2000 technology to generate 100 bp single-end reads. Data quality was assessed using the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html) and the reads were trimmed and filtered using a quality threshold of 25 (base calling) and a minimal size of 60 bp. Only reads in which more than 75% of nucleotides had a minimal quality threshold of 20 were retained. Table S1 shows the number of raw reads generated per sample and the number of reads remaining after trimming and filtering (cleaned reads). The cleaned reads were mapped to the *Ectocarpus* sp. genome (Cock *et al.*, 2010) (available at ORCAE; Sterck *et al.*, 2012) using TopHat2 with the bowtie2 aligner (Kim *et al.*, 2013). More than 90% of the sequencing reads for each library mapped to the genome.

De novo assembly of the pooled RNA-seq data from the ten libraries was carried out using Trinity (Grabherr *et al.*, 2011) in normalized mode with default parameters. Weakly expressed transcripts (isoform percentage <1 and RPKM <1) were removed from the dataset. The remaining transcripts were aligned against the *Ectocarpus* reference genome (Ec32) using GenomeThreader (Gremme *et al.*, 2005) with a maximum intron length of 26,000 bp, a minimum coverage of 75% and a minimum alignment score of 90%.

Gene prediction and comparison of the RNA-seq-based gene predictions with the v1 annotation

Gene prediction was carried out using the EuGène program (Schiex *et al.*, 2001), as described previously (Cock *et al.*, 2010). In addition to the previous data used for the first annotation, alignments of the Trinity RNA-seq-derived transcripts against the *Ectocarpus* sp. reference genome were added to the Eugène pipeline. The mapped Trinity transcripts were compared with the gene

structures of the v1 annotation using AEGeAn (Standage & Brendel, 2012) and a combination of automated and manual approaches used to select the optimal gene structures.

Functional annotation of new gene models

Functional annotation of the new predicted gene models was carried out based on the identification of protein domains using the InterProScan (Jones *et al.*, 2014).

Detection of non-protein-coding genes

The detection of microRNA, ribosomal RNA and snoRNA loci has been described previously (Tarver *et al.*, 2015).

Improvement of the large-scale genome assembly based on a RAD-seq-based genetic map

To improve the large-scale assembly of the *Ectocarpus* genome scaffolds, the SNP markers recently used to generate a high-density, RAD-seq-based genetic map (unpublished data) were located on the sequence scaffolds and the linkage information for these markers used to order the sequence scaffolds into pseudochromosomes and to orientate scaffolds with respect to their corresponding pseudochromosome.

Genome-wide identification of sequence variants to generate a genetic marker resource

Genome sequence data for the female outcrossing line Ec568 (CCAP 1310⁄334, isolated from Arica, northern Chile; Peters *et al.*, 2010) was generated using Illumina HiSeq2500 technology (Fasteris, Switzerland), which produced 25,976,388,600 bp of 2x100 bp paired-end sequence. Sequence variants were detected as described previously (Godfroy *et al.*, 2015). Prinseq (Schmieder & Edwards, 2011) was used to clean and trim the sequence data. Bases with quality scores less than 20 were trimmed from both ends and only reads with a mean quality of at least 25 and which were longer than 50 nucleotides after trimming were retained. Bowtie2 (Langmead *et al.*, 2009) was used to map the reads to a 196,942,248 bp reference genome sequence that consisted of the 1,561 scaffolds (195.8 Mbp) of the Ec32 genome (Cock *et al.*, 2010) plus 39 scaffolds (0.9 Mbp) corresponding to the female haplotype of the sex-determining region from strain Ec597 (Ahmed *et al.*, 2014). The Indel Realigner and Base Score Recalibration programs of the GATK suite (McKenna *et al.*, 2010; DePristo *et al.*, 2011) were then implemented to improve read alignment and quality parameters, respectively. The Samtools depth program was used to estimate sequencing depth per base based on the mapping. Sequence variants were identified using a combination of three different variant-calling programs: Samtools mpileup and bcftools, SHORE qVar and the GATK UnifiedGenotyper. The following filters were then applied to retain only high quality sequence variants: variant loci were selected if 1) coverage was to a depth of between 20 and 50, 2) the variant sequence was at a frequency of 0.95 or higher and 3) the Phred-scaled variant quality score was over 50. These filters were either applied during variant calling (SHORE qVar) or afterwards (Samtools mpileup and Unified Genotyper) using bcftools. VCF files were then compared using the VFCtools suite (vcf-isec command) and only sequence variants identified by at least two of the programs were retained.

To determine whether sequence variants behaved as Mendelian loci, a cross between a UV-mutagenised derivative of the reference genome strain Ec32 (strain Ec722) and the female outcrossing line Ec568 (Heesch *et al.*, 2010) was used to generate a population of 180 progeny segregating the two parental alleles of each of the variant loci. Two libraries were constructed with pools of 84 and 96 haploid, partheno-sporophyte individuals and sequenced using Illumina HiSeq2500 technology (Fasteris, Switzerland) to generate 20,785,058,400 bp and 23,429,143,400 bp

of 2x100 bp paired-end sequences, respectively. The reads were trimmed and cleaned with Prinseq using the parameters described above. Bowtie2 was used to map the reads to the genome sequence scaffolds of the Ec32 reference strain (Cock *et al.*, 2010) plus the scaffolds for the female sex-determining region from strain Ec597 (Ahmed *et al.*, 2014) and the mapping quality was then improved using GATK Indel Realigner and Base Score Recalibration. SHORE convert was used to re-format the mapping files and variants were called using the SHORE consensus program to generate, for each of the two libraries, three files containing SNPs, deletions and insertions. The SNP variants were retained for identification of Mendelian genetic markers. The VarScan compare tool was used to identify SNPs shared by the two pools of haploid individuals. For each of these SNPs the sum of the variant frequencies observed in the two pools was calculated, and only those for which this sum was between 0.8 and 1.2 were retained. VarScan compare was then used to extract the Ec568 variants from the list of Mendelian segregating SNPs.

Database curation of the v2 annotation

A Genome Browser was implemented based on Jbrowse (Standage & Brendel, 2012) using a Chado database (Mungall & Emmert, 2007). The browser integrates both v1 and v2 reference gene models, raw gene models predicted by Eugène, transcripts predicted by Cufflinks and EST and RNA-seq read data.

References

Ahmed S, Cock JM, Pessia E, Luthringer R, Cormier A, Robuchon M, Sterck L, Peters AF, Dittami SM, Corre E, Valero M, Aury JM, Roze D, Van de Peer Y, Bothwell J, Marais GA, Coelho SM. 2014. A Haploid System of Sex Determination in the Brown Alga *Ectocarpus* sp. *Curr Biol* **24**(17): 1945-1957.

Bartsch I, Wiencke C, Bischof K, Buchholz C, Buck B, Eggert A, Feuerpfeil P, Hanelt D, Jacobsen S, Karez R, Karsten U, Molis M, Roleda M, Schubert H, Schumann R, Valentin K, Weinberger F, Wiese J. 2008. The genus *Laminaria sensu lato*: recent insights and developments. *Eur J Phycol* **43**: 1-86.

Charrier B, Coelho S, Le Bail A, Tonon T, Michel G, Potin P, Kloareg B, Boyen C, Peters A, Cock J. 2008. Development and physiology of the brown alga *Ectocarpus siliculosus*: two centuries of research. *New Phytol* **177**(2): 319-332.

Cock JM, Collén J 2015. Independent emergence of complex multicellularity in the brown and red algae. In: Ruiz-Trillo I, Nedelcu AM eds. *Evolutionary transitions to multicellular life*: Springer Verlag, 335-361.

Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, Amoutzias G, Anthouard V, Artiguenave F, Aury J, Badger J, Beszteri B, Billiau K, Bonnet E, Bothwell J, Bowler C, Boyen C, Brownlee C, Carrano C, Charrier B, Cho G, Coelho S, Collén J, Corre E, Da Silva C, Delage L, Delaroque N, Dittami S, Doulbeau S, Elias M, Farnham G, Gachon C, Gschloessl B, Heesch S, Jabbari K, Jubin C, Kawai H, Kimura K, Kloareg B, Küpper F, Lang D, Le Bail A, Leblanc C, Lerouge P, Lohr M, Lopez P, Martens C, Maumus F, Michel G, Miranda-Saavedra D, Morales J, Moreau H, Motomura T, Nagasato C, Napoli C, Nelson D, Nyvall-Collén P, Peters A, Pommier C, Potin P, Poulain J, Quesneville H, Read B, Rensing S, Ritter A, Rousvoal S, Samanta M, Samson G, Schroeder D, Ségurens B, Strittmatter M, Tonon T, Tregear J, Valentin K, von Dassow P, Yamagishi T, Van de Peer Y, Wincker P. 2010. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature* **465**(7298): 617-621.

Coelho SM, Godfroy O, Arun A, Le Corguillé G, Peters AF, Cock JM. 2011. *OUROBOROS* is a master regulator of the gametophyte to sporophyte life cycle transition in the brown alga *Ectocarpus*. *Proc Natl Acad Sci U S A* **108**: 11518-11523.

Coelho SM, Scornet D, Rousvoal S, Peters N, Dartevelle L, Peters AF, Cock JM. 2012. *Ectocarpus*: A model organism for the brown algae. *Cold Spring Harbor Protoc* **2012**: 193-198.

Dayton P. 1985. Ecology of Kelp Communities. *Annu Rev Ecol Syst* **16**: 215–245.

Degroeve S, Saeys Y, De Baets B, Rouzé P, Van de Peer Y. 2005. SpliceMachine: predicting splice sites from high-dimensional local context representations. *Bioinformatics* **21**(8): 1332-1338.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**(5): 491-498.

Dittami S, Scornet D, Petit J, Ségurens B, Da Silva C, Corre E, Dondrup M, Glatting K, König R, Sterck L, Rouzé P, Van de Peer Y, Cock J, Boyen C, Tonon T. 2009. 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**(6): R66.

Godfroy O, Peters AF, Coelho SM, Cock JM. 2015. Genome-wide comparison of ultraviolet and ethyl methanesulphonate mutagenesis methods for the brown alga Ectocarpus. *Mar Genomics*.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**(7): 644-652.

Gremme G, Brendel V, Sparks ME, Kurtz S. 2005. Engineering a software tool for gene structure prediction in higher organisms. *Information and Software Technology* **47**(15): 965-978.

Gschloessl B, Guermeur Y, Cock J. 2008. HECTAR: a method to predict subcellular targeting in heterokonts. *BMC Bioinf* **9**: 393.

Heesch S, Cho GY, Peters AF, Le Corguillé G, Falentin C, Boutet G, Coëdel S, Jubin C, Samson G, Corre E, Coelho SM, Cock JM. 2010. A sequence-tagged genetic map for the brown alga *Ectocarpus siliculosus* provides large-scale assembly of the genome sequence. *New Phytol* **188**(1): 42-51.

Hughes AD, Kelly MS, Black KD, Stanley MS. 2012. Biogas from Macroalgae: is it time to revisit the idea? *Biotechnol Biofuels* **5**: 86.

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**(9): 1236-1240.

Kijjoa A, Sawangwong P. 2004. Drugs and Cosmetics from the Sea. *Mar Drugs* **2**: 73–82.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**(4): R36.

Klinger T. 2015. The role of seaweeds in the modern ocean. *Perspect Phycol* **2**: 31-39.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**(3): R25.

Le Bail A, Billoud B, Le Panse S, Chenivesse S, Charrier B. 2011. *ETOILE* Regulates Developmental Patterning in the Filamentous Brown Alga *Ectocarpus siliculosus*. *Plant Cell* **23**: 1666-1678.

Lipinska A, Cormier A, Luthringer R, Peters AF, Corre E, Gachon CM, Cock JM, Coelho SM. 2015. Sexual Dimorphism and the Evolution of Sex-Biased Gene Expression in the Brown Alga *Ectocarpus*. *Mol Biol Evol*.

Lipinska AP, D'hondt S, Van Damme EJM, De Clerck O. 2014. Uncovering the genetic basis for early isogamete differentiation: a case study of *Ectocarpus siliculosus*. *BMC Genomics* **in press**.

Luthringer R, Lipinska A, Cormier A, Peters AF, Roze D, Cock JM, Coelho SM. 2015. The pseudoautosomal region of the *Ectocarpus* UV sex chromosome. *Mol Biol Evol* **in press**.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**(9): 1297-1303.

Meslet-Cladière L, Delage L, Leroux CJ, Goulitquer S, Leblanc C, Creis E, Gall EA, Stiger-Pouvreau V, Czjzek M, Potin P. 2013. Structure/Function Analysis of a Type III Polyketide Synthase in the Brown Alga *Ectocarpus siliculosus* Reveals a Biochemical Pathway in Phlorotannin Monomer Biosynthesis. *Plant Cell* **25**(8): 3089-3103.

Müller DG. 1966. Untersuchungen zur Entwicklungsgeschichtebder Braunalge *Ectocarpus siliculosus* aus Neapel. *Planta* **68**: 57-68.

Müller DG. 1967. Generationswechsel, Kernphasenwechsel und Sexualität der Braunalge *Ectocarpus siliculosus* im Kulturversuch. *Planta* **75**: 39-54.

Mungall CJ, Emmert DB. 2007. A Chado case study: an ontology-based modular schema for representing genome-associated biological information. *Bioinformatics* **23**(13): i337-346.

Peters AF, Mann AD, Córdova CA, Brodie J, Correa JA, Schroeder DC, Cock JM. 2010. Genetic diversity of *Ectocarpus* (Ectocarpales, Phaeophyceae) in Peru and northern Chile, the area of origin of the genome-sequenced strain. *New Phytol* **188**(1): 30-41.

Peters AF, Marie D, Scornet D, Kloareg B, Cock JM. 2004. Proposal of *Ectocarpus siliculosus* (Ectocarpales, Phaeophyceae) as a model organism for brown algal genetics and genomics. *J Phycol* **40**(6): 1079-1088.

Peters AF, Scornet D, Ratin M, Charrier B, Monnier A, Merrien Y, Corre E, Coelho SM, Cock JM. 2008. Life-cycle-generation-specific developmental processes are modified in the *immediate upright* mutant of the brown alga *Ectocarpus siliculosus*. *Development* **135**(8): 1503-1512.

Prigent S, Collet G, Dittami SM, Delage L, Ethis de Corny F, Dameron O, Eveillard D, Thiele S, Cambefort J, Boyen C, Siegel A, Tonon T. 2014. The genome-scale metabolic network of Ectocarpus siliculosus (EctoGEM): a resource to study brown algal physiology and beyond. *Plant J* **80**(2): 367-381.

Ritter A, Ubertini M, Romac S, Gaillard F, Delage L, Mann A, Cock JM, Tonon T, Correa JA, Potin P. 2010. Copper stress proteomics highlights local adaptation of two strains of the model brown alga *Ectocarpus siliculosus*. *Proteomics* **10**(11): 2074-2088.

Schiex T, Moisan A, Rouzé P 2001. EuGène: An Eucaryotic Gene Finder that combines several sources of evidence. In: Gascuel O, Sagot M-F eds. *Lect. Notes in Comput. Sci. 2066*, 111-125.

Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**(6): 863-864.

Smit AJ. 2004. Medicinal and pharmaceutical uses of seaweed natural products: A review. *J Appl Phycol* **16**: 245–262.

Standage DS, Brendel VP. 2012. ParsEval: parallel comparison and analysis of gene structure annotations. *BMC Bioinformatics* **13**: 187.

Starr RC, Zeikus JA. 1993. UTEX-The culture collection of algae at the University of Texas at Austin. *J Phycol* **29 (Suppl.)**: 1-106.

Steneck RS, Graham MH, Bourque BJ, Corbett D, Erlandson JM, Estes JA, Tegner MJ. 2002. Kelp forest ecosystems: biodiversity, stability, resilience and future. *Environ Conserv* **29**: 436–459.

Sterck L, Billiau K, Abeel T, Rouzé P, Van de Peer Y. 2012. ORCAE: online resource for community annotation of eukaryotes. *Nat Methods* **9**(11): 1041.

Tarver JE, Cormier A, Pinzón N, Taylor RS, Carré W, Strittmatter M, Seitz H, Coelho SM, Cock JM. 2015. microRNAs and the evolution of complex multicellularity: identification of a large, diverse complement of microRNAs in the brown alga *Ectocarpus*. *Nucl Acids Res* **43**: 6384-6398.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**(5): 511-515.

Tseng C. 2001. Algal biotechnology industries and research activities in China. *J. Appl. Phycol.* **13**: 375–380.

Wahl M, Molis M, Hobday AJ, Dudgeon S, Neumann R, Steinberg P, Campbell AH, Marzinelli E, Connell S. 2015. The responses of brown macroalgae to environmental change from local to global scales: direct versus ecologically mediated effects. *Perspect Phycol* **2**: 11 - 29.

Yandell M, Ence D. 2012. A beginner's guide to eukaryotic genome annotation. *Nat Rev Genet* **13**(5): 329-342.

Ye N, Zhang X, Miao M, Fan X, Zheng Y, Xu D, Wang J, Zhou L, Wang D, Gao Y, Wang Y, Shi W, Ji P, Li D, Guan Z, Shao C, Zhuang Z, Gao Z, Qi J, Zhao F. 2015. Saccharina genomes provide novel insight into kelp biology. *Nat Commun* **6**: 6986.

Zambounis A, Elias M, Sterck L, Maumus F, Gachon CM. 2012. Highly dynamic exon shuffling in candidate pathogen receptors... What if brown algae were capable of adaptive immunity? *Mol Biol Evol* **29**: 1263-1276.

Tables

Table 1. Overview of the modifications to the v1 annotation during the production of the v2 annotation of the *Ectocarpus* genome

	Number of genes
v1 and v2 gene models identical (cds level)	10,426
v1 gene model updated	5,336
New gene models in v2	2,030
v1 models fused in the v2	784
v1 models split in the v2	18
Models refactored	512
v1 gene model removed	123

Table 2. Comparison genome-wide statistics for the v1 and v2 annotations of the *Ectocarpus* genome

	Annotation-v1	Annotation-v2
Genes		
Number of coding genes	16,256	17,426
Mean gene length (bp)	6,859	7,542
Longest gene (bp)	122,137	123,931
Shortest gene (bp)	134	150
Exons		
Total number	129,875	138,690
Mean number per gene	7.3	7.96
Max number per gene	171	173
Mean length (bp)	242.2	299.80
Introns		
Total number	113,619	121,264
Mean length (bp)	703.8	738.87
Max length (bp)	25,853	36,147
UTRs		
Genes with 5' UTR	1,098	918
Genes with 3' UTR	4,766	3,056
Genes with 5'3' UTR	2,484	9,737
Genes without UTR	7,598	3,715
Mean length 5' UTR (bp)	120.60	139.61
Mean length 3' UTR (bp)	674.74	901.66

Gene annotation		
Genes with predicted functions	5,583	10,688
Genes with associated GO terms	5,989	7,383
miRNA loci	26	64
rRNA loci	n/a	5
snoRNA loci	n/a	656
lncRNA loci	n/a	2442

Figures

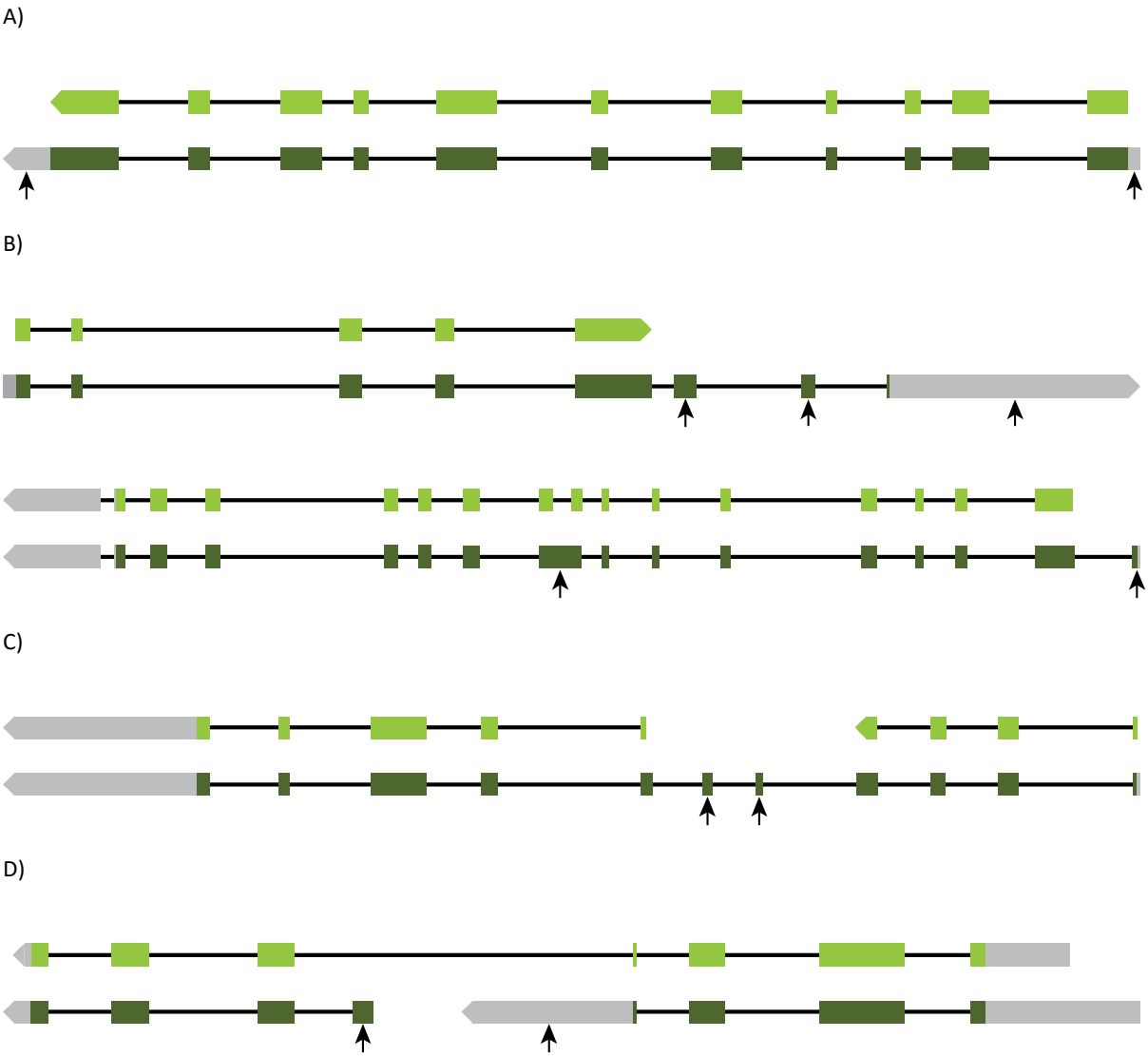


Figure 1. Representative comparisons of v1 and v2 annotation gene predictions illustrating the major types of annotation correction carried out during the transition between the two versions. Protein coding exons are in light or dark green for genome annotation versions v1 and v2, respectively, UTRs are in grey and introns are indicated by thin black lines. A) analysis of the RNA-seq data allowed the identification of UTRs for gene Ec-27_006370. B) v2 genes Ec-27_006520 and Ec-05_002440 have been extended and modified compared to their v1 equivalents. C) v1 genes Esi0002_0099 and Esi0002_0101 were fused to create a single locus, Ec-01_007860. D) v1 gene Esi0002_0311 was split to create two loci, Ec-01_006420 and Ec-01_006421. Arrows indicate gene features that were not identified or misidentified by the v1 annotation.

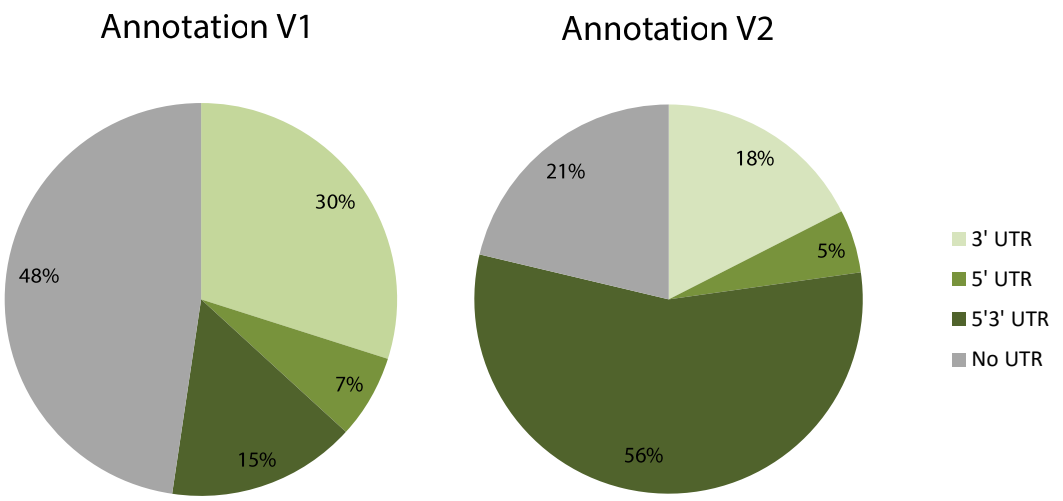


Figure 2. Comparison of the degree of completeness of gene annotations in the v1 and v2 versions of the *Ectocarpus* genome annotation.

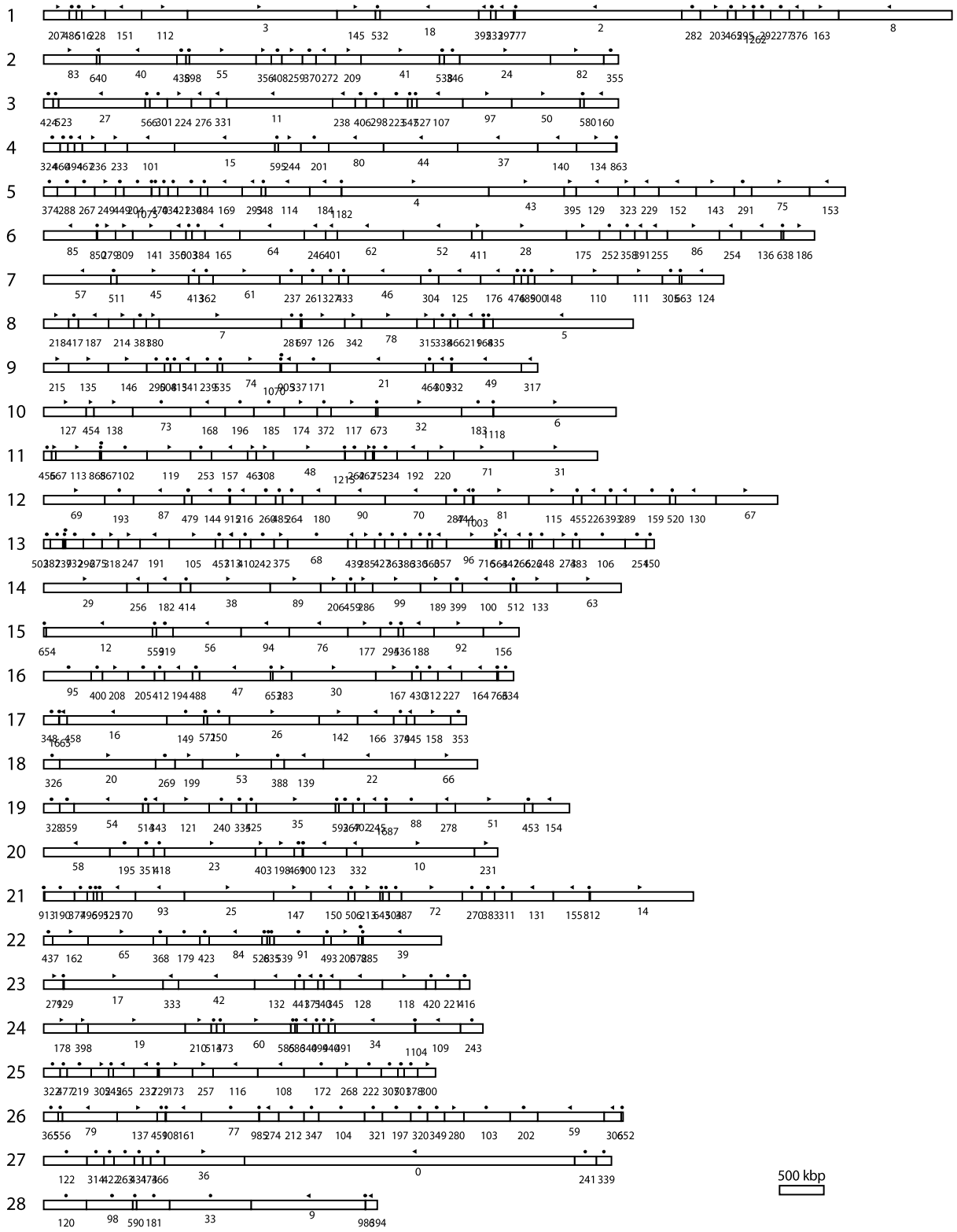


Figure 3. Large-scale assembly of the *Ectocarpus* scaffolds into pseudochromosomes based on a high-density, RAD-seq-based genetic map. Each bar represents one of the 28 chromosomes. Sequence scaffolds (supercontigs) are drawn to scale and identified with numbers (e.g. 207, sctg_207). Left or right pointing arrowheads indicate that the scaffolds have been orientated with respect to the chromosome (i.e. scaffolds with at least two markers separated by a recombination event); unorientated scaffolds are indicated with a spot.

Supplementary Figures

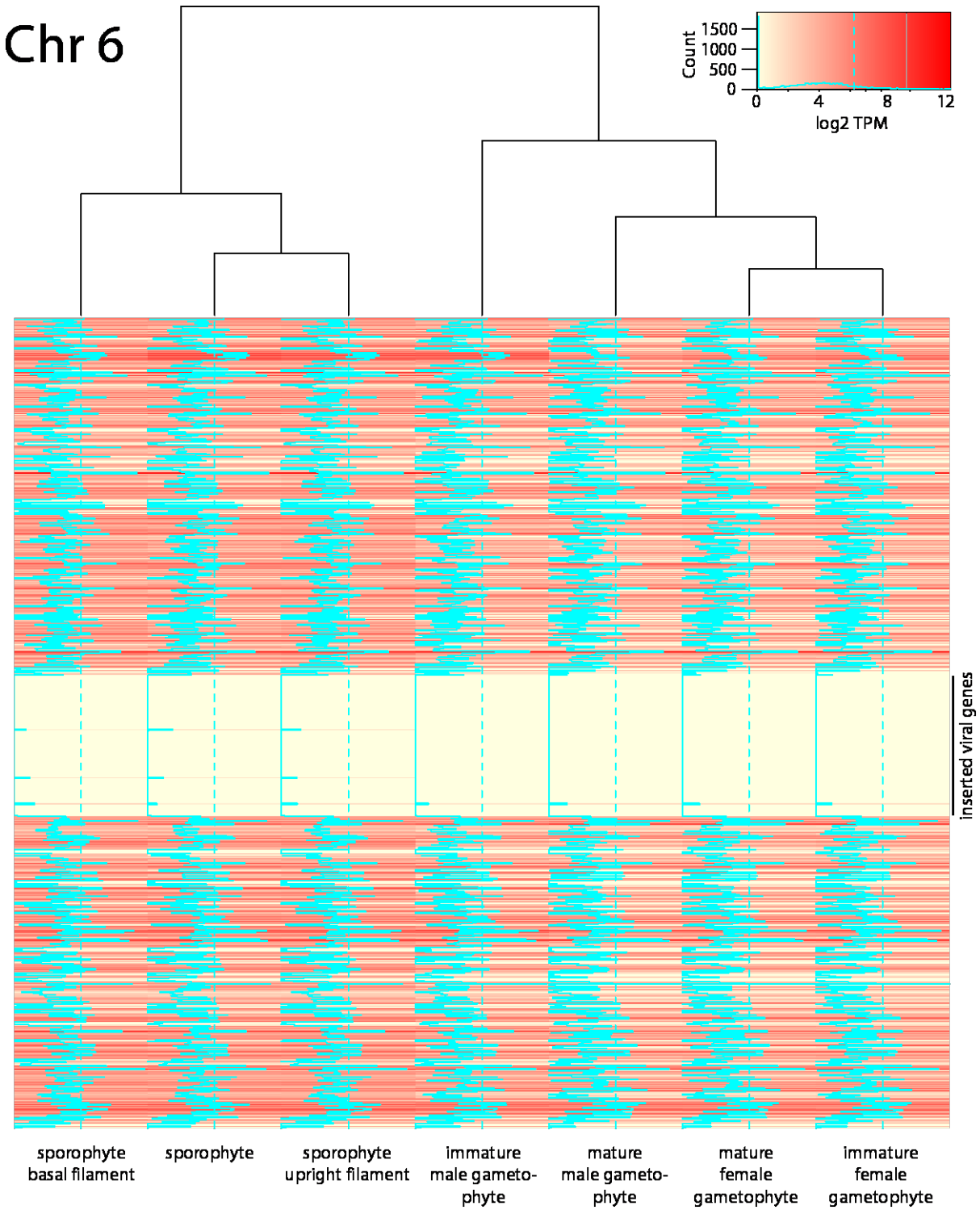


Figure S1. Suppressed transcription from a viral genome inserted into chromosome 6. Heatmap representing transcript abundances (RNA-seq log2 TPM) for all the genes on chromosome 6 in six different tissue samples.

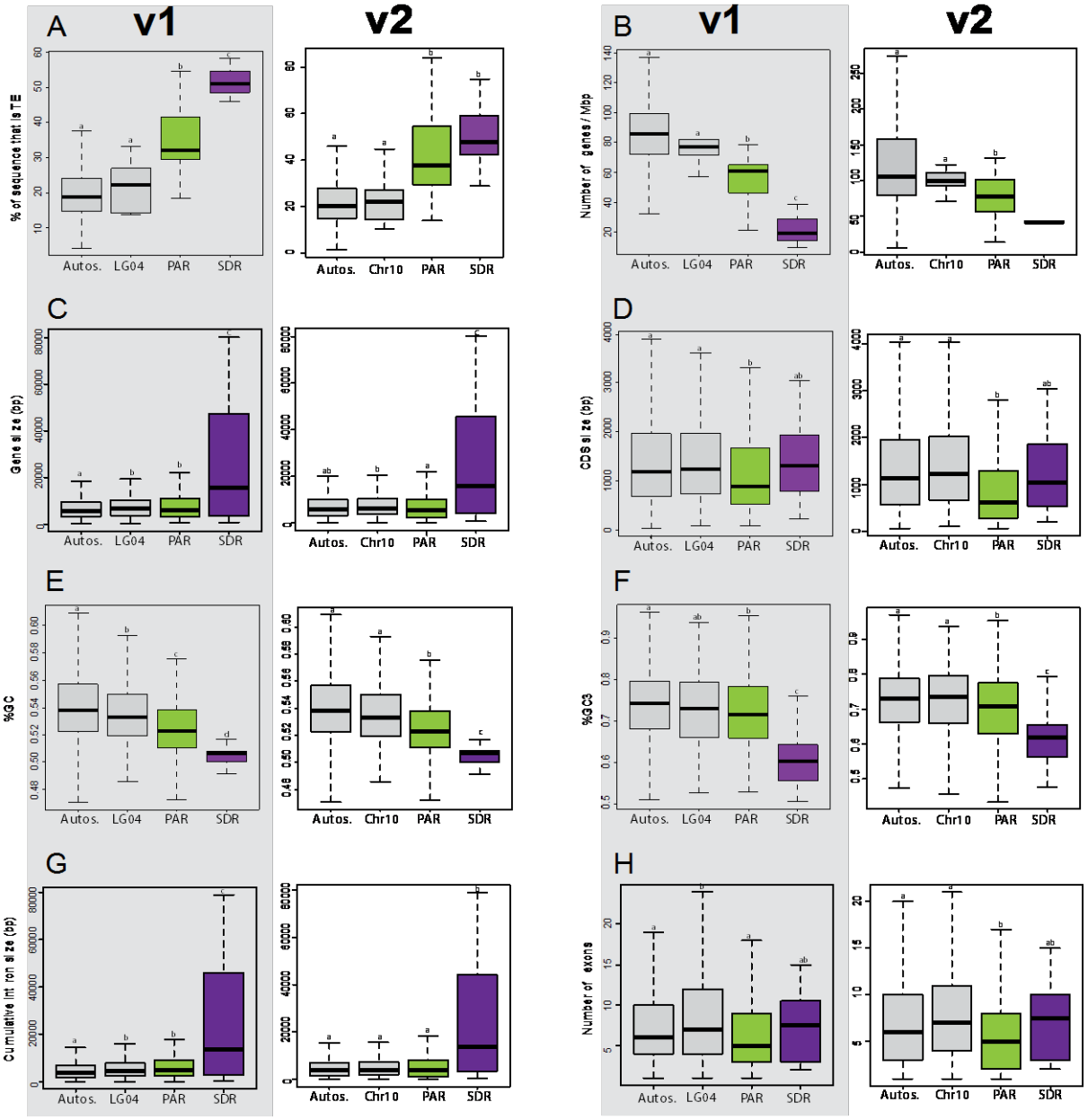


Figure S2. Comparisons of structural characteristics of the sex-determining and pseudoautosomal regions of the sex chromosome with both a representative autosome and with all autosomes for both the v1 and v2 versions of the *Ectocarpus* genome annotation. A) % TE calculated per supercontig; B) gene density per supercontig; C) gene size; D) CDS size; E) % GC per gene; F) %GC3 per coding sequence; G) total intron length per gene; H) number of exons per gene. Statistical differences were tested using the pairwise Mann-Whitney U-test; letters (a, b, c) shared between groups indicate no significant difference. Outlying points have been removed for clarity. Autos., autosomes; LG04, linage group 4; Chr10, chromosome 10; PAR, pseudoautosomal region; SDR, sex-determining region.

Supplementary methods

Manual annotation of genes through the Orcae database

The v2 annotation took into account the functional and structural annotation of 325 and 410 genes, respectively, carried out through the Orcae database (http://bioinformatics.psb.ugent.be/orcae/overview/Ectsi; Sterck *et al.*, 2012) since the publication of the v1 annotation. Many of the structural annotations were based on the same set of RNA-seq data that was used for the genome-wide gene structure prediction but exploited transcripts that had been generated using a reference-guided approach with Tophap2 and Cufflinks2 (Trapnell *et al.*, 2010; Kim *et al.*, 2013). Tophat2 was able to map 92% of the cleaned reads to the genome sequence and 36,565 transcripts were assembled by Cufflinks2 (including multiple transcripts for some loci) using the mapping information and the initial gene models as guides.

Supplementary tables

Table S1: *Ectocarpus* RNA-seq data used in this study

Species	Strain	Stage	Sex	Library	Raw reads	Cleaned reads	Genbank accession number	Reference
Ectocarpus sp.	Ec32	Mature gametophyte	Male	GPO-1	25,119,067	22,428,865	SRR1166429	(Ahmed *et al.*, 2014)
				GPO-2	26,873,490	23,642,187	SRR1166430	(Ahmed *et al.*, 2014)
			Female	GPO-3	21,005,896	18,668,732	SRR1166441	(Ahmed *et al.*, 2014)
				GPO-4	32,150,185	28,667,939	SRR1166452	(Ahmed *et al.*, 2014)
		Immature gametophyte	Male	GBP-24	80,602,259	78,459,187	SRR1660829	(Lipinska *et al.*, 2015)
				GBP-25	85,602,259	83,125,361	SRR1660830	(Lipinska *et al.*, 2015)
			Female	GBP-22	75,827,247	73,723,385	SRR1660827	(Lipinska *et al.*, 2015)
				GBP-23	93,562,945	90,903,680	SRR1660828	(Lipinska *et al.*, 2015)
		Partheno-sporophyte	n/a	GBP-7	37,221,214	37,018,065	n/a	This study
				GBP-8	29,670,293	29,491,668	n/a	This study

Discussion et perspectives

Le génome d'*Ecotcarpus* représente une importante ressource pour l'étude du groupe des algues brunes dans lequel relativement peu d'informations génomiques sont disponibles. Ce travail a permis de notablement augmenter la qualité de la ressource génomique d'*Ectocarpus*. D'une part, par une amélioration de l'assemblage du génome avec la mise à disposition d'une carte génétique hautement résolutive, permettant de fournir un assemblage du génome à grande échelle avec une organisation des super-contigs en pseudo-chromosomes. D'autre part, via l'amélioration des annotations structurales et fonctionnelles au niveau des anciens modèles de gènes, par une structure exonique plus fine et par une nette amélioration de la prédiction des UTR, mais aussi par l'annotation de nouveaux gènes. Enfin, la nouvelle annotation fonctionnelle des gènes a aussi permis d'améliorer les connaissances sur les fonctions des gènes chez *Ectocarpus*.

L'ensemble de ces informations a été intégré dans une nouvelle base de données hébergée au sein de la plateforme Orcae. De plus, les annotations des rRNA, snoRNA et miRNA elles sont désormais intégrées dans la plateforme d'annotation Orcae. La mise à jour de l'ensemble du génome et des annotations, ainsi que leur mise à disposition, permet de fournir l'un des génomes les mieux annotés au sein du groupe des algues brunes, pouvant servir de génome de référence pour des projets d'analyse de génomes.

Cependant, des travaux sont toujours en cours afin de proposer une annotation structurale la plus exhaustive possible. Deux points sont encore en cours d'analyses, la détection et la quantification des isoformes pour les mRNA et la recherche des lncRNA.

Dans le cas de l'annotation des isoformes, le pipeline utilisé comprend le mapping des reads, réalisé avec TopHat2, sur le génome et l'assemblage des reads en transcrits, avec StringTie. Après un premier filtre des résultats de StringTie, afin de ne conserver que les potentielles isoformes, il apparait que les gènes chez *Ectocarpus* ont en moyenne 1,6 isoformes. Il reste à vérifier manuellement que les prédictions de StringTie sont correctes, sur un sous échantillon de gènes, ainsi que d'analyser les changements induits par la présence de ces isoformes au niveau de la structure des protéines et de leur fonction.

La prédiction des lncRNA a été réalisée avec l'outil FEELnc (non publié) (https://github.com/tderrien/FEELnc) qui se base sur les résultats de StringTie afin de déterminer

parmi les transcrits assemblés, la possibilité qu'il s'agisse d'un lncRNA. Un total de 1828 lncRNA ont pu être identifiés, comprenant 2688 transcrits, soit environ 1,5 isoformes par lncRNA. La taille moyenne de ces lncRNA est de 1593pb, avec une médiane à 1121pb. La suite des analyses comprendra le comptage des reads sur ces lncRNA, afin de vérifier leur présence, et pour réaliser une analyse de l'expression différentielle, entre mâles et femelles, pour les stades immatures et matures chez les gamétophytes. Les résultats seront comparés avec ceux de l'analyse de l'expression différentielle au niveau des mRNA pour identifier de possibles régulations *cis*. Enfin, une prédiction des lncRNA chez *Saccharina japonica* devrait être réalisée afin de pouvoir comparer les lncRNA avec ceux d'*Ectocarpus*, dans le but d'étudier le niveau de conservation des lncRNA entre ces espèces.

Article 2 - MicroRNAs and the evolution of complex multicellularity: identification of a large, diverse complement of microRNAs in the brown alga *Ectocarpus*

Introduction

Le processus de réannotation du génome d'*Ectocarpus* ne s'est pas limité à l'amélioration des annotations des gènes codants, mais il s'est aussi attaché à valider la présence des miRNA détectés lors de la première version de l'annotation (Cock et al. 2010) ainsi que les miRNA prédits de manière *in silico* (Billoud et al. 2014).

Afin de permettre une détection fine des miRNA, quatre librairies RNA-seq dédiées aux smallRNA ont été séquencées à partir d'individus mâles (Ec603) et femelles (Ec602) de deux souches quasi isogéniques, au stade de gamétophytes matures, avec deux librairies par sexe. De plus, les données utilisées lors de la première détection, et correspondant à des gamétophytes et des sporophytes, ont aussi été utilisée (Cock et al. 2010).

Ma contribution à cet article a été de réaliser le mapping des reads contre le génome de référence, avec Bowtie2 (Langmead and Salzberg 2012), dont seul les reads mappés sur la totalité de leur longueur ont été conservés. Les données de mapping ont été utilisées par miRDepp2 (Friedländer et al. 2012) afin de réaliser la prédiction des miRNA chez *Ectocarpus*.

Une autre partie de ma contribution a été d'obtenir un comptage du nombre de reads au niveau de plusieurs types de structures génomiques (exons, introns, rRNA, tRNA, snoRNA, régions intergéniques) afin de déterminer la proportion de reads dans chaque groupe. Dans le cas des snoRNA, j'ai généré une nouvelle prédiction afin de prendre en compte le supplément d'information apporté par les données RNA-seq, avec l'aide de l'outil snoSeeker (Yang et al. 2006).

Enfin, ma dernière contribution à cet article a été la recherche des cibles potentielles des miRNA l'aide de la version web de TAPIR (Bonnet et al. 2010)

(http://bioinformatics.psb.ugent.be/webtools/tapir/) en se basant sur l'importante complémentarité de séquence entre le miRNA et le mRNA cible.

Le résultat de ce travail a été intégré avec d'autres analyses dans l'article suivant, publié dans Nucleic Acids Research en mai 2015.

Article

microRNAs and the evolution of complex multicellularity: identification of a large, diverse complement of microRNAs in the brown alga *Ectocarpus*

Auteurs : James E. Tarver[1,2], Alexandre Cormier[3], Natalia Pinzón[4], Richard S. Taylor[1], Wilfrid Carré[3], Martina Strittmatter[3], Hervé Seitz[4], Susana M. Coelho[3], J. Mark Cock[3*]

Affiliations :

[1] School of Earth Sciences, University of Bristol, Life Sciences Building, 24 Tyndall Avenue, Bristol BS8 1TQ, UK;

[2] Genome Evolution Laboratory, Department of Biology, The National University of Ireland, Maynooth, Kildare, Ireland;

[3] Sorbonne Université, UPMC Univ Paris 06, CNRS, Algal Genetics Group, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688 Roscoff, France;

[4] Institute of Human Genetics, UPR 1142, CNRS, 34396 Montpellier Cedex 5, France

Correspondance : cock@sb-roscoff.fr

Article publié dans Nucleic Acids Research

doi : 10.1093/nar/gkv578

Nucleic Acids Research Advance Access published June 22, 2015

doi: 10.1093/nar/gkv578

microRNAs and the evolution of complex multicellularity: identificatio of a large, diverse complement of microRNAs in the brown alga *Ectocarpus*

James E. Tarver[1,2], Alexandre Cormier[3], Natalia Pinzón[4], Richard S. Taylor[1], Wilfrid Carré[3], Martina Strittmatter[3], Hervé Seitz[4], Susana M. Coelho[3] and J. Mark Cock[3,*]

[1]School of Earth Sciences, University of Bristol, Life Sciences Building, 24 Tyndall Avenue, Bristol BS8 1TQ, UK, [2]Genome Evolution Laboratory, Department of Biology, The National University of Ireland, Maynooth, Kildare, Ireland, [3]Sorbonne Université, UPMC Univ Paris 06, CNRS, Algal Genetics Group, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688 Roscoff, France and [4]Institute of Human Genetics, UPR 1142, CNRS, 34396 Montpellier Cedex 5, France

Received February 11, 2015; Revised April 19, 2015; Accepted May 21, 2015

ABSTRACT

There is currently convincing evidence that microRNAs have evolved independently in at least six different eukaryotic lineages: animals, land plants, chlorophyte green algae, demosponges, slime molds and brown algae. MicroRNAs from different lineages are not homologous but some structural features are strongly conserved across the eukaryotic tree allowing the application of stringent criteria to identify novel microRNA loci. A large set of 63 microRNA families was identifie in the brown alga *Ectocarpus* based on mapping of RNA-seq data and nine microRNAs were confirme by northern blotting. The *Ectocarpus* microRNAs are highly diverse at the sequence level with few multi-gene families, and do not tend to occur in clusters but exhibit some highly conserved structural features such as the presence of a uracil at the fi st residue. No homologues of *Ectocarpus* microRNAs were found in other stramenopile genomes indicating that they emerged late in stramenopile evolution and are perhaps specifi to the brown algae. The large number of microRNA loci in *Ectocarpus* is consistent with the developmental complexity of many brown algal species and supports a proposed link between the emergence and expansion of microRNA regulatory systems and the evolution of complex multicellularity.

INTRODUCTION

MicroRNAs (miRNAs) are small, 20–24 nucleotide RNA molecules (exceptionally up to 26 nucleotides) that regulate gene expression by affecting the translation or the stability of target gene transcripts. These small RNA molecules are generated from the double stranded regions of hairpin-containing transcripts by the action of RNAseIII endonucleases such as Drosha and Dicer and are then incorporated into RNA-induced silencing complexes (RISCs), which use the miRNAs as guides to recognize and bind to specific RNA targets. miRNAs have been shown to play key roles in the regulation of many important processes in both plants and animals (1,2) and it has been suggested that the acquisition of these versatile regulatory molecules may have been a key factor in the evolution of complex multicellularity (3–5).

The lack of sequence similarity between plant and animal miRNA families and marked differences between the pathways that generate miRNAs in the two groups suggest that these molecules evolved independently in the two lineages (6–9). In contrast, key components of the miRNA system, such as Dicer endonucleases and Argonaute (which is the central component of RISCs), are found in diverse eukaryotic lineages and are thought to be very ancient and perhaps common to all eukaryotes (10,11). These proteins are thought to have evolved originally as components of systems involving other classes of small RNA, such as the small interfering RNAs (siRNAs), and only later to have been recruited as components of miRNA pathways (9). Like miRNAs, siRNAs are small RNA molecules generated by endonuclease digestion but they may be derived from diverse sources of double stranded RNA such as viral genomes,

*To whom correspondence should be addressed. Tel: +33 2 98 29 23 60; Fax: +33 2 98 29 23 85; Email: cock@sb-roscoff.fr
Present address: Martina Strittmatter, The Scottish Association for Marine Science, Scottish Marine Institute, Oban, Argyll PA37 1QA, UK.

long transcribed inverted repeats or the products of convergent transcription. miRNAs on the other hand, are derived by endonuclease digestion of self-complementary precursor RNAs that form hairpin structures.

Although miRNAs were originally identified in land plants and animals, it has become increasingly clear in recent years that these are not the only eukaryotic lineages to have evolved regulatory systems based on these small RNA molecules. Within the animal lineage, the miRNAs of demosponges are unrelated to those of other animal groups and may have evolved independently (12). Similarly, the unicellular green alga *Chlamydomonas reinhardtii* also appears to possess an miRNA system that is unrelated to that of the land plant lineage (13,14), and no miRNA gene has been shown to be shared between land plants and green algae (15). There is also convincing evidence for the presence of miRNA systems in the brown alga *Ectocarpus* (16) and the social amoeba *Dictyostelium* (17,18). In addition, microRNA-like molecules have been reported in the fungus *Neurospora crassa* (19). Together these reports suggest that at least six or seven different eukaryotic groups possess miRNA systems. Moreover, these miRNA systems appear to have evolved independently in each group because no miRNAs are shared across groups and many intermediate lineages do not possess miRNAs (20). For example, miRNAs have been reported in the brown algae but no strong candidate miRNA loci have been identified in the genomes of three diatoms, which represent another lineage within the stramenopiles (21,22). A common evolutionary origin for the miRNA systems of diverse eukaryotic lineages therefore seems highly unlikely, as it would have required widespread loss of miRNA systems from intermediate lineages and extensive sequence divergence of shared miRNA loci. It is possible, however, that additional eukaryotic groups possess miRNA systems that have not yet been characterized. Indeed, putative miRNAs have been described in several additional lineages, although closer examination of the reported molecules has often failed to support their classification as miRNAs (20). Given the key roles of miRNAs as regulatory molecules in a broad range of processes and their implication in major evolutionary transitions such as the emergence of complex multicellularity (3–5), it is important both to experimentally confirm and characterize miRNA regulatory systems in groups where these systems exist and to clearly confirm their absence from other lineages. Here, we used deep sequencing of small RNA molecules, together with northern blot analysis, to identify and characterize miRNA loci in the filamentous model brown alga *Ectocarpus* and applied a set of stringent criteria to distinguish strong candidate miRNAs from other genomic sources of small RNAs such as siRNA loci. This analysis demonstrated that a recently described set of candidate miRNAs (23) are highly unlikely to correspond to miRNA loci and are more likely siRNAs, but also identified a large repertoire of 63 miRNA families in the *Ectocarpus* genome, the large majority of which had not been described previously. The complexity of the miRNA system in *Ectocarpus* is discussed in the light of the emergence of complex multicellularity in the brown algal lineage. We also discuss the importance of applying stringent criteria to validate candidate miRNA loci in the context of understanding miRNA emergence and evolution across the eukaryotic tree.

MATERIALS AND METHODS

Ectocarpus strains and culture

Two near-isogenic, male and female inbred lines Ec602 and Ec603 were derived from the male strain Ec137 and the female strain Ec25 by repeatedly crossing male and female sibling progeny for six generations (see Ahmed et al. for a detailed pedigree (24)). Ec137 (which carries the *immediate upright* mutation) (25), and Ec25 are siblings of the genome sequenced strain Ec32 (16). Two replicates of gametophytes for each sex were cultivated under standard conditions (26) and frozen at maturity (4 weeks old). Males bore many plurilocular gametangia, females were larger with fewer plurilocular gametangia. All material was examined under binocular and light microscopes to verify the presence of plurilocular gametangia and pools of about 400 individuals from these synchronous cultures were frozen in liquid nitrogen for each replicate.

Small RNA sequencing

The generation of 3 203 265 and 3 911 417 small RNA sequence reads for the sporophyte and gametophyte generations of *Ectocarpus*, respectively, has been described previously (16). An additional 77 702 501 small RNA reads (46 161 660 male and 31 540 841 female) were generated for the duplicate, near-isogenic male and female gametophyte samples (Supplementary Table S1). For the latter, small RNAs were isolated and prepared for sequencing by Fasteris (Planles-Ouates, Switzerland). Between four and 12 μg of total RNA was extracted for each replicate using the Qiagen Mini kit. RNA was separated on a polyacrylamide gel and the 15–30 nucleotide fraction isolated by excision. Addition of single-stranded adapters and PCR amplification was carried out using the DGE-Small RNA kit (Illumina, San Diego, USA) and small RNAs were sequenced on a HiSeq 2000 (Illumina). The sRNA sequence data can be accessed in the SRA Knowledge Base with the accession number SRP052304.

Adaptor sequence was removed from the raw sequence reads in Galaxy (27) and sequences of <18 or >26 nucleotides or which contained one or more unknown nucleotides were discarded.

Mapping of sRNA sequence reads to the *Ectocarpus* genome and transcriptome

The filtered reads were mapped against the *Ectocarpus* genome using Bowtie2 (28) with default parameters. Only fully mapped reads were retained (–end-to-end option in Bowtie2). Read coverage for genomic feature (exons, introns, rRNA, tRNA, snoRNA and intergenic regions) was obtained using Samtools (29). *Ectocarpus* snoRNA loci were predicted using ACAseeker and CDseeker. Coordinates of other genomic features, including rRNA and tRNA loci, were obtained from the *Ectocarpus* genome database at Orcae (http://bioinformatics.psb.ugent.be/orcae/overview/Ectsi) (30).

Sliding window analysis of sRNA read coverage was calculated using a custom script and a non-overlapping sliding window of 25 kb. The data is presented as sRNA read counts per window. Visual analysis of the mapping pattern of the sRNA reads onto the genome indicated that it was not consistent with more than a very limited level of contamination by degraded mRNA fragments. This conclusion was also supported by the fact that 47% of the reads that mapped to mRNA-encoding regions of the genome mapped to the antisense strand compared to the mRNA transcript (data not shown).

Expression levels (transcript abundance) of protein-coding genes in male and female gametophytes were determined using the Illumina RNA-seq dataset described by Ahmed *et al.* (24).

Identification of *Ectocarpus* protein-coding genes with potential roles in small RNA pathways

Ectocarpus homologues of plant, animal and fungal protein-coding genes that have been implicated in various aspects of sRNA biogenesis and function were identified by screening for species to species best reciprocal Blastp matches.

Identification and characterization of miRNA loci in *Ectocarpus*

Ectocarpus is distantly related to both land plants and animals. Screens for miRNAs therefore employed both miRDeep2 (31), which implemented criteria for the identification of animal miRNAs, and miRDeep-p (32), a modified version of miRDeep that was adapted for the identification of plant miRNAs by allowing extended precursor sequences. After filtering, the reads from all six samples (Supplementary Table S1) were combined into a single dataset and provided as input for each program. Candidate miRNA precursors were then extracted from the output files and the miRDeep-2 and miRDeep-p outputs compared using Blast to identify and remove redundant candidate miRNA precursors that had been identified by both programs.

Custom scripts, which incorporated Bowtie (28), were used to align all the sRNA sequence reads to the candidate precursor miRNA loci, with no mismatches allowed. For each miRNA locus, the sRNA species with the highest read count was compared with miRBase using Blast and the most similar match recovered if matches were detected. The entire precursor sequence was folded with Vienna RNAfold (33) and a further script was implemented to combine the output of this analysis with the sRNA read mapping results and miRBase Blast search results.

Similar analyses of sRNA read mapping were also carried out for 23 *Ectocarpus* miRNA loci recently reported by Billoud *et al.* (23).

Investigation of the genomic origin of the *Ectocarpus* miRNA loci

To identify miRNA families, ungapped alignments of either the mature miRNA sequences or just the seed regions (nucleotides 2–8) were generated with Muscle (34) and pairwise sequence identity calculated using MEGA (35). Pre-miRNA sequences were analysed with RepeatMasker (http://repeatmasker.org) against Repbase to detect sequence relationships with repeated elements. Similarity with other genomic regions was detected using Blastn and the pre-miRNA sequences as a query against the *Ectocarpus* genome sequence (http://bioinformatics.psb.ugent.be/orcae/overview/Ectsi) (30). The principal aim of the latter analysis was to determine whether the *Ectocarpus* miRNA loci might have been derived from duplicated copies of protein-coding genes.

Searches for homologues of *Ectocarpus* miRNA loci in other stramenopile genomes

Searches were carried out for homologues of *Ectocarpus* miRNA loci in the genomes of four stramenopile species: *Thalassiosira pseudonana* (diatom; Thaps3 assembled and unmapped scaffolds, http://genome.jgi-psf.org/Thaps3/Thaps3.download.ftp.html) (36), *Phaeodactylum tricornutum* (diatom; Phatr2 assembled and unmapped scaffolds, http://genome.jgi-psf.org/Phatr2/Phatr2.download.ftp.html) (37), *Aureococcus anophagefferens* (Pelagophyceae; http://genome.jgi-psf.org/Auran1/Auran1.download.ftp.html) (38) and *Nannochloropsis oceanica* (Eustigmatophyceae; https://bmb.natsci.msu.edu/BMB/assets/File/benning/genome_assembly.txt) (39). Two different strategies were used. The first involved comparing the entire pre-miRNA sequences with the genomes using Blastn and then analysing the results manually for extended regions of similarity that preferentially included the miRNA and miRNA* regions of the pre-miRNA. The second method involved querying both the miRNA and miRNA* sequences against the genomes and retaining matches with less than four mismatches. The region surrounding each match was then recovered from the subject genome sequence and tested for the ability to form a hairpin loop with sufficient complementary base-pairing between the candidate miRNA and miRNA* sequences.

Comparisons with miRNA loci from other eukaryotic lineages

Structural features of the *Ectocarpus* miRNAs were compared with those of miRNAs from species belonging to other eukaryotic lineages. The sets of miRNAs from the other eukaryotic lineages had been validated previously (40,41) using the same four criteria that we employed in this study to select valid *Ectocarpus* miRNAs (see the Results and Discussion section for details of the four criteria). The species used for the comparisons were *Drosphila melanogaster*, *Danio rerio* (animals), *Amphimedon queenslandica* (demosponges), *Dictyostelium discoideum* (slime molds), *Arabidopsis thaliana, Physcomitrella patens* (land plants), *Chlamydomonas reinhardtii* (chlorophyte green algae). The miRNA expression data were recovered from miRBase (8). Foldback lengths (42) were calculated for the miRNAs from each species using precursor sequences deposited in miRBase v21 that had both the 5′ and 3′ products annotated and have been previously validated as genuine (40,41), and from the annotated *Ectocarpus* miRNAs

described herein. The region corresponding to each miRNA precursor was identified in the respective genome sequence using Blast and the region, together with 100 nucleotides of flanking sequence in both directions, was recovered. RNAfold (33) was used to predict secondary structure, and the foldback was deemed to have ended either at the first occurrence of three consecutive unbound nucleotides or at the occurrence of another secondary structure.

Northern blot analysis

Samples of either 50 or 63 μg of total RNA from male or female *Ectocarpus* strains were subjected to northern blot analysis as previously described (43). DNA oligonucleotide probes complementary to the miRNAs of interest were radioactively labelled at the 5′-end using T4 polynucleotide kinase.

Searches for potential target genes of *Ectocarpus* miRNA loci

Potential target genes of *Ectocarpus* miRNAs were identified using the web version of TAPIR (http://bioinformatics.psb.ugent.be/webtools/tapir/) in precise mode with the default options.

RESULTS AND DISCUSSION

Sequence analysis of gametophyte small RNAs

The first description of miRNA loci in the brown alga *Ectocarpus* was based on the analysis of about seven million sRNA sequences generated using both sporophyte and gametophyte tissue (16). For the present study, an additional 78 million sRNA sequence reads were generated using replicate samples of RNA from male and female gametophytes. Mapping of the sRNA sequence reads to the genome indicated that they were derived from all chromosomes, with no obvious bias towards particular linkage groups or regions within linkage groups (Supplementary Figure S1). After exclusion of reads corresponding to ribosomal RNA (rRNA), transfer RNAs (tRNAs) and small nucleolar RNAs (snoRNAs), the highest coverage of mapped sRNA reads per base pair was for transposable elements (Table 1). This confirms an earlier observation (16) and suggests a possible role for these sRNAs in maintaining genome stability by repressing transposition. Small RNAs have been associated with silencing of transposable elements in a broad range of eukaryotic organisms, including animals, plants and fungi (44). Thirty seven percent of the mapped reads corresponded to regions annotated as genes, with the exon regions being covered slightly more densely than the introns (1.5-fold).

One unusual structural feature of the *Ectocarpus* genome is that the coding strands of adjacent protein-coding genes exhibit a strong tendency to alternate between the two strands of the DNA as one scans along the chromosome, a feature that is normally associated with very small eukaryotic genomes (16). One consequence of this is that 9508 of the 16 192 genes in the *Ectocarpus* genome are part of a convergently transcribed gene pair, i.e. the two genes are located adjacent to one another on the chromosome and transcribed convergently. Pairs of convergent transcription units have been reported to be an important source of sRNAs in both animals and land plants (45–47). This is thought to be because overlap between the pairs of transcripts generates regions of double-stranded RNA. In *Ectocarpus*, the number of sRNA reads that mapped to genes that were members of convergent gene pairs (median FPKM 0.20) was slightly, but significantly (Kruskal–Wallis test, P-value < 8.1e−09), greater than for the other genes in the genome (median FPKM 0.18). However, analysis of mRNA-seq expression data showed that convergent genes were also expressed at a slightly higher level than non-convergent genes (mRNA median FPKM of 10.1 compared with 8.8, Kruskal–Wallis test, $P = 6.8e{-}16$) and when the number of sRNA reads per gene was normalized for this difference there was no significant difference between genes that were members of convergent pairs and the other genes in the genome (Kruskal–Wallis test, $P = 0.77$). This indicates that convergent gene pairs are not a preferential source of sRNAs in *Ectocarpus*.

An analysis was also carried out to identify protein-coding genes with potential roles in small RNA pathways in *Ectocarpus*. Reciprocal best Blast analysis identified >30 homologues of plant, animal and fungal genes that have been implicated in various aspects of sRNA biogenesis and function (Table 2).

Ectocarpus has a large and diverse repertoire of microRNAs

A screen was carried out for miRNA loci using the algorithms miRDeep2 and miRDeep-p, which are optimized to detect animal-like and plant-like miRNAs respectively, together with custom scripts. This analysis identified 1882 candidate miRNA loci, which were then manually filtered following established criteria based on highly conserved features common to both animal and plant miRNA loci (20,40,41): (i) at least 15 nucleotides of the miRNA must pair with the opposite arm of the hairpin, (ii) there should be evidence for the expression of both the miRNA and the miRNA*, (iii) the 3p product should extend two nucleotides beyond the 5p product at its 3′ end (with a corresponding extension at the 3′ end of the 5p product), (iv) 5′ cleavage of the miRNA must be precise, with the clear majority of the reads (at least 66%) starting at the same nucleotide.

The final set of 63 microRNA families (representing 64 loci) included six of the miRNA families previously described by Cock *et al.* (16), together with 57 newly identified families (Supplementary Table S2, Figure 1 and Supplementary Figure S2). Northern blot analysis was carried out to independently validate a subset of nine of these miRNA loci using RNA from a separate set of RNA samples. sRNA species of the expected size were detected in both male and female gametophyte RNA samples for all of the nine miRNA loci (Figure 2). The relative abundances of the miRNAs, estimated from the northern blot analysis, corresponded approximately with estimations based on RNA-seq, with some miRNAs, such as esi-MIR11396a and esi-MIR11368, being expressed at high to very high levels and others, such as esi-MIR11377 and esiMiR3458, being less abundant.

A striking feature of *Ectocarpus* miRNAs is their remarkable diversity, with almost every miRNA constituting a distinct miRNA gene family. When the seed regions of the miRNAs (nucleotides 2–8) (48) were compared, only one

Table 1. Mapping of sRNA reads to different fractions of the *Ectocarpus* genome

Genome fraction or feature	sRNA read count	Cumulative size (bp)	Read coverage (reads per bp)
Exons	3 469 027	25 662 441	0.14
Introns	10 424 142	81 093 270	0.13
Intergenic	7 050 763	73 482 052	0.10
Transposons	8 915 590	10 605 262	0.84
tRNA	755 534	21 829	34.61
rRNA	7 092 058	7903	897.39
snoRNA	57 845	88 311	0.66

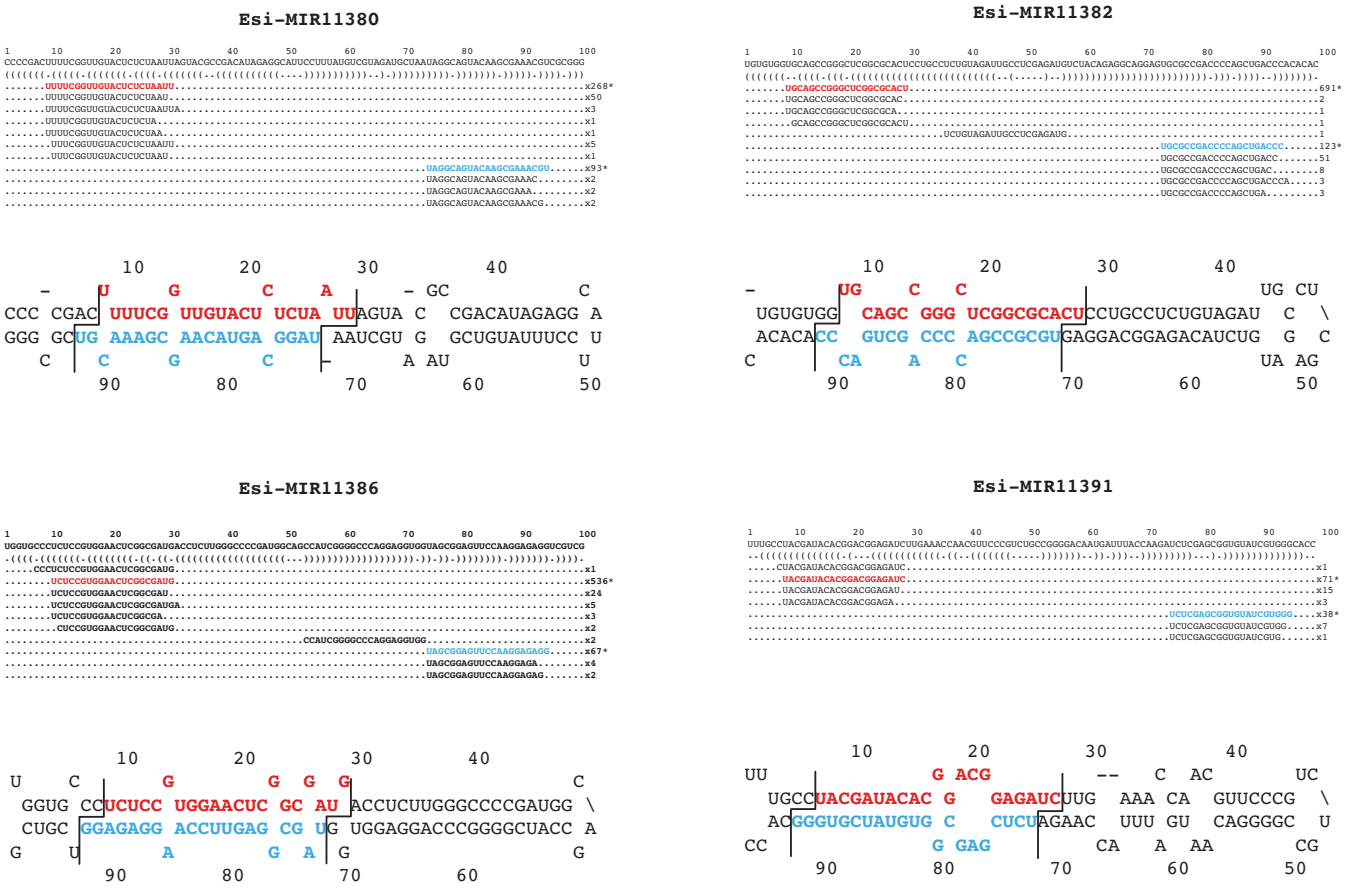


Figure 1. Representative *Ectocarpus* miRNA loci. Representation of read data mapping and the positions of the miRNA (red) and miRNA* (blue) on the predicted hairpin for four representative *Ectocarpus* miRNA loci. Note the high degree of homogeneity of 5′ ends. Lines cutting across the hairpins indicate the two nucleotide offset typical of Dicer processing. Similar diagrams for the full set of 64 miRNAs are shown in Supplementary Figure S2.

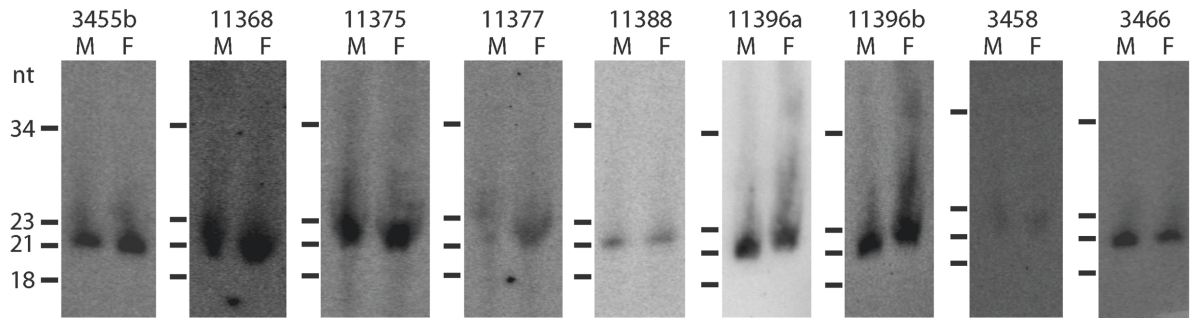


Figure 2. Northern blot analysis of miRNA expression in male and female gametophytes. Hybridization to 50 (esi-MIR3455b, esi-MIR11368, esi-MIR11375, esi-MIR11377, esi-MIR11388, Esi-miR3458, Esi-miR3466) or 62.3 (esi-MIR11396a, esi-MIR11396b) μg of male or female total RNA per lane. Exposure times were the same for all samples except for esi-MIR11396a, which is highly abundant and was exposed for one day rather than 4 days.

Table 2. *Ectocarpus* homologues of proteins involved in microRNA function or related small RNA pathways in other species

Query species	Query gene	Function	Accession number	*Ectocarpus* best Blastp E-value	Reciprocal best blast (species to species)	*Ectocarpus* homologue
Ath	DCL2	Dicer	NP_566199.4	1E−12	Yes	Esi0039_0031
Aga	Ago1	Argonaute	EAA00062.4	3E−94	Yes	Esi0203_0032
Ddi	AgnA	Argonaute (piwi)	EAL69296.1	2E−29	No	No homologue
Ath	RDR1	RNA-dependent RNA polymerase	NP_172932	4E−30	Yes	Esi0512_0001
Ath	RDR6	RNA-dependent RNA polymerase	NP_190519	4E−88	Yes	Esi0100_0017
Ath	SDE3	SDE3/MOV10/Armitage	AAK40099.1	3E−62	Yes	Esi0216_0047
Ath	DAWDLE	pri-miRNA generation	NP_188691.1	3E−47	Yes	Esi0132_0041
Ath	SQUINT	pre-miRNA processing	Q9C566	1E−54	No	Multiple cyclophilins
Ath	HSP90	pre-miRNA processing	AED96244.1	0	Yes	Esi0138_0009
Ath	HASTY	Nuclear export (exportin5/MSN5/HASTY)	Q84UC4	0.00002	Yes	Esi0059_0032
Ath	SERRATE	RNA binding protein that may maintain hairpin structure or direct Dicer	Q9ZVD0	8E−10	Yes	Esi0289_0007
Ath	HYL1	RNA binding protein that may maintain hairpin structure or direct Dicer	NP_563850.1	No hit	n/a	No homologue
Ath	TOUGH	RNA binding protein that may maintain hairpin structure or direct Dicer	AAR99647.1	1E−23	Yes	Esi0125_0056
Ath	HEN1	2′-O-Methylation of miRNAs	NP_567616.1	No hit	n/a	No homologue
Ath	SUO	miRNA-mediated translational repression	NP_190388.2	No hit	n/a	No homologue
Ath	MOS2	miRNA processing	NP_174617.1	2E−23	Yes	Esi0084_0044
Ath	PRL5	miRNA processing	NP_193325.1	6E−131	Yes	Esi0025_0074
Ath	CDC5	miRNA processing	NP_172448.1	3E−69	Yes	Esi1122_0001
Ath	SICKLE	miRNA biogenesis	NP_567704.1	No hit	n/a	No homologue
Ath	KTF1/RDM3/SPT5-like	AGO4 interactor	NP_196049.1	2E−16	No	No homologue
Ath	CBP20	Cap binding complex	NP_199233.1	5E−39	Yes	Esi0206_0003
Ath	CBP80	Cap binding complex	NP_565356.1	6E−28	Yes	Esi0155_0015
Ath	DECAPPING1	Decapping complex	NP_563814.1	4E−16	Yes	Esi0489_0024
Ath	DECAPPING2	Decapping complex	Q8GW31	3E−40	Yes	Esi0010_0022
Ath	VARICOSE	Decapping complex	AEE75331.1	3E−19	Yes	Esi0205_0050
Ath	3-HYDROXY-3-METHYLGLUTARYL CoA REDUCTASE	Isoprenoid synthesis protein that affects miRNA action	NP_177775.2	3E−83	Yes	Esi0027_0087
Ath	HYDRA1	Isoprenoid synthesis protein that affects miRNA action	NP_173433.1	0.00000001	No	No homologue
Ath	SMALL RNA DEGRADING NUCLEASE 1	miRNA degradation	AEE78626.1	1E−25	Yes	Esi0118_0050
Ath	HESO	miRNA uridylation	NP_181504.2	3E−14	No	Uridyltransferases eg. Esi0771_0003
Ath	AMP1	Inhibition of protein production	NP_567007.1	1E−87	Yes	Esi0122_0005
Ath	KATANIN	Cytoskeleton genes that affect miRNA action	NP_178151.1	4E−93	Yes	Esi0007_0029
Ath	dsRNA BINDING PROTEIN4	tasiRNA biogenesis	Q8H1D4	No hit	n/a	No homologue
Ath	SGS3	RNA-directed DNA methylation	AAF73960.1	No hit	n/a	No homologue
Ath	C-TERMINAL DOMAIN PHOSPHATASE-LIKE1	Phosphorylation role in dsRNA gene regulation	NP_193898.3	0.023	No	No homologue
Ath	CLSY1	Generation of 24nt rasiRNAs	NP_189853.1	3E−17	No	No homologue
Ath	PolIV	siRNA synthesis	NP_176490.2	7E−42	No	No homologue
Ath	NRPE1	siRNA synthesis	NP_181532.2	9E−43	No	No homologue
Dme	Pasha	Drosha complex	AAF57175.1	No hit	n/a	No homologue
Hsa	EWSR1	Drosha complex	NP_053733.1	1E−16	Yes	Esi0222_0008
Hsa	p68/DDX5	Drosha complex	NP_004387.1	3E−87	Yes	Esi0013_0199
Hsa	p72/DDX17	Drosha complex	NP_001091974.1	5E−157	Yes	Esi0007_0206
Hsa	Fus	Drosha complex	AAC35285.1	2E−12	No	No homologue
Hsa	ADAR	pri-and/or pre-miRNA editing	EAW53187.1	0.000003	No	No homologue
Hsa	TRBP	pre-miRNA processing	Q15633.3	No hit	n/a	No homologue
Hsa	PACT	pre-miRNA processing	AAL68925.1	1E−26	No	No homologue
Dme	loquacious	pre-miRNA processing	AAY40789.1	No hit	n/a	No homologue
Hsa	KSRP	Promoter of miRNA biogenesis	AAB53222.1	2E−11	No	No homologue
Hsa	Lin28	Drosha/Dicer inhibitor	AAH28566.1	No hit	n/a	No homologue
Hsa	TNRC6 (GW182)	RISC component	NP_055309.2	0.84	No	No homologue
Hsa	TNRC6A	Ago interactor	Q8NDV7	No hit	n/a	No homologue
Hsa	TNRC6B/KIAA1093	Ago interactor	TNC6B_HUMAN	No hit	n/a	No homologue
Hsa	TNRC6C	Ago interactor	Q9HCJ0	No hit	n/a	No homologue
Hsa	TRIM65	Ubiquitination of TNRC6	NP_775818.2	0.000001	No	No homologue
Dme	R2D2	Double-stranded RNA binding protein	Q9VLW8	No hit	n/a	No homologue
Dme	FMR1	miRNA biogenesis	Q9NFU0	No hit	n/a	No homologue
Dme	BEL	ATP-dependent RNA helicase	Q9VHP0	1E−134	Yes	Esi0186_0022
Dme	RM62	DEAD-box RNA helicase	P19109	1E−140	No	Multiple RNA helicases
Cel	ERI1	RNA exonuclease	O44406	6E−26	Yes	Esi0039_0083
Cel	RDE-4	siRNA production	Q22617	No hit	n/a	No homologue
Cel	SID1	Systemic RNA interference	Q9GZC8	No hit	n/a	No homologue
Spo	Hrr1	RNA-directed RNA polymerase complex	O74465	2E−41	No	No homologue
Spo	Cid12	Poly(A) polymerase	O74518	0.00000002	No	Esi0053_0139 (Poly(A) polymerase)
Spo	Chp1	RNAi pathway	Q10103	No hit	n/a	No homologue
Spo	Tas3	RNAi pathway	O94687	No hit	n/a	No homologue

For Dicer, Argonaute and RNA-dependent RNA polymerase, searches were carried out with multiple sequences from diverse eukaryote lineages (10). Ath, *Arabidopsis thaliana*; Aga, *Anopheles gambiae*, Ddi, *Dictyostelium discoideum*; Dme, *Drosophila melanogaster*; Hsa, *Homo sapiens*; Cel, *Caenorhabditis elegans*; Spo, *Schizosaccharomyces pombe*; n/a, not applicable.

pair of genes (esi-MIR11384a/esi-MIR11384b) was classed as belonging to the same family. The same result was obtained when a criterion of at least 85% identity between entire, mature miRNAs (15) was used to define members of a gene family. Even when this latter criterion was considerably relaxed to at least 75% identity, only three families, each with two members, were identified. These observations suggest that miRNA gene duplication has not played an important role in the generation of new miRNA loci in the brown algal lineage. This is in stark contrast to the role that both individual gene and whole genome duplications have played in miRNA family expansion in both animals (49,50) and land plants (15). The low number of paralogues within miRNA families in *Ectocarpus* is consistent with both the lack of evidence for any whole genome duplication events in the lineage leading to this organism and the unusually low number of tandem duplications of protein coding genes (823) identified in this species (16).

Mapping of the 64 miRNA loci to the *Ectocarpus* genome indicated that they were distributed randomly across the chromosomes (Supplementary Figure S1). Clusters of miRNAs (defined here as being within 5 kb (15)) are common in both animals and land plants (15,49,51,52). In contrast, the *Ectocarpus* miRNA loci exhibited very little tendency to cluster in the genome, with only two pairs of loci being separated by <5 kb. The miRNAs encoded by one of these pairs of clusters shared 76% identity, suggesting that they may have been derived from a tandem duplication event. However, such local duplication events appear to have been very rare.

There is evidence that some miRNA loci in both animals and plants produce more than one pair of miRNA-like molecules from a single pre-miRNA hairpin structure (53–56). These additional miRNA-like molecules are often in phase with the miRNA/miRNA* pair, in which case they have been called miRNA-offset RNAs (moRNAs). There is accumulating evidence that these additional miRNA-like molecules have biological functions (55,56) and, therefore, they may contribute significantly to the total size of the miRNA repertoire in some species. In plants these miRNA-like molecules tend to exhibit a strong preference for a U or A nucleotide at the 5′ end (90% in *Arabidopsis*) (56) but this does not appear to be the case in animals (53). We did not obtain evidence that this type of miRNA-like molecule occurs commonly in *Ectocarpus*, but esi-MIR11352 was of interest because a putative moRNA (UCUUUGAUCGGACAUGUUUCU) with a 5′ U nucleotide and 5′ processing homogeneity was detected for this locus, along with a potential 'star' product (Supplementary Figure S2).

In addition to the 64 miRNA loci identified, we also noted the presence of a large number of loci that were identified by miRdeep2 and/or miRdeep-p and fulfilled the majority of the criteria we used to define miRNA loci but were located in genomic regions consisting of complex, extensive palindromic sequences that generated multiple sRNA species over a region of several hundred base pairs (see Figure 3 for an example of such a locus). Sixty-five of these additional loci, which we classified as weak miRNA candidates, are shown in Supplementary Figure S3. The analysis of these loci highlighted the importance of manually checking the genomic context of a candidate miRNA even if loci are computationally predicted with high confidence. Further analysis will be required to determine whether these loci actually produce functional miRNAs, but it is possible that they may represent so-called transitional miRNAs (57), i.e. newly emerging miRNA loci.

Quantitative PCR is not a suitable strategy for identifying novel miRNAs

Detailed analysis of the mapped sRNA reads allowed us to demonstrate that 23 *Ectocarpus* miRNA loci recently described by Billoud *et al.* (23) failed to pass the quality control criteria applied here. These candidate miRNAs were part of a larger set (500–1500 depending on criteria) that had been identified using a bioinformatic approach. A subset of 72 candidates were analysed by Billoud *et al.* using quantitative PCR and 23 were subsequently reported as miRNAs. However, we were unable to validate any of these 23 candidate miRNAs using our sRNA read data. For five of the candidates no sRNA reads mapped to the loci, for an additional thirteen candidate miRNAs the most abundant class of read was not the annotated miRNA product and for the final five candidates the miRNA/miRNA* pairing was clearly incorrect (Supplementary Figure S4). Thus the candidate miRNAs identified by Billoud *et al.* are most likely siRNAs.

Quantitative PCR is commonly used to validate candidate miRNAs identified by bioinformatic approaches due to its low cost. The analysis carried out in this study identified limitations of this approach and demonstrated the importance of validation using sRNA read data. sRNA read data allows key criteria such as evidence for the existence of both miRNA and miRNA* species, homogeneity of 5′ processing and pre-miRNA processing consistent with dicer activity, to be tested. Whilst quantitative PCR is clearly useful for the quantification of known miRNAs, as a tool to validate novel candidate miRNA loci it suffers from the weakness of not being able to distinguish miRNAs from rare RNA species, siRNAs or degraded products of diverse RNA transcripts.

Expression patterns of *Ectocarpus* miRNAs

Expression levels (Supplementary Table S2) varied between 0.24 and 8387.33 RPM for the miRNA and between 0.01 and 131.95 RPM for the miRNA* (the miRNA being defined as the most strongly expressed of the two species (58)).

Sex-biased expression of miRNA loci has been reported for both animals (59–62) and land plants (63). Statistical tests, implemented with DEseq and EdgeR, were therefore carried out to determine whether any of the miRNA loci were differentially expressed in male and female individuals, but no statistically significant differences were detected. Similarly, there was no evidence that the miRNAs were differentially expressed between the sporophyte and gametophyte generations of the life cycle. Note that the RNA blot analysis did not provide any evidence for differential expression of the miRNA loci between male and female individuals (Figure 2), in agreement with the analysis of the RNA-seq data.

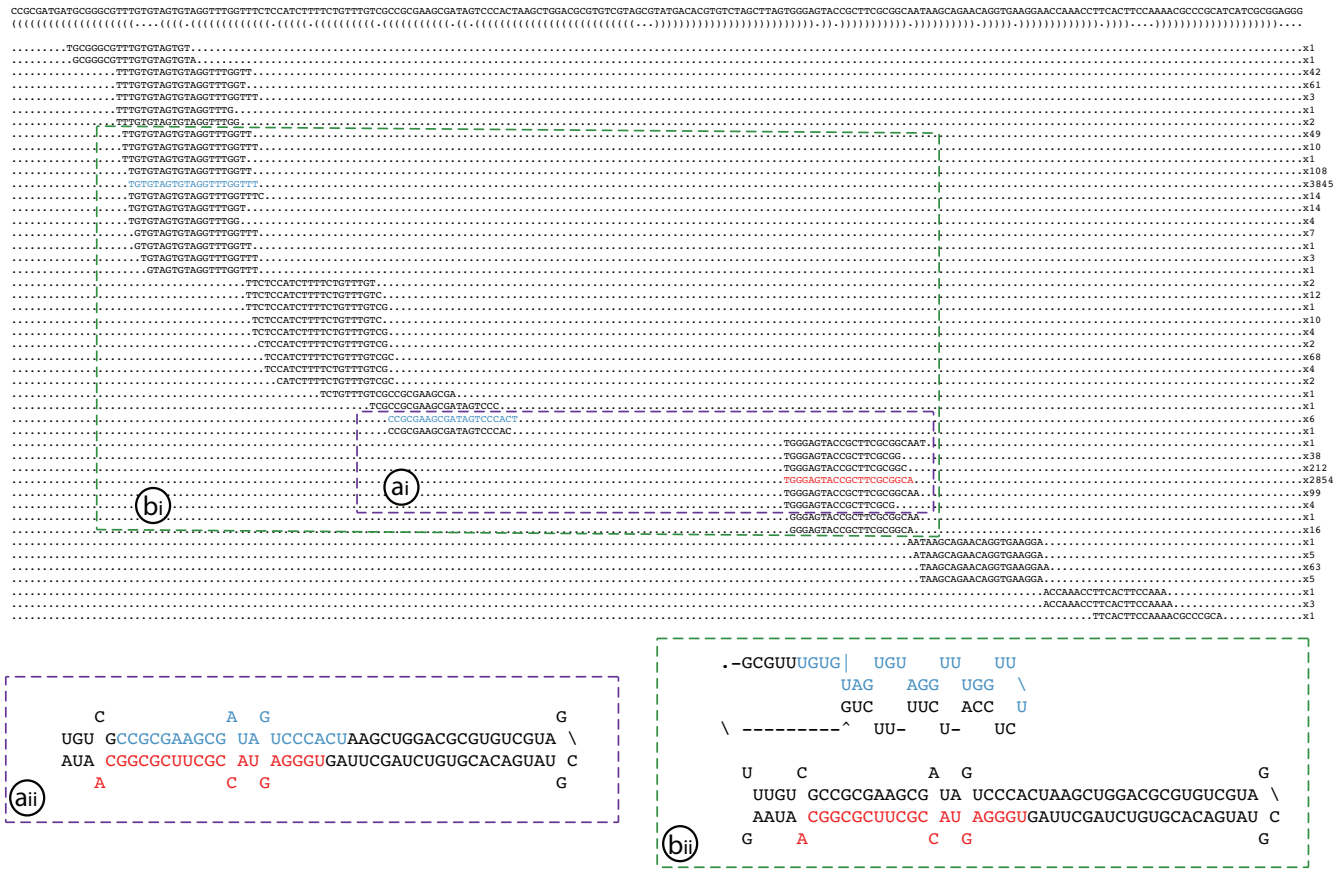


Figure 3. Example of a weak candidate miRNA. Weak candidate miRNA 8 was identified by miRDeep2 (31) with a high score (1.6e+3). The locus encodes potential miRNA (2854 reads) and miRNA* (6 reads) products (ai), with the expected 2 nucleotide offset and a characteristic hairpin loop (aii). However, when the precursor sequence was extended, two regions were identified on the 5′ side that exhibited higher expression (3845 and 68 reads, respectively) than the miRNA* sequence originally annotated by miRDeep2 (bi). When this longer precursor was folded (bii) it no longer formed a characteristic hairpin, and the two products lacked both the required offset and sufficient complementary base pairing.

Prediction of miRNA target genes

Sequence complementarity between miRNAs and their target mRNAs varies across eukaryotic groups, with plant and green algal miRNAs tending to have a high level of complementarity with their target genes and animal miRNA, in contrast, tending to have low complementarity (although there is evidence that plant miRNAs can also have low complementarity targets (64)). As a first step towards identifying putative targets of *Ectocarpus* miRNAs, we carried out a search, using TAPIR (65), based on the assumption that complementarity between the miRNA and mRNA target was high. This analysis identified 160 potential target genes in the *Ectocarpus* genome (Supplementary Table S3), with individual miRNAs being matched to between zero (17 miRNAs) and 13 target genes. Experimental validation will be required to verify that these genes are actually targets of the *Ectocarpus* miRNAs.

Interestingly, seven of the 160 genes were predicted to be targeted by two miRNAs. In four of these seven cases, the two miRNAs had different seed regions and targeted different regions of the gene. Note however that, in general, the high diversity of the seed regions of the *Ectocarpus* miRNAs suggests that there is unlikely to be a high level of target redundancy in this species, i.e. in most cases target genes are unlikely to be targeted by multiple miRNA loci.

Functions could be predicted for 104 of the putative target genes based on sequence information and this analysis indicated that they were involved in a broad range of cellular processes. Strongly represented cellular processes included cellular signalling and regulation (11 genes), proteolysis (11 genes), membrane function (10 genes) and genes with a probable role in defence (10 genes), with an additional 18 genes involved in general protein-protein interactions.

Genomic origin of the *Ectocarpus* miRNAs

Several mechanisms have been described for the generation of new miRNA loci; these include: (i) duplication of existing miRNA loci (66), (ii) generation of miRNA loci from duplicated copies of protein-coding genes (67,68), (iii) evolution from transposable elements (17,69,70) and (iv) evolution from the many hairpin regions scattered throughout the genome (52,71,72). The near absence of miRNA families and miRNA clusters in *Ectocarpus* suggests that duplication of miRNA loci has not been not a major mechanism for the generation of new miRNA loci in this species. Simi-

larly, comparison of the *Ectocarpus* pre-miRNA sequences with transposon sequences using RepeatMasker and with the *Ectocarpus* protein-coding genes using Blast did not detect any evidence that the miRNAs were derived from the latter features. By deduction, therefore, these analyses suggest that hairpin regions in the genome may have been an important source of new miRNA loci in this lineage. Hairpin regions within introns may have been favoured during this process because they had the advantage of already being transcribed. Evolution of miRNA loci from genomic hairpins is thought to have been an important mechanism of miRNA genesis in animals, and it has been suggested that this mechanism, as opposed to recruitment of duplicated fragments of future target genes, may have been favoured by the low level of sequence similarity between animal miRNAs and their targets (15). It remains to be determined whether this is also the case for *Ectocarpus*. The search carried out in this study identified potential targets that shared high similarity with the *Ectocarpus* miRNAs but further analysis will be required to validate these potential target genes.

The majority of the *Ectocarpus* miRNA loci are located within protein-coding genes (75%). This contrasts with the situation observed in land plants, where most (84%) miRNA loci are located in intergenic regions (15) and is more similar to that of several animals including humans and *Drosophila*, where nearly half of the miRNA genes occur in introns (52,73). One of the factors that may explain the observed distribution of *Ectocarpus* miRNA loci is that protein coding genes, and particularly intron sequence, constitute an exceptionally large proportion of the genome sequence in this species (16). Indeed, the *Ectocarpus* miRNAs that occur within protein-coding genes were found principally located within introns, the only exceptions being three miRNAs that were located in untranslated regions. All but three of the miRNA loci that were located within protein coding genes were transcribed from the same strand as the host gene (Supplementary Table S2), indicating that the intronic miRNAs could be co-transcribed as part of their host gene mRNA. However, no correlation was detected between the abundances of these miRNAs (RPM) and the abundance of their host gene's mRNA (FPKM) in the duplicate male and female gametophyte samples (Spearman's rank correlation coefficient $\rho = 0.0019$, P-value $= 0.98$). Lack of correlation between the abundances of intronic miRNAs and their host gene mRNA transcripts has also been observed in animal systems (74) (but see also (75)). When the abundances of these two molecules are not correlated, this may be either because their relative abundances are significantly influenced by post-transcriptional processes affecting the processing or stability of at least one of the two types of molecule or because the two features are transcribed independently (or a combination of these two phenomena). Several studies have indicated that a significant proportion of human intronic miRNAs possess their own promoters, which could function independently of the host gene promoter (e.g. (76,77)). It is possible that many of the intronic *Ectocarpus* miRNAs are also transcribed independently of the host gene. Note that, while the emergence of new miRNA loci may be favoured in regions of the genome that are already transcribed such as introns, subsequent acquisition of an independent promoter would confer greater flexibility of expression. In this respect it is interesting to note that, in animals, evolutionarily old intronic miRNA loci appear to be more likely to possess their own promoter region than young intronic miRNA loci (76).

None of the intronic miRNAs were mirtrons. The intron that contains miRNA esi-MIR11390 (intron 3 of gene Esi0084_0039) is predicted to form a stem-loop that involves the entire sequence of the intron but the miRNA/miRNA* duplex is not located next to the splice site.

Evolutionary origins of the *Ectocarpus* miRNA loci

Comparisons of miRNA complements of diverse species within both the land plant and animal lineages has shown that these loci accumulate gradually over evolutionary time, that their sequences are strongly conserved and that they are rarely lost once acquired (40,41,78). Where loss of miRNA loci has occurred, this can often be correlated with genome reduction and phenotypic simplification, for example in lineages that have adopted a parasitic life history (79). It has recently been suggested that miRNA loss is a more common phenomenon than previously reported (80) but this latter study did not adequately take into account the widely appreciated phenomenon of apparent loss of miRNA loci due to the use of low coverage genome and/or small RNA sequence data, which can lead to considerable over-estimation of the rate of miRNA loss during evolution (40,81,82). Current evidence therefore indicates that a certain proportion of miRNA loci are conserved over long periods of evolutionary time. Based on this observation, we carried out a search for homologues of the *Ectocarpus* miRNAs in other stramenopile lineages.

At present, the *Ectocarpus* genome is the only complete genome sequence available for the brown algae, but the genomes of several other members of the stramenopile supergroup have been sequenced. A search was carried out for sequences homologous to the 64 *Ectocarpus* pre-miRNA regions in the genomes of two diatom species, *Thalassiosira pseudonana* (36) and *Phaeodactylum tricornutum* (37) and members of the Pelagophyceae (*Aureococcus anophagefferens*) (38) and the Eustigmatophyceae (*Nannochloropsis oceanica*) (39). The latter two classes are more closely related to the brown algae than the diatoms (83). Blastn search results were analysed for matching regions that exhibited at least partial conservation of the miRNA and/or miRNA* sequences and could potentially encode RNAs with hairpin structures, but no clear matches were found in any of the four species analysed. Recent estimates indicate that these four species of stramenopiles may all have diverged from the brown algal lineage more than 400 Mya (83). It is therefore possible that extensive divergence over this length of evolutionary time may have obscured homologies. However, given that subsets of both animal and land plant miRNA loci have been strongly conserved over similar periods of time (15,40,41,49), this is unlikely to have been the case for all of the miRNA loci. Moreover, recent extensive searches of three diatom genomes failed to find any strong candidate miRNA loci, indicating that this stramenopile group does not possess a miRNA regulatory system (21,22). Taken together, these observations suggest that

the *Ectocarpus* miRNA loci have evolved since the brown algal lineage diverged from that of the Eustigmatophyceae.

There is currently convincing evidence for the existence of miRNA loci in six diverse eukaryotic groups: metazoans, demosponges, slime molds, land plants, chlorophyte green algae (*Chlamydomonas*) and brown algae (1,2,12–14,16,17). Despite considerable conservation of miRNAs within lineages, there are no well-supported cases of miRNA loci being shared between lineages, suggesting that miRNA systems have evolved independently in each lineage, presumably from existing systems such as siRNAs. Interestingly, almost all of the organisms that have been shown to possess miRNAs exhibit some form of multicellularity (*Chlamydomonas* being an exception) and, conversely, the eukaryotic groups that exhibit the highest levels of multicellular complexity—animals, land plants and brown algae (3)—all possess miRNA systems. This correlation between complex multicellularity and the presence of regulatory systems based on miRNAs has led several authors to suggest that the latter may have played a key role in the evolution of the former (4,5). This suggestion is supported by the fact that, in animals at least, developmental complexity (estimated either based on numbers of different cell types or by scoring morphological characters) is approximately correlated with the complexity of the miRNA component of the genome (50,84,85). A similar correlation can be made across eukaryotic groups. We show here that the three eukaryotic lineages that exhibit the highest levels of developmental complexity— animals, land plants and brown algae—also have considerably more complex miRNA repertoires (at least 60 miRNA loci) than less developmentally complex organisms. For example, *Drosophila*, *Arabidopsis* and *Ectocarpus* possess 110, 64 and 63 miRNA loci, respectively ((40,41) and this study). In contrast, organisms from lineages with a lower level of developmentally complexity, such as *Amphimedon* (eight miRNAs), *Dictyostelium* (11 miRNAs) and *Chlamydomonas* (10 miRNAs), have markedly fewer miRNA loci (40,41).

Comparison of miRNA structural features across eukaryotic lineages

If the miRNA systems of diverse eukaryotic lineages evolved independently from a common, ancestral small-RNA-based regulatory system (Table 2) then we would expect the different, extant miRNA systems to exhibit marked differences due to their independent evolutionary histories. To explore this prediction, structural features of the *Ectocarpus* miRNA loci were compared with those of miRNA loci identified in other lineages. On average, the *Ectocarpus* miRNA foldbacks were longer than those of any of the other eukaryotic lineages (170 nt) but were more similar to the long foldbacks of land plant (e.g. *Arabidopsis*, 136 nt), green algal (*Chlamydomonas*, 140 nt) and slime mold (*Dictyostelium*, 132 nt) miRNA loci than to the markedly shorter foldbacks (~82 nt) of eumetazoan miRNA loci (Figure 4). Note that the foldbacks of the *Amphimedon* miRNA loci were significantly longer than those of *Drosophila* or zebrafish, supporting an independent origin for the miRNAs in this lineage.

The majority of the *Ectocarpus* miRNAs were 21 nucleotides in length (84.3%), the remaining ten loci producing miRNAs of 20 (one locus) or 22 nucleotides (Figure 4). Land plants, *Chlamydomonas* and *Dictyostelium* show a similar preference for 21 nucleotide miRNAs, whereas animal and demosponge miRNAs do not show this bias. As expected, the size ranges of miRNA*s from different species followed a similar pattern to that of the miRNAs (Figure 4).

The *Ectocarpus* miRNAs also showed an exceptionally strong tendency to have a U residue at the first position (92%) whereas this was considerably less marked for the miRNA* sequences (36%). This bias was observed for all miRNAs independent of whether they corresponded to the 5p or the 3p product. The preference for U at the first position was variable across the other eukaryotic lineages (Figure 4). A strong bias was also observed for *Chlamydomonas* (80%), land plant (e.g. 74% for *Arabidopsis*), demosponge (75%) and animal (e.g. 73% for *Drosophila* and around 40% for animals in general (86)) miRNAs, whereas no bias (22%) was observed for *Dictyostelium*. None of these organisms showed a bias for a particular residue at the first position of the miRNA* (Figure 4). Note, however, that the lack of a strong bias does not necessarily mean that the miRNA* species are not selected as guide strands because different argonaute proteins may have different sequence preferences (87).

Analysis of the crystal structure of human Ago2 protein bound to miRNA has indicated that a short loop within the middle (MID) domain, called the nucleotide specificity loop, is likely to play a key role in determining preference for specific 5′ miRNA nucleotides (preference for U and A over G and C). The *Ectocarpus* genome encodes one Argonaute homologue (Esi0203_0032, Table 2), which is 39.8% identical (66.2% similar) to human Ago2. Residues involved in non-specific binding of the 5′ miRNA nucleotide, such as Ago2 Y529, Q545 and K570 are conserved in the *Ectocarpus* protein but the region corresponding to the nucleotide specificity loop is highly divergent. Structural analysis of AGO/miRNA complexes will therefore be required to determine whether steric constraints imposed by the AGO protein underlie the bias towards 5′ U residues in brown algal miRNAs.

In *Ectocarpus*, there was a weak preference for the miRNA to be located in the 3p rather than the 5p position (66%). This was also the case for *Dictyostelium* (73%) *Drosophila* (61%) and *Chlamydomonas* (60%), whereas miRNAs tended to be evenly distributed between the two positions in *Arabidopsis* (48%) and *Amphimedon* (50%).

When these various structural features are taken together, the miRNA repertoires of each eukaryotic lineage exhibit different ranges of characteristics, a pattern that is consistent with each miRNA system having an independent evolutionary origin. The *Ectocarpus* miRNA loci are more similar to land plant miRNAs in terms of their structure but resemble animal miRNA in other respects, such as their strong tendency to be located within genes for example. We also noted that the structures of animal miRNA loci are quite distinct from those of miRNA loci from all the other eukaryotic groups, in particular foldbacks are significantly shorter. This unusual structure feature of animal miRNAs may reflect a molecular constraint specific to that lineage,

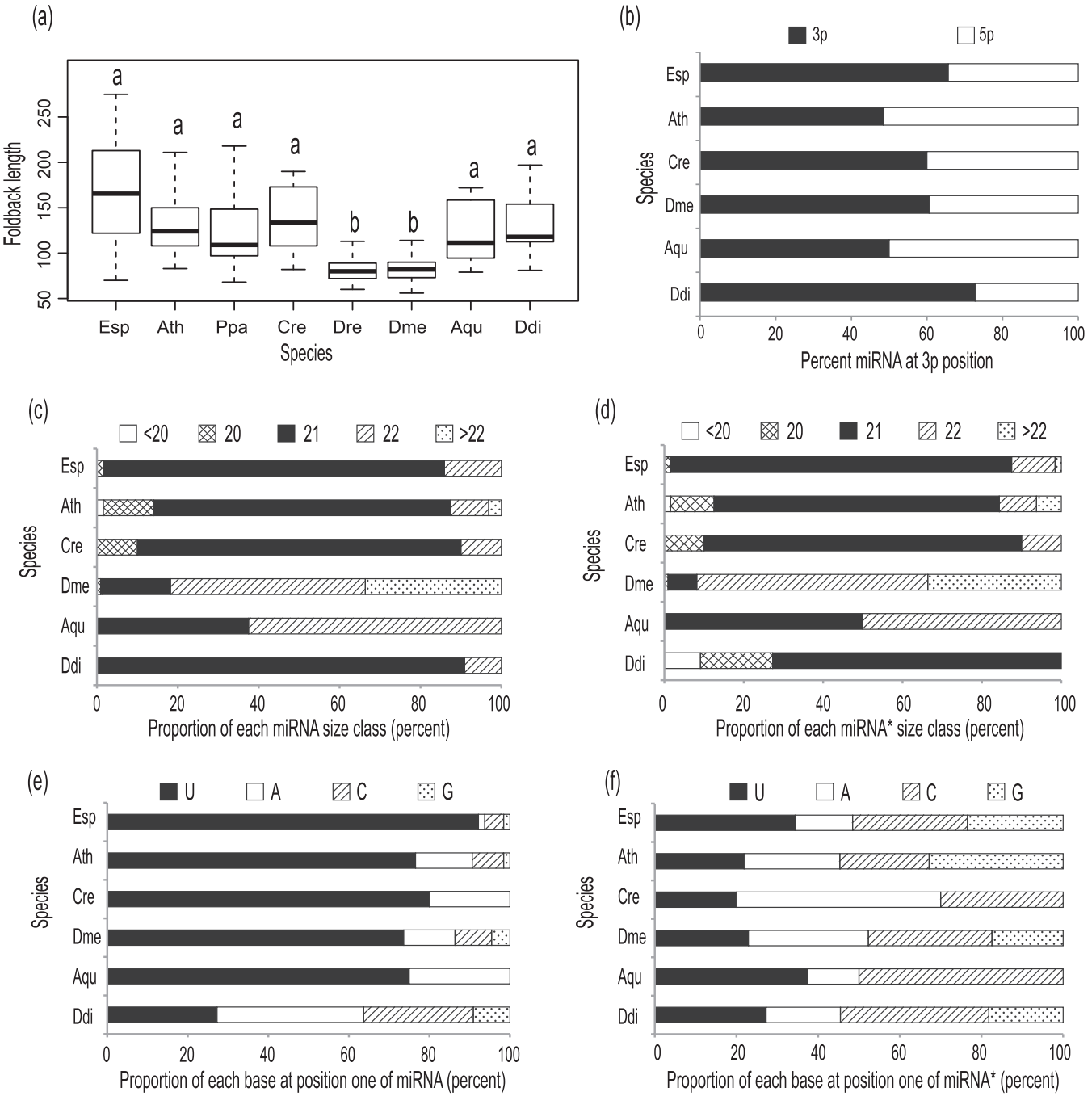


Figure 4. Structural characteristics of miRNA loci from different eukaryotic lineages. (**a**) Variation in foldback length, a and b indicate statistically different size ranges (Kruskal–Wallis test, padj = 1.2e−10), (**b**) position of the miRNA (3p or 5p) on the hairpin, (**c**, **d**) miRNA and miRNA* size distributions, (**e**, **f**) proportions of U, A, C and G at the first residue in miRNAs and miRNA*s from different lineages. The ranges of miRNA size (Kruskal-Wallis test, padj = 2.2e−16), miRNA* size (Kruskal-Wallis test, padj = 2.2e−16) and preference for a uracil residue at position one of the miRNA (Fisher exact test, $P = 0.0002$) were significantly different across species. Aqu, *Amphimedon queenslandica* (number of miRNAs = 8); Ath, *Arabidopsis thaliana* ($n = 69$); Cre, *Chlamydomonas reinhardtii* ($n = 10$); Dre, *Danio rerio* ($n = 166$); Ddi, *Dictyostelium discoideum* ($n = 11$); Dme, *Drosophila melanogaster* ($n = 110$); Esp, *Ectocarpus* sp. ($n = 64$); Ppa, *Physcomitrella patens* ($n = 40$).

such as the involvement of a dual RNAseIII Drosha/Dicer system in miRNA processing for example.

miRNA loci from different eukaryotic lineages also exhibited differences in terms of their expression. For example, on average, the miRNA product of an *Ectocarpus* miRNA locus was 446 times more abundant than the miRNA* product, allowing the two products to be clearly distinguished. Similar marked preferences for the miRNA product were observed for *Chlamydomonas* and *Arabidopsis* miRNA loci (425x and 225x, respectively) but the situation was different in *Dictyostelium* and in *Drosophila*, where mean miRNA/miRNA* abundance ratios were only 18× and 83×, respectively. The low ratio observed for *Drosophila* is consistent with the observation that both miRNA and miRNA* species have been shown to be involved in gene regulation in this species (88).

Interestingly, the 65 weak candidate *Ectocarpus* miRNAs shared a number of structural characteristics with the 64 genuine miRNAs, including a tendency to be located within protein-coding genes (67%), a strong bias towards having a U residue at the first position of the miRNA (95% for the miRNA but only 30% for the miRNA*) and a strong bias towards miRNAs that are 21 nucleotides in length (92%). These observations support the hypothesis that the weak candidate loci may represent evolving or nascent miRNA loci (7,57,89,90).

CONCLUSIONS

Analysis of sRNA read mapping and application of a set of strict criteria allowed us to demonstrate that a previously identified set of 23 *Ectocarpus* loci that had been thought to be sources of miRNAs are more probably siRNA sources. However, the same analysis also allowed the identification of a large number of previously undescribed miRNA loci bringing the total number of well-supported miRNA loci in *Ectocarpus* to 64. The identification of these new loci considerably expands the size of the miRNA complement in this organism and provides additional support for the presence of *bone fide* miRNAs in the brown algae. The 64 *Ectocarpus* miRNA loci were classified into 63 families indicating an exceptionally high level of sequence diversity compared with miRNA repertoires from other eukaryotic lineages. The *Ectocarpus* miRNA loci exhibited a number of other exceptional features including the long lengths of their foldback loops, a very strong preference for a uracil at the start of the miRNA and a very marked difference between the abundances of the miRNA and the miRNA* species.

Ectocarpus miRNA loci share features with both animal and plant miRNAs but are not homologous to the miRNAs in these other lineages, consistent with the hypothesis that miRNAs have evolved independently in each of these three lineages. This hypothesis is further supported by the absence of homologues of *Ectocarpus* miRNAs in other stramenopile genomes, suggesting that the brown algal miRNA repertoire evolved after the diversification of this eukaryotic supergroup. Given the developmental complexity of some brown algal species, the discovery of this large repertoire of miRNA loci in *Ectocarpus* also reinforces the proposed link between the acquisition of miRNAs and the emergence of complex multicellularity (3–5). It is particularly striking that the three eukaryotic lineages that exhibit the highest levels of multicellularity complexity appear to possess significantly more miRNAs than species from lineages that exhibit less developmental complexity.

An important aim for the future will be to develop methodologies to investigate the mechanism of biogenesis and to identify the cellular functions of the *Ectocarpus* miRNA loci. This study did not find any evidence for differential expression of miRNA loci in males or females or in the different generations of the life cycle. Additional analyses will be required to determine whether these genes are regulated in response to other stimuli or coincidentally with other developmental events. Another important future question concerns the evolutionary origins of these loci. Are the miRNA loci conserved in other brown algal species? Did their emergence in the stramenopile lineage predate the evolution of complex multicellularity in this group? At present, genome sampling within the stramenopiles is too sparse to allow this type of question to be addressed, but this situation is likely to change rapidly in the coming years.

Finally, there is a danger that the proliferation, in recent years, of poorly substantiated reports of miRNAs from diverse eukaryotic species, often based on the application of inappropriate methodologies, will obscure the deep evolutionary history of these key regulatory molecules. We demonstrate here the importance of combining deep sRNA read data with stringent selection criteria and a reference genome sequence for the unambiguous detection and validation of miRNA loci. We hope that this study will contribute towards the development of a generally adopted, rigorous miRNA validation mechanism and thereby, in the longer term, to an improved understanding of miRNA evolution within the eukaryotic tree.

NOTE ADDED IN PROOF

Following submission to miRBase, an additional family of two members (esi-MIR11396a and esi-MIR11396b) was identified based on similarity between hairpin sequences bringing the number of miRNA families to 62.

ACCESSION NUMBER

SRP052304.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank two anonymous reviewers for their helpful comments and suggestions for improvements to the manuscript.

FUNDING

Centre National de la Recherche Scientifique, the University Pierre and Marie Curie, the Interreg program France (Channel) – England (project Marinexus); University Pierre and Marie Curie (Emergence program); Agence National

de la Recherche (Future Investment program project Idealg and Young Investigator project Sexseaweed); Irish Research Council EMPOWER Postdoctoral Fellowship (to J.E.T.). Funding for open access charge: Agence National de la Recherche (ANR, French National Research Agency).
Conflict of interest statement. None declared.

REFERENCES

1. Voinnet,O. (2009) Origin, biogenesis, and activity of plant microRNAs. *Cell*, **136**, 669–687.
2. Carthew,R.W. and Sontheimer,E.J. (2009) Origins and mechanisms of miRNAs and siRNAs. *Cell*, **136**, 642–655.
3. Cock,J.M. and Collén,J. (2015) In Ruiz-Trillo,I. and Nedelcu,A.M. (eds.), *Evolutionary Transitions to Multicellular Life*. Springer Verlag, pp. 335–361.
4. Mattick,J.S. (2004) RNA regulation: a new genetics? *Nat. Rev. Genet.*, **5**, 316–323.
5. Peterson,K.J., Dietrich,M.R. and McPeek,M.A. (2009) MicroRNAs and metazoan macroevolution: insights into canalization, complexity, and the Cambrian explosion. *Bioessays*, **31**, 736–747.
6. Grimson,A., Srivastava,M., Fahey,B., Woodcroft,B.J., Chiang,H.R., King,N., Degnan,B.M., Rokhsar,D.S. and Bartel,D.P. (2008) Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature*, **455**, 1193–1197.
7. Axtell,M.J., Westholm,J.O. and Lai,E.C. (2011) Vive la difference: biogenesis and evolution of microRNAs in plants and animals. *Genome Biol.*, **12**, 221.
8. Kozomara,A. and Griffiths-Jones,S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.*, **39**, D152–157.
9. Ghildiyal,M. and Zamore,P.D. (2009) Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.*, **10**, 94–108.
10. Cerutti,H. and Casas-Mollano,J.A. (2006) On the origin and functions of RNA-mediated silencing: from protists to man. *Curr. Genet.*, **50**, 81–99.
11. Mukherjee,K., Campos,H. and Kolaczkowski,B. (2013) Evolution of animal and plant dicers: early parallel duplications and recurrent adaptation of antiviral RNA binding in plants. *Mol. Biol. Evol.*, **30**, 627–641.
12. Sperling,E.A., Robinson,J.M., Pisani,D. and Peterson,K.J. (2010) Where's the glass? Biomarkers, molecular clocks, and microRNAs suggest a 200-Myr missing Precambrian fossil record of siliceous sponge spicules. *Geobiology*, **8**, 24–36.
13. Molnar,A., Schwach,F., Studholme,D.J., Thuenemann,E.C. and Baulcombe,D.C. (2007) miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*. *Nature*, **447**, 1126–1129.
14. Zhao,T., Li,G., Mi,S., Li,S., Hannon,G.J., Wang,X.J. and Qi,Y. (2007) A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. *Genes Dev.*, **21**, 1190–1203.
15. Nozawa,M., Miura,S. and Nei,M. (2012) Origins and evolution of microRNA genes in plant species. *Genome Biol. Evol.*, **4**, 230–239.
16. Cock,J.M., Sterck,L., Rouzé,P., Scornet,D., Allen,A.E., Amoutzias,G., Anthouard,V., Artiguenave,F., Aury,J., Badger,J. *et al.* (2010) The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature*, **465**, 617–621.
17. Avesson,L., Reimegard,J., Wagner,E.G. and Soderbom,F. (2012) MicroRNAs in Amoebozoa: deep sequencing of the small RNA population in the social amoeba *Dictyostelium discoideum* reveals developmentally regulated microRNAs. *RNA*, **18**, 1771–1782.
18. Hinas,A., Reimegard,J., Wagner,E.G., Nellen,W., Ambros,V.R. and Soderbom,F. (2007) The small RNA repertoire of *Dictyostelium discoideum* and its regulation by components of the RNAi pathway. *Nucleic Acids Res.*, **35**, 6714–6726.
19. Lee,H.C., Li,L., Gu,W., Xue,Z., Crosthwaite,S.K., Pertsemlidis,A., Lewis,Z.A., Freitag,M., Selker,E.U., Mello,C.C. *et al.* (2010) Diverse pathways generate microRNA-like RNAs and Dicer-independent small interfering RNAs in fungi. *Mol. Cell*, **38**, 803–814.
20. Tarver,J.E., Donoghue,P.C. and Peterson,K.J. (2012) Do miRNAs have a deep evolutionary history? *Bioessays*, **34**, 857–866.
21. Rogato,A., Richard,H., Sarazin,A., Voss,B., Cheminant Navarro,S., Champeimont,R., Navarro,L., Carbone,A., Hess,W.R. and Falciatore,A. (2014) The diversity of small non-coding RNAs in the diatom *Phaeodactylum tricornutum*. *BMC Genomics*, **15**, 698.
22. Lopez-Gomollon,S., Beckers,M., Rathjen,T., Moxon,S., Maumus,F., Mohorianu,I., Moulton,V., Dalmay,T. and Mock,T. (2014) Global discovery and characterization of small non-coding RNAs in marine microalgae. *BMC Genomics*, **15**, 697.
23. Billoud,B., Nehr,Z., Le Bail,A. and Charrier,B. (2014) Computational prediction and experimental validation of microRNAs in the brown alga *Ectocarpus siliculosus*. *Nucleic Acids Res.*, **42**, 417–429.
24. Ahmed,S., Cock,J.M., Pessia,E., Luthringer,R., Cormier,A., Robuchon,M., Sterck,L., Peters,A.F., Dittami,S.M., Corre,E. *et al.* (2014) A haploid system of sex determination in the brown alga *Ectocarpus* sp. *Curr Biol*, **24**, 1945–1957.
25. Peters,A.F., Scornet,D., Ratin,M., Charrier,B., Monnier,A., Merrien,Y., Corre,E., Coelho,S.M. and Cock,J.M. (2008) Life-cycle-generation-specific developmental processes are modified in the *immediate upright* mutant of the brown alga *Ectocarpus siliculosus*. *Development*, **135**, 1503–1512.
26. Coelho,S.M., Scornet,D., Rousvoal,S., Peters,N.T., Dartevelle,L., Peters,A.F. and Cock,J.M. (2012) How to cultivate *Ectocarpus*. *Cold Spring Harb. Protoc.*, **2012**, 258–261.
27. Goecks,J., Nekrutenko,A. and Taylor,J. (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.*, **11**, R86.
28. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
29. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
30. Sterck,L., Billiau,K., Abeel,T., Rouzé,P. and Van de Peer,Y. (2012) ORCAE: online resource for community annotation of eukaryotes. *Nat. Methods*, **9**, 1041.
31. Friedlander,M.R., Mackowiak,S.D., Li,N., Chen,W. and Rajewsky,N. (2012) miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.*, **40**, 37–52.
32. Yang,X. and Li,L. (2011) miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. *Bioinformatics*, **27**, 2614–2615.
33. Lorenz,R., Bernhart,S.H., Honer Zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 26.
34. Edgar,R. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
35. Tamura,K., Peterson,D., Peterson,N., Stecher,G., Nei,M. and Kumar,S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.*, **28**, 2731–2739.
36. Armbrust,E., Berges,J., Bowler,C., Green,B., Martinez,D., Putnam,N., Zhou,S., Allen,A., Apt,K., Bechner,M. *et al.* (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science*, **306**, 79–86.
37. Bowler,C., Allen,A., Badger,J., Grimwood,J., Jabbari,K., Kuo,A., Maheswari,U., Martens,C., Maumus,F., Otillar,R. *et al.* (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature*, **456**, 239–244.
38. Gobler,C.J., Berry,D.L., Dyhrman,S.T., Wilhelm,S.W., Salamov,A., Lobanov,A.V., Zhang,Y., Collier,J.L., Wurch,L.L., Kustka,A.B. *et al.* (2011) Niche of harmful alga *Aureococcus anophagefferens* revealed through ecogenomics. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 4352–4357.
39. Vieler,A., Wu,G., Tsai,C.H., Bullard,B., Cornish,A.J., Harvey,C., Reca,I.B., Thornburg,C., Achawanantakun,R., Buehl,C.J. *et al.* (2012) Genome, functional gene annotation, and nuclear transformation of the heterokont oleaginous alga *Nannochloropsis oceanica* CCMP1779. *PLoS Genet.*, **8**, e1003064.
40. Tarver,J.E., Sperling,E.A., Nailor,A., Heimberg,A.M., Robinson,J.M., King,B.L., Pisani,D., Donoghue,P.C. and Peterson,K.J. (2013) miRNAs: small genes with big potential in metazoan phylogenetics. *Mol. Biol. Evol.*, **30**, 2369–2382.

41. Taylor,R.S., Tarver,J.E., Hiscock,S.J. and Donoghue,P.C. (2014) Evolutionary history of plant microRNAs. *Trends Plant Sci.*, **19**, 175–182.
42. Cuperus,J.T., Fahlgren,N. and Carrington,J.C.(2011) Evolution and functional diversification of MIRNA genes. *Plant Cell*, **23**, 431–442.
43. Pall,G.S., Codony-Servat,C., Byrne,J., Ritchie,L. and Hamilton,A. (2007) Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot. *Nucleic Acids Res.*, **35**, e60.
44. Wheeler,B.S. (2013) Small RNAs, big impact: small RNA pathways in transposon control and their effect on the host stress response. *Chromosome Res.*, **21**, 587–600.
45. Czech,B., Malone,C.D., Zhou,R., Stark,A., Schlingeheyde,C., Dus,M., Perrimon,N., Kellis,M., Wohlschlegel,J.A., Sachidanandam,R. *et al.* (2008) An endogenous small interfering RNA pathway in *Drosophila*. *Nature*, **453**, 798–802.
46. Ghildiyal,M., Seitz,H., Horwich,M.D., Li,C., Du,T., Lee,S., Xu,J., Kittler,E.L., Zapp,M.L., Weng,Z. *et al.* (2008) Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science*, **320**, 1077–1081.
47. Celton,J.M., Gaillard,S., Bruneau,M., Pelletier,S., Aubourg,S., Martin-Magniette,M.L., Navarro,L., Laurens,F. and Renou,J.P. (2014) Widespread anti-sense transcription in apple is correlated with siRNA production and indicates a large potential for transcriptional and/or post-transcriptional control. *New Phytol.*, **203**, 287–299.
48. Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
49. Hertel,J., Lindemeyer,M., Missal,K., Fried,C., Tanzer,A., Flamm,C., Hofacker,I.L. and Stadler,P.F. (2006) The expansion of the metazoan microRNA repertoire. *BMC Genomics*, **7**, 25.
50. Heimberg,A.M., Sempere,L.F., Moy,V.N., Donoghue,P.C. and Peterson,K.J. (2008) MicroRNAs and the advent of vertebrate morphological complexity. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 2946–2950.
51. Marco,A., Hooks,K. and Griffiths-Jones,S. (2012) Evolution and function of the extended miR-2 microRNA family. *RNA Biol.*, **9**, 242–248.
52. Nozawa,M., Miura,S. and Nei,M. (2010) Origins and evolution of microRNA genes in *Drosophila* species. *Genome Biol. Evol.*, **2**, 180–189.
53. Shi,W., Hendrix,D., Levine,M. and Haley,B. (2009) A distinct class of small RNAs arises from pre-miRNA-proximal regions in a simple chordate. *Nat. Struct. Mol. Biol.*, **16**, 183–189.
54. Langenberger,D., Bermudez-Santana,C., Hertel,J., Hoffmann,S., Khaitovich,P. and Stadler,P.F. (2009) Evidence for human microRNA-offset RNAs in small RNA sequencing data. *Bioinformatics*, **25**, 2298–2301.
55. Bortoluzzi,S., Biasiolo,M. and Bisognin,A. (2011) MicroRNA-offset RNAs (moRNAs): by-product spectators or functional players? *Trends Mol. Med.*, **17**, 473–474.
56. Zhang,W., Gao,S., Zhou,X., Xia,J., Chellappan,P., Zhang,X. and Jin,H. (2010) Multiple distinct small RNAs originate from the same microRNA precursors. *Genome Biol.*, **11**, R81.
57. Berezikov,E., Robine,N., Samsonova,A., Westholm,J.O., Naqvi,A., Hung,J.H., Okamura,K., Dai,Q., Bortolamiol-Becet,D., Martin,R. *et al.* (2011) Deep annotation of *Drosophila melanogaster* microRNAs yields insights into their processing, modification, and emergence. *Genome Res.*, **21**, 203–215.
58. Lau,N.C., Lim,L.P., Weinstein,E.G. and Bartel,D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, **294**, 858–862.
59. Marco,A. (2014) Sex-biased expression of microRNAs in *Drosophila melanogaster*. *Open Biol.*, **4**, 140024.
60. Luo,G.Z., Hafner,M., Shi,Z., Brown,M., Feng,G.H., Tuschl,T., Wang,X.J. and Li,X. (2012) Genome-wide annotation and analysis of zebra finch microRNA repertoire reveal sex-biased expression. *BMC Genomics*, **13**, 727.
61. Marco,A., Kozomara,A., Hui,J.H., Emery,A.M., Rollinson,D., Griffiths-Jones,S. and Ronshaugen,M. (2013) Sex-biased expression of microRNAs in *Schistosoma mansoni*. *PLoS Negl. Trop. Dis.*, **7**, e2402.
62. Ciaudo,C., Servant,N., Cognat,V., Sarazin,A., Kieffer,E., Viville,S., Colot,V., Barillot,E., Heard,E. and Voinnet,O. (2009) Highly dynamic and sex-specific expression of microRNAs during early ES cell differentiation. *PLoS Genet.*, **5**, e1000620.
63. Aryal,R., Jagadeeswaran,G., Zheng,Y., Yu,Q., Sunkar,R. and Ming,R. (2014) Sex specific expression and distribution of small RNAs in papaya. *BMC Genomics*, **15**, 20.
64. Brodersen,P., Sakvarelidze-Achard,L., Bruun-Rasmussen,M., Dunoyer,P., Yamamoto,Y.Y., Sieburth,L. and Voinnet,O. (2008) Widespread translational inhibition by plant miRNAs and siRNAs. *Science*, **320**, 1185–1190.
65. Bonnet,E., He,Y., Billiau,K. and Van de Peer,Y. (2010) TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. *Bioinformatics*, **26**, 1566–1568.
66. Li,A. and Mao,L. (2007) Evolution of plant microRNA gene families. *Cell Res.*, **17**, 212–218.
67. Rajagopalan,R., Vaucheret,H., Trejo,J. and Bartel,D.P. (2006) A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes Dev.*, **20**, 3407–3425.
68. Allen,E., Xie,Z., Gustafson,A.M., Sung,G.H., Spatafora,J.W. and Carrington,J.C. (2004) Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nat. Genet.*, **36**, 1282–1290.
69. Piriyapongsa,J. and Jordan,I.K. (2007) A family of human microRNA genes from miniature inverted-repeat transposable elements. *PLoS One*, **2**, e203.
70. Piriyapongsa,J. and Jordan,I.K. (2008) Dual coding of siRNAs and miRNAs by plant transposable elements. *RNA*, **14**, 814–821.
71. Felippes,F.F., Schneeberger,K., Dezulian,T., Huson,D.H. and Weigel,D. (2008) Evolution of *Arabidopsis thaliana* microRNAs from random sequences. *RNA*, **14**, 2455–2459.
72. Lu,J., Shen,Y., Wu,Q., Kumar,S., He,B., Shi,S., Carthew,R.W., Wang,S.M. and Wu,C.I. (2008) The birth and death of microRNA genes in *Drosophila*. *Nat. Genet.*, **40**, 351–355.
73. Campo-Paysaa,F., Semon,M., Cameron,R.A., Peterson,K.J. and Schubert,M. (2011) microRNA complements in deuterostomes: origin and evolution of microRNAs. *Evol. Dev.*, **13**, 15–27.
74. Biasiolo,M., Sales,G., Lionetti,M., Agnelli,L., Todoerti,K., Bisognin,A., Coppe,A., Romualdi,C., Neri,A. and Bortoluzzi,S. (2011) Impact of host genes and strand selection on miRNA and miRNA* expression. *PLoS One*, **6**, e23854.
75. Liang,Y., Ridzon,D., Wong,L. and Chen,C. (2007) Characterization of microRNA expression profiles in normal human tissues. *BMC Genomics*, **8**, 166.
76. Ozsolak,F., Poling,L.L., Wang,Z., Liu,H., Liu,X.S., Roeder,R.G., Zhang,X., Song,J.S. and Fisher,D.E. (2008) Chromatin structure analyses identify miRNA promoters. *Genes Dev.*, **22**, 3172–3183.
77. Berezikov,E. (2011) Evolution of microRNA diversity and regulation in animals. *Nat. Rev. Genet.*, **12**, 846–860.
78. Wheeler,B.M., Heimberg,A.M., Moy,V.N., Sperling,E.A., Holstein,T.W., Heber,S. and Peterson,K.J. (2009) The deep evolution of metazoan microRNAs. *Evol. Dev.*, **11**, 50–68.
79. Fromm,B., Worren,M.M., Hahn,C., Hovig,E. and Bachmann,L. (2013) Substantial loss of conserved and gain of novel MicroRNA families in flatworms. *Mol. Biol. Evol.*, **30**, 2619–2628.
80. Thomson,R.C., Plachetzki,D.C., Mahler,D.L. and Moore,B.R. (2014) A critical appraisal of the use of microRNA data in phylogenetics. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E3659–3668.
81. Field,D.J., Gauthier,J.A., King,B.L., Pisani,D., Lyson,T.R. and Peterson,K.J. (2014) Toward consilience in reptile phylogeny: miRNAs support an archosaur, not lepidosaur, affinity for turtles. *Evol. Dev.*, **16**, 189–196.
82. Heimberg,A.M., Cowper-Sal-lari,R., Semon,M., Donoghue,P.C. and Peterson,K.J. (2010) microRNAs reveal the interrelationships of hagfish, lampreys, and gnathostomes and the nature of the ancestral vertebrate. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 19379–19383.
83. Brown,J.W. and Sorhannus,U. (2010) A molecular genetic timescale for the diversification of autotrophic stramenopiles (Ochrophyta): substantive underestimation of putative fossil ages. *PLoS One*, **5**.
84. Erwin,D.H., Laflamme,M., Tweedt,S.M., Sperling,E.A., Pisani,D. and Peterson,K.J. (2011) The Cambrian conundrum: early divergence and later ecological success in the early history of animals. *Science*, **334**, 1091–1097.
85. Sempere,L.F., Cole,C.N., McPeek,M.A. and Peterson,K.J. (2006) The phylogenetic distribution of metazoan microRNAs: insights into

evolutionary complexity and constraint. *J. Exp. Zool. B Mol. Dev. Evol.*, **306**, 575–588.
86. Wang,B. (2013) Base Composition Characteristics of Mammalian miRNAs. *J. Nucleic Acids*, **95**, 1570.
87. Ghildiyal,M., Xu,J., Seitz,H., Weng,Z. and Zamore,P.D. (2010) Sorting of *Drosophila* small silencing RNAs partitions microRNA* strands into the RNA interference pathway. *RNA*, **16**, 43–56.
88. Okamura,K., Phillips,M.D., Tyler,D.M., Duan,H., Chou,Y.T. and Lai,E.C. (2008) The regulatory activity of microRNA* species has substantial influence on microRNA and 3′ UTR evolution. *Nat. Struct. Mol. Biol.*, **15**, 354–363.
89. Berezikov,E. (2011) Evolution of microRNA diversity and regulation in animals. *Nat. Rev. Genet.*, **12**, 846–860.
90. Axtell,M.J. (2013) Classification and comparison of small RNAs from plants. *Annu. Rev. Plant Biol.*, **64**, 137–159.

Discussion et perspectives

Les résultats de la prédiction réalisée avec miRDeep2 et les vérifications réalisées en Northern Blot ont permis de valider l'existence de soixante-trois miRNA chez *Ectocarpus* tout en invalidant une partie des prédictions obtenues lors de la première annotation du génome, qui correspondent plus probablement à des siRNA (short interfering RNA).

L'utilisation des données RNA-seq des sRNA, pour l'analyse de l'expression différentielle avec DESeq (Anders and Huber 2010) et EdgeR (Robinson et al. 2010) entre mâles et femelles chez les gamétophytes matures, n'a pas permis de détecter des miRNA différentiellement exprimés de manière significative. De même, entre sporophytes et gamétophytes, aucune évidence pour que des miRNA soient différentiellement exprimés n'a été trouvée.

La détection des cibles réalisée avec TAPIR s'est basée sur l'hypothèse que la complémentarité de séquence entre le miRNA et les mRNA cibles était élevée. L'analyse a permis de détecter cent soixante cibles potentielles dans le génome d'*Ectocarpus*, avec une partie des miRNA sans mRNA cible détecté (17), et certains miRNA possédant jusqu'à treize cibles. De manière intéressante, sept des cent soixante mRNA cibles identifiées étaient la cible de deux miRNA. L'analyse fonctionnelle réalisée sur les cibles a montré qu'un nombre important de processus cellulaires était ciblé par les différentes miRNA, avec une plus forte proportion de gènes impliqués dans des processus de régulation et de signalisation, de protéolyse, de fonction membranaire ou bien encore dans la défense.

Le peu d'homologie entre les miRNA d'*Ectocarpus* et les miRNA présentent chez les animaux et les plantes, tend à valider l'hypothèse que les miRNA ont évolué de manière indépendante dans ces différents clades. Cependant chez *Ectocarpus*, les mécanismes de biogenèse des miRNA restent encore à déterminer, de même que les fonctions cellulaires des différents miRNA détectés.

Conclusions générales et perspectives

Les différents processus d'amélioration au niveau structural et fonctionnel du génome d'*Ectocarpus* ont permis de fournir une amélioration significative de ce modèle des algues brunes. Ces différents travaux soulignent l'importance de maintenir et de faire évoluer de manière régulière l'annotation et la structure du génome afin de tenir compte de l'amélioration des connaissances dans de nombreux domaines, mais aussi dans l'évolution des technologies, afin de répondre aux demandes de chercheurs.

Les travaux réalisés sur *Ectocarpus* ont ainsi permis de fournir, cinq ans après la publication du génome, une deuxième version de l'assemblage du génome et de ses annotations. Au niveau du génome, ce dernier est désormais structuré en vingt-huit pseudo-chromosomes grâce à la carte génétique hautement résolutive, obtenue à partir de données RAD-seq. Pour l'annotation structurale des gènes, la nouvelle version apporte une très nette amélioration de la structure des gènes, principalement au niveau de la prédiction des UTR, mais aussi au niveau de la structure exonique des gènes, ainsi que la détection de nouveaux gènes. Une autre amélioration est la prédiction des isoformes associées à chaque gène, travail toujours en cours de réalisation. En plus d'une mise à jour de la structure des mRNA, la nouvelle version de l'annotation structurale comprend l'ajout des données sur différents ncRNA, en l'occurrence les snoRNA, miRNA et lncRNA. En effet, le développement des techniques de séquençage a permis de faciliter la découverte et la détection de ces différents types de ncRNA. Ainsi pour *Ectocarpus*, cela a permis de valider la présence de miRNA dans le génome et possiblement de siRNA, mais aussi la détection de lncRNA. Beaucoup cependant reste à faire afin d'étudier plus précisément les mécanismes de biogenèse et les fonctions biologiques de ces ncRNA, et les comparer aux données acquises avec celles d'autres espèces dans différents clades.

Afin de prendre en compte les modifications dans la structure des gènes codants dans la deuxième version de l'annotation et l'augmentation des informations disponibles dans les différentes bases de données, l'annotation fonctionnelle a été actualisée. Cela a permis de mettre à disposition une annotation plus précise et plus complète et d'avoir une vue plus précise de la fonction d'une grande quantité des gènes.

Finalement, cette amélioration globale des données génomiques et d'annotation du modèle *Ectocarpus* représente une base solide pour de futurs développements et de nouvelles analyses qui

permettront de poursuivre l'exploration de ce groupe phylogénétiquement original que sont les algues brunes.

Bibliographie

Abeel T, Van Parys T, Saeys Y, Galagan J, Van de Peer Y. 2012. GenomeView: a next-generation genome browser. Nucleic Acids Res. 40:e12.

Achawanantakun R, Chen J, Sun Y, Zhang Y. 2015. LncRNA-ID: Long non-coding RNA IDentification using balanced random forests. Bioinformatics:btv480.

Ahmed S, Cock JM, Pessia E, Luthringer R, Cormier A, Robuchon M, Sterck L, Peters AF, Dittami SM, Corre E, et al. 2014. A Haploid System of Sex Determination in the Brown Alga Ectocarpus sp. Curr. Biol. 24:1945–1957.

Albritton SE, Kranz A-L, Rao P, Kramer M, Dieterich C, Ercan S. 2014. Sex-Biased Gene Expression and Evolution of the X Chromosome in Nematodes. Genetics:genetics.114.163311.

Allen CE. 1917. A CHROMOSOME DIFFERENCE CORRELATED WITH SEX DIFFERENCES IN SPHAeROCARPOS. Science 46:466–467.

Amaral PP, Clark MB, Gascoigne DK, Dinger ME, Mattick JS. 2011. lncRNAdb: a reference database for long noncoding RNAs. Nucleic Acids Res. 39:D146–D151.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. Genome Biol. 11:R106.

Anders S, Pyl PT, Huber W. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169.

Ankeny RA, Leonelli S. 2011. What's so special about model organisms? Stud. Hist. Philos. Sci. Part A 42:313–323.

Arendsee ZW, Li L, Wurtele ES. 2014. Coming of age: orphan genes in plants. Trends Plant Sci. 19:698–708.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25:25–29.

Assis R, Zhou Q, Bachtrog D. 2012. Sex-biased transcriptome evolution in Drosophila. Genome Biol. Evol.

Au KF, Jiang H, Lin L, Xing Y, Wong WH. 2010. Detection of splice junctions from paired-end RNA-seq data by SpliceMap. Nucleic Acids Res. 38:4570–4578.

Au KF, Sebastiano V, Afshar PT, Durruthy JD, Lee L, Williams BA, van Bakel H, Schadt EE, Reijo-Pera RA, Underwood JG, et al. 2013. Characterization of the human ESC transcriptome by hybrid sequencing. Proc. Natl. Acad. Sci. U. S. A. 110:E4821–E4830.

Axtell MJ, Westholm JO, Lai EC. 2011. Vive la différence: biogenesis and evolution of microRNAs in plants and animals. Genome Biol. 12:221.

Bachtrog D. 2005. Sex chromosome evolution: Molecular aspects of Y-chromosome degeneration in Drosophila. Genome Res. 15:1393–1401.

Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat. Rev. Genet. 14:113–124.

Bachtrog D, Kirkpatrick M, Mank JE, McDaniel SF, Pires JC, Rice W, Valenzuela N. 2011. Are all sex chromosomes created equal? Trends Genet. 27:350–357.

Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman T-L, Hahn MW, Kitano J, Mayrose I, Ming R, et al. 2014. Sex Determination: Why So Many Ways of Doing It? PLoS Biol.

Badouin H, Hood ME, Gouzy J, Aguileta G, Siguenza S, Perlin MH, Cuomo CA, Fairhead C, Branca A, Giraud T. 2015. Chaos of Rearrangements in the Mating-Type Chromosomes of the Anther-Smut Fungus Microbotryum lychnidis-dioicae. Genetics 200:1275–1284.

Bail AL, Charrier B, De Smet I Editor. 2012. Culture Methods and Mutant Generation in the Filamentous Brown Algae Ectocarpus siliculosus. Plant Organog. Methods Protoc.:323.

Baldauf SL. 2003. The Deep Roots of Eukaryotes. Science 300:1703–1706.

Baldauf SL. 2008. An overview of the phylogeny and diversity of eukaryotes. 46:263–273.

Bartel DP. 2004. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. Cell 116:281–297.

Bartel DP. 2009. MicroRNAs: Target Recognition and Regulatory Functions. Cell 136:215–233.

Bartsch I, Wiencke C, Bischof K, Buchholz CM, Buck BH, Eggert A, Feuerpfeil P, Hanelt D, Jacobsen S, Karez R, et al. 2008. The genus Laminaria sensu lato : recent insights and developments. Eur. J. Phycol. 43:1–86.

Becheikh S, Michaud M, Thomas F, Raibaut A, Renaud F. 1998. Roles of resource and partner availability in sex determination in a parasitic copepod. Proc. R. Soc. B Biol. Sci. 265:1153–1156.

Bergero R, Charlesworth D. 2009. The evolution of restricted recombination in sex chromosomes. Trends Ecol. Evol. 24:94–102.

Bergero R, Charlesworth D. 2011. Preservation of the Y Transcriptome in a 10-Million-Year-Old Plant Sex Chromosome System. Curr. Biol. 21:1470–1474.

Bergero R, Forrest A, Kamau E, Charlesworth D. 2007. Evolutionary strata on the X chromosomes of the dioecious plant Silene latifolia: evidence from new sex-linked genes. Genetics 175:1945–1954.

Bier E, Mcginnis W. 2004. 3 Model Organisms in the Study of Development and Disease.

Billoud B, Nehr Z, Bail AL, Charrier B. 2014. Computational prediction and experimental validation of microRNAs in the brown alga Ectocarpus siliculosus. Nucleic Acids Res. 42:417–429.

Blavet N, Blavet H, Muyle A, Käfer J, Cegan R, Deschamps C, Zemp N, Mousset S, Aubourg S, Bergero R, et al. 2015. Identifying new sex-linked genes through BAC sequencing in the dioecious plant Silene latifolia. BMC Genomics

Böhne A, Sengstag T, Salzburger W. 2014. Comparative Transcriptomics in East African Cichlids Reveals Sex- and Species-Specific Expression and New Candidates for Sex Differentiation in Fishes. Genome Biol. Evol. 6:2567–2585.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Bolker JA. 2009. Exemplary and surrogate models: two modes of representation in biology. Perspect. Biol. Med. 52:485–499.

Bolton JJ. 1983. Ecoclinal variation in Ectocarpus siliculosus (Phaeophyceae) with respect to temperature growth optima and survival limits. Mar. Biol. 73:131–138.

Bonnet E, He Y, Billiau K, Peer YV de. 2010. TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. Bioinformatics 26:1566–1568.

Brent MR. 2008. Steady progress and recent breakthroughs in the accuracy of automated genome annotation. Nat. Rev. Genet. 9:62–73.

Brosseau GE. 1960. Genetic Analysis of the Male Fertility Factors on the Y Chromosome of Drosophila Melanogaster. Genetics 45:257–274.

Bull JJ. 1978. Sex Chromosomes in Haploid Dioecy: A Unique Contrast to Muller's Theory for Diploid Dioecy. Am. Nat.:245.

Bull JJ. 1983. Evolution of sex determining mechanisms. Benjamin/Cummings Pub. Co.

Bull JJ. 1985. Sex determining mechanisms: An evolutionary perspective. Experientia 41:1285–1296.

Burt DW. 2002. Origin and evolution of avian microchromosomes. Cytogenet. Genome Res. 96:97–112.

Busch S, Schmid R. 2001. Enzymes associated with beta-carboxylation in Ectocarpus siliculosus (Phaeophyceae): Are they involved in net carbon acquisition? Eur. J. Phycol. 36:61–70.

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 25:1915–1927.

Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M. 2008. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 18:188–196.

Carbone A, Zinovyev A, Képès F. 2003. Codon adaptation index as a measure of dominating codon bias. Bioinformatics 19:2005–2015.

Carvalho AB, Clark AG. 2013. Efficient identification of Y chromosome sequences in the human and Drosophila genomes. Genome Res. 23:1894–1907.

Carvalho AB, Dobo BA, Vibranovski MD, Clark AG. 2001. Identification of five new genes on the Y chromosome of Drosophila melanogaster. Proc. Natl. Acad. Sci. 98:13225–13230.

Carvalho AB, Lazzaro BP, Clark AG. 2000. Y chromosomal fertility factors kl-2 and kl-3 of Drosophila melanogaster encode dynein heavy chain polypeptides. Proc. Natl. Acad. Sci. 97:13239–13244.

Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, Holland TA, Keseler IM, Kothari A, Kubo A, et al. 2014. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res. 42:D459–D471.

Cavalier-Smith T. 2002. Origins of the machinery of recombination and sex. Heredity 88:125–141.

Chang PL, Dunham JP, Nuzhdin SV, Arbeitman MN. 2011. Somatic sex-specific transcriptome differences in Drosophila revealed by whole transcriptome sequencing. BMC Genomics 12:364.

Charlesworth B. 1991. The evolution of sex chromosomes. Science 251:1030–1033.

Charlesworth D, Charlesworth B. 1980. Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. Genet. Res. 35:205–214.

Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. Heredity 95:118–128.

Charrier B, Coelho SM, Le Bail A, Tonon T, Michel G, Potin P, Kloareg B, Boyen C, Peters AF, Cock JM. 2008. Development and physiology of the brown alga Ectocarpus siliculosus: two centuries of research. New Phytol. 177:319–332.

Chen G, Wang C, Shi T. 2011. Overview of available methods for diverse RNA-Seq data analyses. Sci. China Life Sci. 54:1121–1128.

Chen N, Bellott DW, Page DC, Clark AG. 2012. Identification of avian W-linked contigs by short-read sequencing. BMC Genomics 13:183.

Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, Johnston WK, Russ C, Luo S, Babiarz JE, et al. 2010. Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes Dev. 24:992–1009.

Chibalina MV, Filatov DA. 2011. Plant Y Chromosome Degeneration Is Retarded by Haploid Purifying Selection. Curr. Biol. 21:1475–1479.

Clark AG. 1988. The Evolution of the Y Chromosome with X-Y Recombination. Genetics 119:711–720.

Cline MS, Kent WJ. 2009. Understanding genome browsing. Nat. Biotechnol. 27:153–155.

Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, Amoutzias G, Anthouard V, Artiguenave F, Aury J-M, Badger JH, et al. 2010. The Ectocarpus genome and the independent evolution of multicellularity in brown algae. Nature 465:617–621.

Coelho SM, Godfroy O, Arun A, Le Corguillé G, Peters AF, Cock JM. 2011a. Genetic regulation of life cycle transitions in the brown alga Ectocarpus. Plant Signal. Behav. 6:1858–1860.

Coelho SM, Godfroy O, Arun A, Le Corguillé G, Peters AF, Cock JM. 2011b. OUROBOROS is a master regulator of the gametophyte to sporophyte life cycle transition in the brown alga Ectocarpus. Proc. Natl. Acad. Sci. U. S. A. 108:11518–11523.

Consortium T 1000 GP. 2010. A map of human genome variation from population-scale sequencing. Nature 467:1061–1073.

Consortium TF, Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, et al. 2005. The Transcriptional Landscape of the Mammalian Genome. Science 309:1559–1563.

Consortium TU. 2015. UniProt: a hub for protein information. Nucleic Acids Res. 43:D204–D212.

Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, Grützner F, Kaessmann H. 2014. Origins and functional evolution of Y chromosomes across mammals. Nature 508:488–493.

Cutter AD, Ward S. 2005. Sexual and temporal dynamics of molecular evolution in C. elegans development. Mol. Biol. Evol. 22:178–188.

Daines B, Wang H, Wang L, Li Y, Han Y, Emmert D, Gelbart W, Wang X, Li W, Gibbs R, et al. 2011. The Drosophila melanogaster transcriptome by paired-end RNA sequencing. Genome Res. 21:315–324.

Davis MPA, van Dongen S, Abreu-Goodger C, Bartonicek N, Enright AJ. 2013. Kraken: A set of tools for quality control and analysis of high-throughput sequence data. Methods 63:41–49.

Davis TA, Volesky B, Mucci A. 2003. A review of the biochemistry of heavy metal biosorption by brown algae. Water Res. 37:4311–4330.

Dayton P. 1985. Ecology of Kelp Communities. Annu. Rev. Ecol. Syst. 16:215–245.

Derrien T, Guigó R, Johnson R. 2012. The Long Non-Coding RNAs: A New (P)layer in the "Dark Matter." Front. Genet.

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. Genome Res. 22:1775–1789.

Dietrich MR, Ankeny RA, Chen PM. 2014. Publication Trends in Model Organism Research. Genetics 198:787–794.

Dillies M-A, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, Keime C, Marot G, Castel D, Estelle J, et al. 2012. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Brief. Bioinform.

Dillwyn LW. 1809. British Confervæ; or , Colored figures and descriptions of the British plants referred by botanists to the genus Conferva. By Lewis Weston Dillwyn. London,: W. Phillips

Dittami SM, Proux C, Rousvoal S, Peters AF, Cock JM, Coppée J-Y, Boyen C, Tonon T. 2011. Microarray estimation of genomic inter-strain variability in the genus Ectocarpus (Phaeophyceae). BMC Mol. Biol. 12:2.

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, et al. 2012. Landscape of transcription in human cells. Nature 489:101–108.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21.

Duret L. 2000. tRNA gene number and codon usage in the C. elegans genome are co-adapted for optimal translation of highly expressed genes. Trends Genet. 16:287–289.

Duret L, Mouchiroud D. 2000. Determinants of Substitution Rates in Mammalian Genes: Expression Pattern Affects Selection Intensity but Not Mutation Rate. Mol. Biol. Evol. 17:68–070.

Eckalbar WL, Hutchins ED, Markov GJ, Allen AN, Corneveaux JJ, Lindblad-Toh K, Palma FD, Alföldi J, Huentelman MJ, Kusumi K. 2013. Genome reannotation of the lizard Anolis carolinensis based on 14 adult and embryonic deep transcriptomes. BMC Genomics 14:49.

Ekblom R, Wolf JBW. 2014. A field guide to whole-genome sequencing, assembly and annotation. Evol. Appl. 7:1026–1042.

Ellegren H, Parsch J. 2007. The evolution of sex-biased genes and sex-biased gene expression. Nat. Rev. Genet. 8:689–698.

Ezaz T, Stiglec R, Veyrunes F, Marshall Graves JA. 2006. Relationships between Vertebrate ZW and XY Sex Chromosome Systems. Curr. Biol. 16:R736–R743.

Fan X-N, Zhang S-W. 2015. lncRNA-MFDL: identification of human long non-coding RNAs by fusing multiple features and using deep learning. Mol. Biosyst. 11:892–897.

Fatica A, Bozzoni I. 2014. Long non-coding RNAs: new players in cell differentiation and development. Nat. Rev. Genet. 15:7–21.

Ferguson-Smith MA, Rens W. 2010. The unique sex chromosome system in platypus and echidna. Russ. J. Genet. 46:1160–1164.

Ferris P, Olson BJSC, De Hoff PL, Douglass S, Casero D, Prochnik S, Geng S, Rai R, Grimwood J, Schmutz J, et al. 2010. Evolution of an expanded sex-determining locus in Volvox. Science 328:351–354.

Fichot EB, Norman RS. 2013. Microbial phylogenetic profiling with the Pacific Biosciences sequencing platform. Microbiome 1:10.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, et al. 2014. Pfam: the protein families database. Nucleic Acids Res. 42:D222–D230.

Frankish A, Uszczynska B, Ritchie GR, Gonzalez JM, Pervouchine D, Petryszak R, Mudge JM, Fonseca N, Brazma A, Guigo R, et al. 2015. Comparison of GENCODE and RefSeq gene annotation and the impact of reference geneset on variant effect prediction. BMC Genomics 16:S2.

Friedländer MR, Chen W, Adamidi C, Maaskola J, Einspanier R, Knespel S, Rajewsky N. 2008. Discovering microRNAs from deep sequencing data using miRDeep. Nat. Biotechnol. 26:407–415.

Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky N. 2012. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res. 40:37–52.

Galibert F, André C, Hitte C. 2004. Le chien, un modèle pour la génétique des mammifères. MS Médecine Sci. 20:761–766.

Gallach M, Betrán E. 2011. Intralocus sexual conflict resolved through gene duplication. Trends Ecol. Evol. 26:222–228.

Gallach M, Domingues S, Betran E. 2011. Gene Duplication and the Genome Distribution of Sex-Biased Genes. Int. J. Evol. Biol.

Gan Q, Chepelev I, Wei G, Tarayrah L, Cui K, Zhao K, Chen X. 2010. Dynamic regulation of alternative splicing and chromatin structure in Drosophila gonads revealed by RNA-seq. Cell Res. 20:763–783.

Garber M, Grabherr MG, Guttman M, Trapnell C. 2011. Computational methods for transcriptome annotation and quantification using RNA-seq. Nat. Methods 8:469–477.

Gethmann RC. 1988. Crossing Over in Males of Higher Diptera (Brachycera). J. Hered. 79:344–350.

Godwin J, Luckenbach JA, Borski RJ. 2003. Ecology meets endocrinology: environmental sex determination in fishes. Evol. Dev. 5:40–49.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotech

Griffith M, Walker JR, Spies NC, Ainscough BJ, Griffith OL. 2015. Informatics for RNA Sequencing: A Web Resource for Analysis on the Cloud. PLoS Comput Biol 11:e1004393.

Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. 2003. Rfam: an RNA family database. Nucleic Acids Res. 31:439–441.

Grützner F, Rens W, Tsend-Ayush E, El-Mogharbel N, O'Brien PCM, Jones RC, Ferguson-Smith MA, Marshall Graves JA. 2004. In the platypus a meiotic chain of ten sex chromosomes shares genes with the bird Z and mammal X chromosomes. Nature 432:913–917.

Gschloessl B, Guermeur Y, Cock JM. 2008. HECTAR: a method to predict subcellular targeting in heterokonts. BMC Bioinformatics 9:393.

Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, et al. 2011. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature 477:295–300.

Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, Fan L, Koziol MJ, Gnirke A, Nusbaum C, et al. 2010. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat. Biotechnol. 28:503–510.

Haas BJ, Delcher AL, Mount SM, Wortman JR, Jr RKS, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, et al. 2003. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 31:5654–5666.

Haas BJ, Wortman JR, Ronning CM, Hannick LI, Smith RK, Maiti R, Chan AP, Yu C, Farzad M, Wu D, et al. 2005. Complete reannotation of the Arabidopsis genome: methods, tools, protocols and the final release. BMC Biol. 3:7.

Hackenberg M, Sturm M, Langenberger D, Falcón-Pérez JM, Aransay AM. 2009. miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. Nucleic Acids Res. 37:W68–W76.

Hambuch TM, Parsch J. 2005. Patterns of synonymous codon usage in Drosophila melanogaster genes with sex-biased expression. Genetics 170:1691–1700.

Ha M, Kim VN. 2014. Regulation of microRNA biogenesis. Nat. Rev. Mol. Cell Biol. 15:509–524.

Hardcastle TJ, Kelly KA. 2010. baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics 11:422.

Hare B, Brown M, Williamson C, Tomasello M. 2002. The Domestication of Social Cognition in Dogs. Science 298:1634–1636.

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, et al. 2012. GENCODE: The reference human genome annotation for The ENCODE Project. Genome Res. 22:1760–1774.

Hartl DL. 2000. Fly meets shotgun: shotgun wins. Nat. Genet. 24:327–328.

Heesch S, Cho GY, Peters AF, Le Corguillé G, Falentin C, Boutet G, Coëdel S, Jubin C, Samson G, Corre E, et al. 2010. A sequence-tagged genetic map for the brown alga Ectocarpus siliculosus provides large-scale assembly of the genome sequence. New Phytol. 188:42–51.

Heesch S, Day JG, Yamagishi T, Kawai H, Müller DG, Küpper FC. 2012. Cryopreservation of the Model Alga Ectocarpus (Phaeophyceae). Cryoletters 33:327–336.

He L, Hannon GJ. 2004. MicroRNAs: small RNAs with a big role in gene regulation. Nat. Rev. Genet. 5:522–531.

Henry VJ, Bandrowski AE, Pepin A-S, Gonzalez BJ, Desfeux A. 2014. OMICtools: an informative directory for multi-omic data analysis. Database J. Biol. Databases Curation 2014.

Hesse MB. 1963. Models and analogies in science. Sheed and Ward

Hestenes D. 1987. Toward a modeling theory of physics instruction. Am. J. Phys. 55:440–454.

Hingamp P, Brochier C, Talla E, Gautheret D, Thieffry D, Herrmann C. 2008. Metagenome Annotation Using a Distributed Grid of Undergraduate Students. PLoS Biol 6:e296.

Hofacker IL. 2002. RNA Secondary Structure Analysis Using the Vienna RNA Package. In: Current Protocols in Bioinformatics. John Wiley & Sons, Inc.

Hoffmann R, Seidl T, Dugas M. 2002. Profound effect of normalization on detection of differentially expressed genes in oligonucleotide microarray data analysis. Genome Biol. 3:research0033.1–research0033.11.

Huang P-J, Liu Y-C, Lee C-C, Lin W-C, Gan RR-C, Lyu P-C, Tang P. 2010. DSAP: deep-sequencing small RNA analysis pipeline. Nucleic Acids Res. 38:W385–W391.

Hughes AD, Kelly MS, Black KD, Stanley MS. 2012. Biogas from Macroalgae: is it time to revisit the idea? Biotechnol. Biofuels 5:86.

Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Graves T, Fulton RS, Dugan S, Ding Y, Buhay CJ, Kremitzki C, et al. 2012. Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. Nature 483:82–86.

Hu LL, Huang Y, Wang QC, Zou Q, Jiang Y. 2012. Benchmark comparison of ab initio microRNA identification methods and software. Genet. Mol. Res. 11:4525–4538.

Immler S, Otto SP. 2015. The evolution of sex chromosomes in organisms with separate haploid sexes. Evolution 69:694–708.

Jaffe DB, Butler J, Gnerre S, Mauceli E, Lindblad-Toh K, Mesirov JP, Zody MC, Lander ES. 2003. Whole-Genome Sequence Assembly for Mammalian Genomes: Arachne 2. Genome Res. 13:91–96.

Janzen FJ, Phillips PC. 2006. Exploring the evolution of environmental sex determination, especially in reptiles. J. Evol. Biol. 19:1775–1784.

Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen C-F, Thomas MA, Haussler D, Jacob HJ. 2004. Comparative Recombination Rates in the Rat, Mouse, and Human Genomes. Genome Res. 14:528–538.

Jiang P, Wu H, Wang W, Ma W, Sun X, Lu Z. 2007. MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features. Nucleic Acids Res. 35:W339–W344.

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinforma. Oxf. Engl. 30:1236–1240.

Jones-Rhoades MW, Bartel DP, Bartel B. 2006. MicroRNAs AND THEIR REGULATORY ROLES IN PLANTS. Annu. Rev. Plant Biol. 57:19–53.

Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28:27–30.

Karro JE, Yan Y, Zheng D, Zhang Z, Carriero N, Cayting P, Harrrison P, Gerstein M. 2007. Pseudogene.org: a comprehensive database and comparison platform for pseudogene annotation. Nucleic Acids Res. 35:D55–D60.

Kawai J, Shinagawa A, Shibata K, Yoshino M, Itoh M, Ishii Y, Arakawa T, Hara A, Fukunishi Y, Konno H, et al. 2001. Functional annotation of a full-length mouse cDNA collection. Nature 409:685–690.

Keith JM ed. 2008. UNAFold - Springer. In: Methods in Molecular Biology™. Humana Press.

Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Pääbo S. 2005. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. Science 309:1850–1854.

Kijjoa A, Sawangwong P. 2004. Drugs and Cosmetics from the Sea. Mar. Drugs 2:73–82.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14:R36.

Kim VN, Han J, Siomi MC. 2009. Biogenesis of small RNAs in animals. Nat. Rev. Mol. Cell Biol. 10:126–139.

Kiuchi T, Koga H, Kawamoto M, Shoji K, Sakai H, Arai Y, Ishihara G, Kawaoka S, Sugano S, Shimada T, et al. 2014. A single female-specific piRNA is the primary determiner of sex in the silkworm. Nature 509:633–636.

Kondo M, Nagao E, Mitani H, Shima A. 2001. Differences in recombination frequencies during female and male meioses of the sex chromosomes of the medaka, Oryzias latipes. Genet. Res. 78:23–30.

Konerding D. 2004. An Essential Guide to the Basic Local Alignment Search Tool: BLAST. Brief. Bioinform. 5:93–94.

Kosswig C. 1964. Polygenic sex determination. Experientia 20:190–199.

Kozomara A, Griffiths-Jones S. 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 42:D68–D73.

Krol J, Loedige I, Filipowicz W. 2010. The widespread regulation of microRNA biogenesis, function and decay. Nat. Rev. Genet. 11:597–610.

Kvam VM, Liu P, Si Y. 2012. A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. Am. J. Bot.

La Barre S, Potin P, Leblanc C, Delage L. 2010. The Halogenated Metabolism of Brown Algae (Phaeophyta), Its Biological Importance and Its Environmental Significance. Mar. Drugs 8:988–1010.

Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35:3100–3108.

Lahn B, Page D. 1999. Four Evolutionary Strata on the Human X Chromosome. Science 286:964–967.

Lahn BT, Page DC. 1997. Functional Coherence of the Human Y Chromosome. Science 278:675–680.

Lahn BT, Page DC. 1999. Four evolutionary strata on the human X chromosome. Science 286:964–967.

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, et al. 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. 40:D1202–D1210.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357–359.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10:R25.

Le Bail A, Dittami SM, de Franco P-O, Rousvoal S, Cock MJ, Tonon T, Charrier B. 2008. Normalisation genes for expression analyses in the brown alga model Ectocarpus siliculosus. BMC Mol. Biol. 9:75.

Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, Buels RM, Stein L, Holmes IH, Elsik CG, Lewis SE. 2013. Web Apollo: a web-based genomic annotation editing platform. Genome Biol. 14:R93.

Lee RC, Feinbaum RL, Ambros V. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75:843–854.

Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BMG, Haag JD, Gould MN, Stewart RM, Kendziorski C. 2013. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics 29:1035–1043.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930.

Li A, Zhang J, Zhou Z. 2014. PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. BMC Bioinformatics 15:311.

Lien S, Szyda J, Schechinger B, Rappold G, Arnheim N. 2000. Evidence for Heterogeneity in Recombination in the Human Pseudoautosomal Region: High Resolution Analysis by Sperm Typing and Radiation-Hybrid Mapping. Am. J. Hum. Genet. 66:557–566.

Liew WC, Orbán L. 2014. Zebrafish sex: a complicated affair. Brief. Funct. Genomics 13:172–187.

Lim LP, Lau NC, Weinstein EG, Abdelhakim A, Yekta S, Rhoades MW, Burge CB, Bartel DP. 2003. The microRNAs of Caenorhabditis elegans. Genes Dev. 17:991–1008.

Lipinska A, Cormier A, Luthringer R, Peters AF, Corre E, Gachon CMM, Cock JM, Coelho SM. 2015. Sexual Dimorphism and the Evolution of Sex-Biased Gene Expression in the Brown Alga Ectocarpus. Mol. Biol. Evol. 32:1581–1597.

Liu Y, Zhou J, White KP. 2014. RNA-seq differential expression studies: more sequence or more replication? Bioinformatics 30:301–304.

Lomsadze A, Burns PD, Borodovsky M. 2014. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. Nucleic Acids Res. 42:e119–e119.

Lorenzi HA, Puiu D, Miller JR, Brinkac LM, Amedeo P, Hall N, Caler EV. 2010. New assembly, reannotation and analysis of the Entamoeba histolytica genome reveal new genomic features and protein content information. PLoS Negl. Trop. Dis. 4:e716.

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

Lu B, Zeng Z, Shi T. 2013. Comparative study of de novo assembly and genome-guided assembly strategies for transcriptome reconstruction based on RNA-Seq. Sci. China Life Sci. 56:143–155.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, et al. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience 1:18.

Luthringer R, Cormier A, Ahmed S, Peters AF, Cock JM, Coelho SM. 2014. Sexual dimorphism in the brown algae. Perspect. Phycol. 1:11–25.

Maier I. 1995. Brown algal pheromones. Prog. Phycol. Res. 11:51–102.

Maier I. 1997a. The fine structure of the male gamete of Ectocarpus siliculosus (Ectocarpales, Phaeophyceae). II. The flagellar apparatus. Eur. J. Phycol. 32:255.

Maier I. 1997b. The fine structure of the male gamete of Ectocarpus siliculosus (Ectocarpales, Phaeophyceae). I. General structure of the cell. Eur. J. Phycol. 32:241–253.

Malone J, Oliver B. 2011. Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biol. 9:34.

Martins MJF, Mota CF, Pearson GA. 2013. Sex-biased gene expression in the brown alga Fucus vesiculosus. BMC Genomics 14:294.

Matasci N, Hung L-H, Yan Z, Carpenter EJ, Wickett NJ, Mirarab S, Nguyen N, Warnow T, Ayyampalayam S, Barker M, et al. 2014. Data access for the 1,000 Plants (1KP) project. GigaScience 3:17.

Mazumder R, Natale DA, Julio JAE, Yeh L-S, Wu CH. 2010. Community annotation in biology. Biol. Direct 5:12.

McCarthy DJ, Humburg P, Kanapin A, Rivas MA, Gaulton K, $author.lastName $author firstName, Cazier J-B, Donnelly P. 2014. Choice of transcripts and software has a large effect on variant annotation. Genome Med. 6:26.

McDaniel SF, Neubig KM, Payton AC, Quatrano RS, Cove DJ. 2013. Recent gene-capture on the UV sex chromosomes of the moss Ceratodon purpureus. Evol. Int. J. Org. Evol. 67:2811–2822.

McDaniel SF, Willis JH, Shaw AJ. 2007. A Linkage Map Reveals a Complex Basis for Segregation Distortion in an Interpopulation Cross in the Moss Ceratodon purpureus. Genetics 176:2489–2500.

McGettigan PA. 2013. Transcriptomics in the RNA-seq era. Curr. Opin. Chem. Biol. 17:4–11.

Meng Y, Shao C, Wang H, Chen M. 2012. Are all the miRBase-registered microRNAs true? RNA Biol. 9:249–253.

Moore EC, Roberts RB. 2013. Polygenic sex determination. Curr. Biol. 23:R510–R512.

Morozova O, Hirst M, Marra MA. 2009. Applications of New Sequencing Technologies for Transcriptome Analysis. Annu. Rev. Genomics Hum. Genet. 10:135–151.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Meth 5:621–628.

Müller B, Grossniklaus U. 2010. Model organisms — A historical perspective. J. Proteomics 73:2054–2063.

Müller DG. 1967. Generationswechsel, Kernphasenwechsel und Sexualität der Braunalge Ectocarpus siliculosus im Kulturversuch. Planta 75:39–54.

Müller DG. 1976. Relative sexuality in Ectocarpus siliculosus. Arch. Microbiol. 109:89–94.

Mungall CJ, Emmert DB. 2007. A Chado case study: an ontology-based modular schema for representing genome-associated biological information. Bioinformatics 23:i337–i346.

Muyle A, Zemp N, Deschamps C, Mousset S, Widmer A, Marais GAB. 2012. Rapid de novo evolution of X chromosome dosage compensation in Silene latifolia, a plant with young sex chromosomes. PLoS Biol. 10:e1001308.

Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ, et al. 2005. A Scan for Positively Selected Genes in the Genomes of Humans and Chimpanzees. PLoS Biol 3:e170.

Ørom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q, et al. 2010. Long Noncoding RNAs with Enhancer-like Function in Human Cells. Cell 143:46–58.

Oshlack A, Robinson MD, Young MD. 2010. From RNA-seq reads to differential expression results. Genome Biol. 11:220.

Oshlack A, Wakefield MJ. 2009. Transcript length bias in RNA-seq data confounds systems biology. Biol. Direct 4:14.

Otto SP, Pannell JR, Peichel CL, Ashman T-L, Charlesworth D, Chippindale AK, Delph LF, Guerrero RF, Scarpino SV, McAllister BF. 2011. About PAR: The distinct evolutionary dynamics of the pseudoautosomal region. Trends Genet. 27:358–367.

Ozsolak F, Milos PM. 2011. RNA sequencing: advances, challenges and opportunities. Nat Rev Genet 12:87–98.

Paillard B. 2011. Un état des lieux de la filière algue : parution du livre “Algues, filières du futur.”

Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression. Nat. Rev. Genet. 14:83–87.

Patel HR, Delbridge ML, Graves JAM. 2010. Organization and Evolution of the Marsupial X Chromosome. In: Deakin JE, Waters PD, Graves JAM, editors. Marsupial Genetics and Genomics. Springer Netherlands. p. 151–171.

Pennisi E. 2000. Ideas Fly at Gene-Finding Jamboree. Science 287:2182–2184.

Perry JC, Harrison PW, Mank JE. 2014. The Ontogeny and Evolution of Sex-Biased Gene Expression in Drosophila melanogaster. Mol. Biol. Evol. 31:1206–1219.

Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. [Internet] advance online publication.

Peters AF, Marie D, Scornet D, Kloareg B, Mark Cock J. 2004. PROPOSAL OF ECTOCARPUS SILICULOSUS (ECTOCARPALES, PHAEOPHYCEAE) AS A MODEL ORGANISM FOR BROWN ALGAL GENETICS AND GENOMICS1,2. J. Phycol. 40:1079–1088.

Peters AF, Scornet D, Ratin M, Charrier B, Monnier A, Merrien Y, Corre E, Coelho SM, Cock JM. 2008. Life-cycle-generation-specific developmental processes are modified in the immediate upright mutant of the brown alga Ectocarpus siliculosus. Dev. Camb. Engl. 135:1503–1512.

Pieau C, Girondot M, Richard-Mercier N, Desvages G, Dorizzi M, Zaborski P. 1994. Temperature sensitivity of sexual differentiation of gonads in the European pond turtle: Hormonal involvement. J. Exp. Zool. 270:86–94.

Pointer MA, Harrison PW, Wright AE, Mank JE. 2013. Masculinization of Gene Expression Is Associated with Exaggeration of Male Sexual Dimorphism. PLoS Genet 9:e1003697.

Prigent S, Collet G, Dittami SM, Delage L, Ethis de Corny F, Dameron O, Eveillard D, Thiele S, Cambefort J, Boyen C, et al. 2014. The genome-scale metabolic network of Ectocarpus siliculosus (EctoGEM): a resource to study brown algal physiology and beyond. Plant J. Cell Mol. Biol. 80:367–381.

Proestou DA, Goldsmith MR, Twombly S. 2008. Patterns of Male Reproductive Success in Crepidula fornicata Provide New Insight for Sex Allocation and Optimal Sex Change. Biol. Bull. 214:194–202.

Pröschel M, Zhang Z, Parsch J. 2006. Widespread Adaptive Evolution of Drosophila Genes With Sex-Biased Expression. Genetics 174:893–900.

Qiu S, Bergero R, Guirao-Rico S, Campos JL, Cezard T, Gharbi K, Charlesworth D. 2015. RAD mapping reveals an evolving, polymorphic and fuzzy boundary of a plant pseudoautosomal region. Mol. Ecol.:n/a – n/a.

Quek XC, Thomson DW, Maag JLV, Bartonicek N, Signal B, Clark MB, Gloss BS, Dinger ME. 2015. lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. Nucleic Acids Res. 43:D168–D173.

Ransom R. 1981. Computers and embryos / models in developmental biology.

Rice WR. 1987. The Accumulation of Sexually Antagonistic Genes as a Selective Agent Promoting the Evolution of Reduced Recombination between Primitive Sex Chromosomes. Evolution 41:911–914.

Rice WR. 1996a. Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. Nature 381:232–234.

Rice WR. 1996b. Evolution of the Y Sex Chromosome in Animals Y chromosomes evolve through the degeneration of autosomes. BioScience 46:331–343.

Richards S, Liu Y, Bettencourt BR, Hradecky P, Letovsky S, Nielsen R, Thornton K, Hubisz MJ, Chen R, Meisel RP, et al. 2005. Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution. Genome Res. 15:1–18.

Rinn JL, Chang HY. 2012. Genome Regulation by Long Noncoding RNAs. Annu. Rev. Biochem. 81:145–166.

Ritter A, Dittami SM, Goulitquer S, Correa JA, Boyen C, Potin P, Tonon T. 2014. Transcriptomic and metabolomic analysis of copper stress acclimation in Ectocarpus siliculosus highlights signaling and tolerance mechanisms in brown algae. BMC Plant Biol. 14:116.

Ritter A, Ubertini M, Romac S, Gaillard F, Delage L, Mann A, Cock JM, Tonon T, Correa JA, Potin P. 2010. Copper stress proteomics highlights local adaptation of two strains of the model brown alga Ectocarpus siliculosus. PROTEOMICS 10:2074–2088.

Roberts A, Pimentel H, Trapnell C, Pachter L. 2011. Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics:btr355.

Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, Mungall K, Lee S, Okada HM, Qian JQ, et al. 2010. De novo assembly and analysis of RNA-seq data. Nat. Methods 7:909–912.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140.

Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11:R25.

Ross JA, Peichel CL. 2008. Molecular cytogenetic evidence of rearrangements on the Y chromosome of the threespine stickleback fish. Genetics 179:2173–2182.

Rouyer F, Simmler M-C, Johnsson C, Vergnaud G, Cooke HJ, Weissenbach J. 1986. A gradient of sex linkage in the pseudoautosomal region of the human sex chromosomes. Nature 319:291–295.

Salse J, Bolot S, Throude M, Jouffe V, Piegu B, Quraishi UM, Calcagno T, Cooke R, Delseny M, Feuillet C. 2008. Identification and Characterization of Shared Duplications between Rice and Wheat Provide New Insight into Grass Genome Evolution. Plant Cell 20:11–24.

Sawyer SA, Parsch J, Zhang Z, Hartl DL. 2007. Prevalence of positive selection among nearly neutral amino acid replacements in Drosophila. Proc. Natl. Acad. Sci. 104:6504–6510.

Sboner A, Mu XJ, Greenbaum D, Auerbach RK, Gerstein MB. 2011. The real cost of sequencing: higher than you think! Genome Biol. 12:125.

Schiex T, Moisan A, Rouzé P. 2001. Eugène: An Eukaryotic Gene Finder That Combines Several Sources of Evidence. In: Gascuel O, Sagot M-F, editors. Computational Biology. Lecture Notes in Computer Science. Springer Berlin Heidelberg. p. 111–125.

Schmid R, Dring MJ. 1993. Rapid, Blue-Light-Induced Acidifications at the Surface of Ectocarpus and Other Marine Macroalgae. Plant Physiol. 101:907–913.

Schmieder R, Edwards R. 2011a. Quality control and preprocessing of metagenomic datasets. Bioinforma. Oxf. Engl. 27:863–864.

Schmieder R, Edwards R. 2011b. Fast Identification and Removal of Sequence Contamination from Genomic and Metagenomic Datasets. PLoS ONE 6:e17288.

Schmieder R, Lim YW, Edwards R. 2012. Identification and removal of ribosomal RNA sequences from metatranscriptomes. Bioinformatics 28:433–435.

Schmitt S, Prestel M, Paro R. 2005. Intergenic transcription through a Polycomb group response element counteracts silencing. Genes Dev. 19:697–708.

Schulz MH, Zerbino DR, Vingron M, Birney E. 2012. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28:1086–1092.

Ser JR, Roberts RB, Kocher TD. 2010. Multiple Interacting Loci Control Sex Determination in Lake Malawi Cichlid Fishes. Evol. Int. J. Org. Evol. 64:486–501.

Seyednasrollah F, Laiho A, Elo LL. 2015. Comparison of software packages for detecting differential expression in RNA-seq studies. Brief. Bioinform. 16:59–70.

Sharma E, Künstner A, Fraser BA, Zipprich G, Kottler VA, Henz SR, Weigel D, Dreyer C. 2014. Transcriptome assemblies for studying sex-biased gene expression in the guppy, Poecilia reticulata. BMC Genomics 15:400.

Sigrist CJA, Castro E de, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I. 2013. New and continuing developments at PROSITE. Nucleic Acids Res. 41:D344–D347.

Silva PC, Basson PW, Moe RL. 1996. Catalogue of the benthic marine algae of the Indian Ocean. Univ. Calif. Publ. Bot. 79:1–1259.

Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. 2014. Sequencing depth and coverage: key considerations in genomic analyses. Nat. Rev. Genet. 15:121–132.

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, et al. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825–837.

Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH. 2009. JBrowse: A next-generation genome browser. Genome Res. 19:1630–1638.

Smeds L, Kawakami T, Burri R, Bolivar P, Husby A, Qvarnström A, Uebbing S, Ellegren H. 2014. Genomic identification and characterization of the pseudoautosomal region in highly differentiated avian sex chromosomes. Nat. Commun.

Smit AJ. 2004. Medicinal and pharmaceutical uses of seaweed natural products: A review. J. Appl. Phycol. 16:245–262.

Speijer D, Lukeš J, Eliáš M. 2015. Sex is a ubiquitous, ancient, and inherent attribute of eukaryotic life. Proc. Natl. Acad. Sci. 112:8827–8834.

StacheCrain B, Muller D, Goff L. 1997. Molecular systematics of Ectocarpus and Kuckuckia (Ectocarpales, Phaeophyceae) inferred from phylogenetic analysis of nuclear- and plastid-encoded DNA sequences. J. Phycol. 33:152–168.

Standage DS, Brendel VP. 2012. ParsEval: parallel comparison and analysis of gene structure annotations. BMC Bioinformatics 13:187.

Stein L. 2001. Genome annotation: from sequence to biology. Nat. Rev. Genet. 2:493–503.

Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, Nickerson E, Stajich JE, Harris TW, Arva A, et al. 2002. The Generic Genome Browser: A Building Block for a Model Organism System Database. Genome Res. 12:1599–1610.

Steneck RS, Graham MH, Bourque BJ, Corbett D, Erlandson JM, Estes JA, Tegner MJ. 2002. Kelp forest ecosystems: biodiversity, stability, resilience and future. Environ. Conserv. null:436–459.

Stewart WDP. 1974. Algal physiology and biochemistry / edited by W. D. P. Stewart. Berkeley : University of California Press, 1974.

Strickler SR, Bombarely A, Mueller LA. 2012. Designing a transcriptome next-generation sequencing project for a nonmodel plant species1. Am. J. Bot. 99:257–266.

Sun F, Liu S, Gao X, Jiang Y, Perera D, Wang X, Li C, Sun L, Zhang J, Kaltenboeck L, et al. 2013. Male-Biased Genes in Catfish as Revealed by RNA-Seq Analysis of the Testis Transcriptome. PLoS ONE 8:e68452.

Tarazona S, García-Alcalde F, Dopazo J, Ferrer A, Conesa A. 2011. Differential expression in RNA-seq: A matter of depth. Genome Res. 21:2213–2223.

Tarver JE, Cormier A, Pinzón N, Taylor RS, Carré W, Strittmatter M, Seitz H, Coelho SM, Cock JM. 2015. microRNAs and the evolution of complex multicellularity: identification of a large, diverse complement of microRNAs in the brown alga Ectocarpus. Nucleic Acids Res. 43:6384–6398.

Tautz D, Domazet-Lošo T. 2011. The evolutionary origin of orphan genes. Nat. Rev. Genet. 12:692–702.

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. 31:46–53.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7:562–578.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotech 28:511–515.

Uebbing S, Künstner A, Mäkinen H, Ellegren H. 2013. Transcriptome Sequencing Reveals the Character of Incomplete Dosage Compensation across Multiple Tissues in Flycatchers. Genome Biol. Evol. 5:1555–1566.

Veyrunes F, Chevret P, Catalan J, Castiglia R, Watson J, Dobigny G, Robinson TJ, Britton-Davidian J. 2010. A novel sex determination system in a close relative of the house mouse. Proc. R. Soc. B Biol. Sci. 277:1049–1056.

Vicoso B, Kaiser VB, Bachtrog D. 2013. Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. Proc. Natl. Acad. Sci. U. S. A. 110:6453–6458.

Vicoso B, Zektser Y, Mahajan S, Bachtrog D. 2013. Comparative Sex Chromosome Genomics in Snakes: Differentiation, Evolutionary Strata, and Lack of Global Dosage Compensation. PLoS Biol 11:e1001643.

Volders P-J, Helsens K, Wang X, Menten B, Martens L, Gevaert K, Vandesompele J, Mestdagh P. 2013. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. Nucleic Acids Res. 41:D246–D251.

Volders P-J, Verheggen K, Menschaert G, Vandepoele K, Martens L, Vandesompele J, Mestdagh P. 2015. An update on LNCipedia: a database for annotated human lncRNA sequences. Nucleic Acids Res. 43:D174–D180.

Wajid B, Serpedin E. 2014. Do it yourself guide to genome assembly. Brief. Funct. Genomics:elu042.

Wang J, Na J-K, Yu Q, Gschwend AR, Han J, Zeng F, Aryal R, VanBuren R, Murray JE, Zhang W, et al. 2012. Sequencing papaya X and Yh chromosomes reveals molecular basis of incipient sex chromosome evolution. Proc. Natl. Acad. Sci. U. S. A. 109:13710–13715.

Wang L, Park HJ, Dasari S, Wang S, Kocher J-P, Li W. 2013. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res. 41:e74–e74.

White MJD. 1973. Animal cytology and evolution [by] M. J. D. White. Cambridge, [Eng.] University Press, 1973.

Whittle C-A, Malik MR, Krochko JE. 2007. Gender-specific selection on codon usage in plant genomes. BMC Genomics 8:169.

Williams TM, Carroll SB. 2009. Genetic and molecular insights into the development and evolution of sexual dimorphism. Nat. Rev. Genet. 10:797–804.

Wright AE, Harrison PW, Montgomery SH, Pointer MA, Mank JE. 2014. INDEPENDENT STRATUM FORMATION ON THE AVIAN SEX CHROMOSOMES REVEALS INTER-CHROMOSOMAL GENE CONVERSION AND PREDOMINANCE OF PURIFYING SELECTION ON THE W CHROMOSOME. Evol. Int. J. Org. Evol. 68:3281–3295.

Wright WG. 1988. Sex change in the Mollusca. Trends Ecol. Evol. 3:137–140.

Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26:873–881.

Wu Y, Wei B, Liu H, Li T, Rayner S. 2011. MiRPara: a SVM-based software tool for prediction of most probable microRNA coding regions in genome scale sequences. BMC Bioinformatics 12:107.

Wyman MJ, Cutter AD, Rowe L. 2012. Gene duplication in the evolution of sexual dimorphism. Evol. Int. J. Org. Evol. 66:1556–1566.

Yamato KT, Ishizaki K, Fujisawa M, Okada S, Nakayama S, Fujishita M, Bando H, Yodoya K, Hayashi K, Bando T, et al. 2007. Gene organization of the liverwort Y chromosome reveals distinct sex chromosome evolution in a haploid system. Proc. Natl. Acad. Sci. U. S. A. 104:6472–6477.

Yandell M, Ence D. 2012. A beginner's guide to eukaryotic genome annotation. Nat. Rev. Genet. 13:329–342.

Yang J-H, Zhang X-C, Huang Z-P, Zhou H, Huang M-B, Zhang S, Chen Y-Q, Qu L-H. 2006. snoSeeker: an advanced computational package for screening of guide and orphan snoRNA genes in the human genome. Nucleic Acids Res. 34:5112–5123.

Ye N, Zhang X, Miao M, Fan X, Zheng Y, Xu D, Wang J, Zhou L, Wang D, Gao Y, et al. 2015. Saccharina genomes provide novel insight into kelp biology. Nat. Commun.

Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W, Hu S, Zeng C, et al. 2005. The Genomes of Oryza sativa: A History of Duplications. PLoS Biol 3:e38.

Yu Q, Tong E, Skelton RL, Bowers JE, Jones MR, Murray JE, Hou S, Guan P, Acob RA, Luo M-C, et al. 2009. A physical map of the papaya genome with integrated genetic map and genome sequence. BMC Genomics 10:371.

Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821–829.

Zhang MQ. 2002. Computational prediction of eukaryotic protein-coding genes. Nat. Rev. Genet. 3:698–709.

Zhang Z, Hambuch TM, Parsch J. 2004. Molecular evolution of sex-biased genes in Drosophila. Mol. Biol. Evol. 21:2130–2139.

Zhang Z, Parsch J. 2005. Positive Correlation Between Evolutionary Rate and Recombination Rate in Drosophila Genes with Male-Biased Expression. Mol. Biol. Evol. 22:1945–1947.

Zhao S. 2014. Assessment of the Impact of Using a Reference Transcriptome in Mapping Short RNA-Seq Reads. PLoS ONE 9:e101374.

Zhao S, Zhang B. 2015. A comprehensive evaluation of ensembl, RefSeq, and UCSC annotations in the context of RNA-seq read mapping and gene quantification. BMC Genomics 16:97.

Zhou Q, Bachtrog D. 2012. Sex-Specific Adaptation Drives Early Sex Chromosome Evolution in Drosophila. Science 337:341–345.

Zhou Q, Zhang J, Bachtrog D, An N, Huang Q, Jarvis ED, Gilbert MTP, Zhang G. 2014. Complex evolutionary trajectories of sex chromosomes across bird taxa. Science 346:1246338.

Annexes

Annexe 1 : Sexual dimorphism in the brown algae

Sexual dimorphism in the brown algae

Auteurs : Luthringer R[1,2], Cormier A[1,2], Ahmed S[1,2], Peters AF[3], Cock JM[1,2], Coelho SM[1,2*]

Affiliations :

[1] Sorbonne Universités, UPMC Univ Paris 06, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688, Roscoff cedex, France;

[2] CNRS, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688, Roscoff cedex, France;

[3] Bezhin Rosko, 29250 Santec, France

Correspondance : coelho@sb-roscoff.fr

Article publié dans Perspectives in Phycology

doi : 10.1127/2198-011X/2014/0002

Perspectives in Phycology, Vol. 1 (2014), Issue 1, p. 11–25
Published online June 2014

Article

Sexual dimorphism in the brown algae

R. Luthringer[1,2], A. Cormier[1,2], S. Ahmed[1,2], A.F. Peters[3], J.M. Cock[1,2] & S.M. Coelho[1,2]*

[1]Sorbonne Universités, UPMC Univ Paris 06, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, 29688, Roscoff cedex, France
[2]CNRS, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, 29688, Roscoff cedex, France
[3]Bezhin Rosko, 29250 Santec, France
*Corresponding author: coelho@sb-roscoff.fr

With 3 figures and 1 table in the text and appendi

Abstract: Sexual dimorphisms have been described in several groups of organisms, but while an important number of investigations have focused on animal and plant systems, much less is known about this phenomena in other eukaryotes. We review here the current knowledge on sexual dimorphisms in the brown algae, a group of multicellular eukaryotes that have been evolving separately from animals and plants for more than a billion years. We discuss the ecological implications of these sexual dimorphisms, describe recent studies aimed at understanding the molecular basis of sex-related differences, and highlight the advantages of the brown algae to study the evolution of sexual dimorphism in a broad evolutionary context.

Keywords: sex, seaweed, evolution, sex chromosomes, isogamy, anisogamy, gamete size

Introduction

Sexual dimorphisms, which can be defined as phenotypic differences between male and female individuals of the same species, have been described to various degrees in many different groups of eukaryotic organisms. In his book on sexual selection Darwin (1871) described many examples where females and males within a single animal species differed dramatically in morphology, colouration, size, and behaviour. He proposed that gender-related differences evolved due to sexual selection resulting from variation in mating success among individuals. In recent years, there has also been a growing interest in plant sexual dimorphism (e.g. Delph et al. 2010, reviewed in Barrett & Hough 2013).

The aim of this short review is to discuss what is currently known about sexual dimorphism in brown algae, a group of multicellular eukaryotes that has evolved independently from animals and plants for more than a billion years, and to explore the potential of this group as a source of alternative model systems to study this phenomenon. We discuss the sexually dimorphic traits that have been identified in brown algae and some of the ecological implications of these dimorphisms. We also look at recent work aimed at investigating the molecular basis of sex-related differences in this group.

The brown algae exhibit a broad range of differences between male and female gametes, including isogamous (gametes of the same size), anisogamous (where the female gamete is larger than the male gamete) and oogamous species (where the female gamete is larger and non-motile). Classically, males and females are defined based on the relative size of the gametes they produce, females producing relatively few, large and usually non-motile gametes (eggs or ovules) and males producing many, small and often motile gametes (sperm or pollen). For the purpose of this review we will use the terms “male” and “female” as employed in the phycology literature, i.e. females are defined as either producing larger gametes or, in the case of morphologically isogamous species, producing gametes that quickly settle and release a pheromone to attract male gametes. Males are defined as producing smaller gametes or gametes that swim for longer, have an exploratory behaviour and respond to the female pheromone (Berthold 1881; Maier 1995). In this context, the term “isogamy” relates strictly to the gamete size, and does not take into account the physiological and behavioural differences that are consistently present in all brown algal “isogamous” lineages.

Dioicy is prevalent in the brown algae

Sexual dimorphism can only be expressed at the level of the whole thallus in species where males and females are separate individuals. Separate males and females can occur

DOI: 10.1127/2198-011X/2014/0002

either during the diploid or during the haploid phase of the life cycle, in which case the species is described as either *dioecious* or *dioicous*, respectively (see App. 1). A survey of representative species from all the main orders of the brown algae suggests that dioicy is the prevalent reproductive system in this phylogenetic group (Fig. 1). This situation contrasts markedly with that described for flowering plants, where only about 6% of species have separate sexes and this state is viewed as an evolutionary dead-end (Richards 1986; Heilbuth et al. 2001). The rarity of dioecy in flowe - ing plants may be related to the existence of widespread self-incompatibility systems in this group, as these systems allow species to be hermaphroditic without incurring problems related to inbreeding due to selfing. To date, there is little evidence for the existence of self-incompatibility systems in the brown algae (but see Gibson, 1994) and this may account at least in part for the observed difference in the frequency of dioicy. Other land plant groups also lack self-incompatibility, including for example gymnosperms, which are mostly *monoecious* but with a few lineages that include both monoecious and dioecious members (Givnish 1980). In mosses, more than half of the species are dioicous, the remainder being *hermaphrodite* (Wyatt & Anderson 1984).

Among gymnosperms, there is a strong correlation between the mode of reproduction (dioecy or monoecy) and the mode of pollen dispersal: monoecious species tend to be wind-dispersed and dioecious species to be animal-dispersed (Givnish 1980). Efforts have been made to identify similar factors that may influence or be related to reproduction mode in brown algae. Reproductive mode may indeed correlate with ecological factors, such as position on the shore, e.g. dioecious Fucales are preferentially found on the middle shore and hermaphrodites higher up the shoreline (Vernet & Harper 1980). Interestingly, it has been noted that monoicy is occasionally accompanied by the loss of sexual reproduction, at least under laboratory conditions (Müller & Meel 1982; Kuhlenkamp & Müller 1985).

Analysis of the distribution of sexual systems across the phylogenetic tree of the brown algae (Fig. 1) suggests that there have been several transitions between modes of reproduction during the evolution of this group. This conclusion is supported by several specific reports of transitions between dioicy/dioecy and monoicy/monoecy (Peters et al. 1997; Cánovas et al. 2011). The occurrence of sterile paraphyses in dioecious female *Fucus* was hypothesized to correspond to relics of the antheridium-bearing paraphyses (Billard et al. 2005), suggestive of a shift from monoecy to dioecy in this genus.

The prevalence of dioicy across the brown algal phylogeny suggests that this may have been the ancestral state for this group. A similar situation has been described for mosses, which are found to be extremely labile in their transitions between dioicy and hermaphroditism. Here, transitions to dioicy were found to occur at twice the rate of transitions to hermaphroditism at the genus level (McDaniel et al. 2013) and dioicy has also been proposed to be the ancestral state for this group (Wyatt 1982).

Traits distinguishing male and female sexes in dioicous and dioecious species of brown algae

Several sexually dimorphic traits have been described in brown algae (Table 1). These can be divided into two main classes: 1) differences between male and female gametes and 2) differences between the male and female gamete-producing stage of the life cycle (the gametophyte generation in species with haploid-diploid life cycles, see Appendix 1). We will treat these two classes of trait separately.

Most sex-related traits that have been described for male and female gametes are related to either the different functions of the two types of gamete or are a consequence of differences in gamete size. For example, during sexual reproduction in many brown algae, female gametes swim for only a short period of time before rapidly adhering to a substratum and starting to produce a sexual pheromone. The pheromone is detected by male gametes, which then swim towards and directly interact with the female gamete (Maier 1995). As a consequence of the different roles of the male and female gametes during this process, they exhibit marked sex-related differences in swimming behaviour, pheromone production, pheromone detection and cell-to-cell interaction.

The various isogamous, anisogamous and oogamous brown algal species represent a broad range of sex-related differences in gamete size. These size differences, which are thought to have evolved as a consequence of the different selection pressures on male and female gametes, also represent sexually dimorphic traits. Anisogamy and oogamy have arisen repeatedly across the eukaryotes and these systems are thought to have been derived from simpler isogamous mating systems in ancestral unicellular species (Parker et al. 1972; Kirk 2006). Somewhat surprisingly, it has also been proposed, based on phylogenetic reconstruction, that oogamy was the ancestral state in brown algae (Silberfeld et al. 2010). If this hypothesis is correct, it suggests that it may be possible for oogamy to evolve towards isogamy, despite the fact that transitions from oogamy towards isogamy are difficult to explain from a theoretical point of view (Togashi et al. 2012). Note, however, in this context that two examples of anisogamy in the primitive fucalean species *Notheia anomala* and the primitive laminarialean species *Akkesiphyus lubricus* suggest that oogamy may have arisen within these two orders (Kawai 1986; Gibson & Clayton 1987).

Differences in gamete size in anisogamous and oogamous brown algal species may influence other characteristics. In particular gamete size is likely to be one of the factors

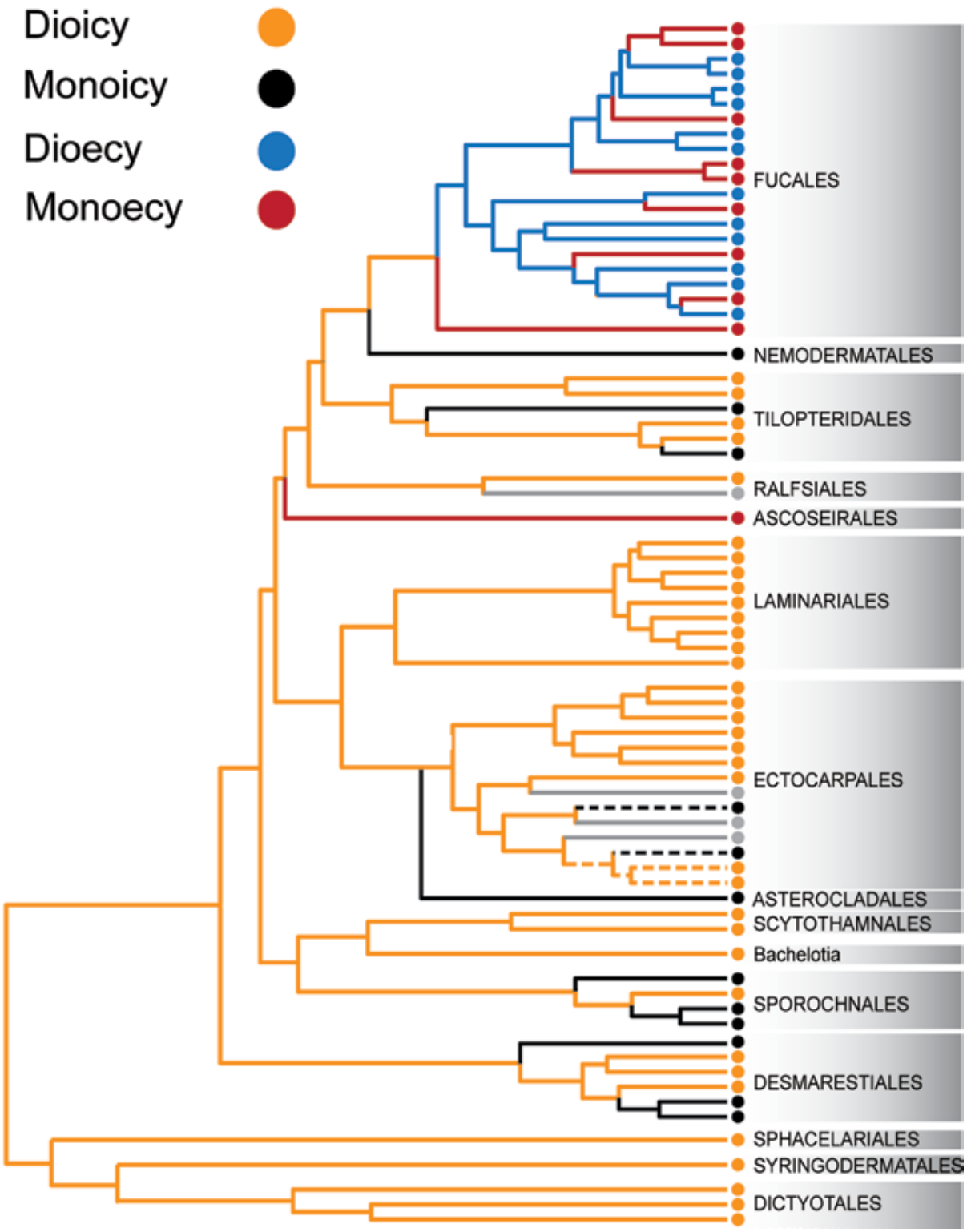


Fig. 1. Distribution of the sexual systems in the different brown algae lineages, based on the phylogenetic tree of Silberfeld et al. (2010). For simplicity, we use the terms monoicy/monoecy and dioicy/dioecy, although in some cases (some *Fucus* species for instance) the term hermaphroditism would be better adapted. The species used for this tree are the same as in Silberfeld et al. (2010) except for the following cases where species without known sexuality were replaced by closely related sexual species: *Hincksia granulosa*, *Leathesia difformis, Asperococcus bullosus, Punctaria latifolia* were replaced respectively by *Feldmannia michelliae, Chordaria linearis, Dictyosiphon foeniculaceus, Striaria attenuata*. Dashed lines were used for these species. Grey indicates lineages in which sexuality is unknown.

Table 1. Sexually dimorphic traits in brown algae. Note that gametes are considered to be parthenogenetic only if they develop into a functional individual (i.e. species whose gametes start to germinate but then degenerate were not scored as parthenogenetic). *apogamous development of sporophytes. **exceptionally yes. ***lineage according to Stache-Crain et al. (1997). [a]Most of the parthenogenetic eggs degenerate after one month. [b]A related species (*P. gracilis*) shows male and female gametophyte dimorphism. [c]A small proportion of male gametes (less than 1%) can grow parthenogenetically.

		Parthenogenesis		Gamete size		Gametophyte		Pheromone		Phototaxis		Reference
Species	**Order**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	
• *Fucus vesiculosus*	Fucales	No	No	Egg	Sperm	n.a.		Production	Attraction	Non motile	Yes	(Thuret, 1854; Overton, 1913; van den Hoek et al., 1995)
• *Notheia anomala*	Fucales	No	No	Large	Small	n.a.		*No data*		Yes	Yes	(Gibson & Clayton, 1987)
• *Nemoderma tingitanum*	Nemodermatales	*No data*	*No data*	Large	Small	n.a.		*No data*		*No data*		(Kuckuck, 1912)
• *Cutleria multifida*	Tilopteridales	Yes	No	Large	Small	*No sexual dimorphism described*		Production	Attraction	Yes	Yes	(Falkenberg, 1879; Müller, 1974; Derenbach et al., 1980; van den Hoek et al., 1995)
• *Saccorhiza polyschides*	Tilopteridales	*No data*		Egg	Sperm	Larger cells; sparsely branched	Cells smaller pale; branched	Production	Gamete release and attraction	Non motile	No	(Norton, 1969; Henry, 1987b)
• *Halosiphon tomentosus*	Tilopteridales	Yes	*No data*	Egg	Sperm	n.a.		Production	Gamete release and attraction	Non motile	Yes	(Maier, 1984; Boo et al., 1999)
• *Phyllariopsis brevipes*	Tilopteridales	*No data*		Egg	Sperm	*No sexual dimorphism described*		*No data*		Non motile	*No data*	(Henry, 1987a)

Table 1 *continued*

Species	Order	Parthenogenesis		Gamete size		Gametophyte		Pheromone		Phototaxis		Reference
		Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	
• *Ascoseira mirabilis*	Ascoseirales	Yes	Yes	Isogamy	Isogamy	Gametophyte absent		Production	Attraction not found	No	No	(Clayton, 1987; Müller et al., 1990)
• *Analipus japonicus*[a]	Ralfsiales	Yes	No	Large	Small	*No sexual dimorphism described*		Production	Gamete release and attraction	Yes	Yes	(Nakamura, 1984; Müller, 1989; Nelson & De Wreede, 1989)
• *Laminaria digitata* (and all Laminariales except *Akkesiphycus)*	Laminariales	Yes	No	Egg	Sperm	Large cells; sparsely branched	Numerous and small cells; highly branched	Production	Gamete release and attraction	Non motile	No	(Sauvageau, 1918; Oppliger et al., 2011; Shan et al., 2013)
• *Akkesiphycus lubricus*	Laminariales	Yes	No**	Large, 4–5 plastids	Small, 1 plastid	*No sexual dimorphism described*		*No data*		Yes	Yes	(Kawai, 1986)
• *Pseudochorda nagaii*[b]	Laminariales	*No data*		Egg	Sperm	*No sexual dimorphism described*		*No data*		*No data*		(Kawai & Nabata, 1990; Kawai et al., 1991)
• *Chnoospora implexa*	Ectocarpales	Yes	Yes	Isogamy	Isogamy	*No sexual dimorphism described*		Production	Attraction	Yes	Yes	(Kogame, 2001)
• *Colpomenia peregrina*	Ectocarpales	Yes	No	Large	Small	12–50µm; 4–8 loculi	20–40µm; 7–12 tiers of loculi (each 3–4 µm)	Production	Attraction	Yes	Yes	(Clayton, 1979; Müller et al., 1985a; Yamagishi & Kogame, 1998)
• *Petalonia fascia*	Ectocarpales	Yes	Yes	Isogamy	Isogamy	*No sexual dimorphism described*		Production	Attraction	Yes	Yes	(Kogame, 1997)
• *Scytosiphon lomentaria*	Ectocarpales	Yes	Yes	Isogamy	Isogamy	*No sexual dimorphism described*		Production	Attraction	Yes	Yes	(Nakamura & Tatewaki, 1975)

Table 1 *continued*

Species	Order	Parthenogenesis		Gamete size		Gametophyte		Pheromone		Phototaxis		Reference
		Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	
• *Scytosiphon canaliculatus*[c]	Ectocarpales	Yes	No	Large	Small	*No sexual dimorphism described*		Production	Attraction	Yes	Yes	Kogame pers. commun; (Kogame, 1996)
• *Ectocarpus siliculosus*[c] *(1a***)*	Ectocarpales	Yes	No	Isogamy	Isogamy	*No sexual dimorphism described*		Production	Attraction	Yes	Yes	(Berthold, 1881; Müller, 1967a; Müller, 1967b)
• *Ectocarpus siliculosus (1c***)*	Ectocarpales	Yes	Yes	Isogamy	Isogamy	*No sexual dimorphism described*		Production	Attraction	Yes	Yes	(Müller, 1967b; Müller, 1967a; Bothwell et al., 2010)
• *Feldmannia mitchelliae*	Ectocarpales	Yes	No	Large	Small	n.a.		Production	Attraction	No	No	(Müller, 1969)
• *Chordaria linearis*	Ectocarpales	Yes	Yes	Isogamy	Isogamy	n.a.		Production	Attraction	No	No	(Peters, 1992a)
• *Dictyosiphon foeniculaceus*	Ectocarpales	Yes	Yes	Isogamy	Isogamy	*No sexual dimorphism described*		Production	Attraction	Yes	Yes	(Peters & Müller, 1985; Peters, 1992b)
• *Striaria attenuata*	Ectocarpales	Yes	Yes	Isogamy	Isogamy	*No sexual dimorphism described*		Production	Attraction	Yes	Yes	(Peters et al., 2004)
• *Scytothamnus australis*	Scytothamnales	Yes	Yes	Isogamy	Isogamy	*No sexual dimorphism described*		Production	Attraction	Yes	Yes	(Clayton, 1986)
• *Splachnidium rugosum*	Scytothamnales	Yes	No	Large	Small	*No sexual dimorphism described*		*No data*		*No data*		(Clayton, 1991)
• *Himantothallus grandifolius*	Desmarestiales	Yes	No	Egg	Sperm	Large cells	Small cells	Production	Gamete release and attraction	Non motile	No	(Wiencke & Clayton, 1990)
• *Perithalia caudata*	Sporochnales	Yes	No*	Egg	Sperm	Large cells	Small cells	Production	Gamete release and attraction	Non motile	No	(Müller et al., 1985b)

Table 1 *continued*

Species	Order	Parthenogenesis		Gamete size		Gametophyte		Pheromone		Phototaxis		Reference
		Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	
• *Desmarestia aculeata (and other dioicous* Desmarestiales*)*	Desmarestiales	Yes	No	Egg	Sperm	Large cells	Small cells	Production	Gamete release and attraction	Non motile	No	(Schreiber, 1932)
• *Desmarestia firma*	Desmarestiales	Yes	No**	Egg	Sperm	Large cells, strongly pigmented	Small cells, poorly pigmented	Production	Gamete release and attraction	Non motile	No	(Anderson, 1982; Ramírez et al., 1986)
• *Phaeurus antarcticus*	Desmarestiales	No*	Yes	Egg	Sperm	Large cells, smaller fil - ment	Small cells, narrower filamen	Production	Gamete release and attraction	Non motile	No	(Clayton & Wiencke, 1990)
• *Syringoderma phinneyi*	Syringodermatales	Yes	Yes	Isogamy	Isogamy	*No sexual dimorphism described*		Production	Attraction	Yes	Yes	(Müller et al., 1982; Henry & Müller, 1983)
• *Sphacelaria rigidula*	Sphacelariales	Yes	No	Large	Small	*No sexual dimorphism described*		Production	Attraction	Yes	Yes	(van den Hoek & Flinterman, 1968)
• *Cladostephus spongiosus*	Sphacelariales	*No data*		Isogamy	Isogamy	n.a		Production	Attraction	Yes	Yes	(Müller et al. 1986; Gibson, 1994)
• *Dictyota dichotoma*	Dictyotales	Yes	No	Egg	Sperm	Broader and smaller interdichotomies	Narrow apical angles; short cortical cells; narrow medullary cells	Production	Attraction	Non motile	Yes	(Phillips et al., 1990; Tronholm et al., 2008)

that determines whether a gamete is capable of undergoing *parthenogenesis* should it fail to encounter a gamete of the opposite sex. In anisogamous and oogamous species this has led to differences between the parthenogenetic capacities of male and female gametes (Table 1). Usually both male and female gametes of isogamous brown algal species are capable of parthenogenesis whereas only the female gametes of anisogamous species are parthenogenetic (i.e. in the latter parthenogenesis is a sexually dimorphic trait). Exceptions to this trend do however exist, e.g. *Desmarestia* (Ramírez et al. 1986) or *Phaeurus* (Clayton & Wiencke 1990). Neither the male nor the female gametes undergo parthenogenesis in many oogamous species (especially in the Fucales), but there are notable exceptions in the Laminariales. Interestingly, flagella remnants have been observed in the egg cells of *Laminaria angustata* suggesting that the gametes of this species may be considered to represent an intermediate state between anisogamy and oogamy (Motomura & Sakai 1988). One interesting possibility that would merit further investigation is that the flagella remnants may play a role in female parthenogenesis in these species by allowing the formation of centrosomes in the unfertilised gamete. Overall, these trends suggest that gamete size influences parthenogenetic capacity up to a point, but that in oogamous species the large female gamete is specialised for zygote production and is no longer capable of initiating parthenogenetic development. Understanding the costs and benefits of these different reproductive strategies, particularly the incorporation of different degrees of parthenogenetic capacity in the sexual cycle, represents an interesting avenue for future research, both experimental and theoretical, and the brown algae would be a suitable group in which to study this phenomenon.

Microscopic dioicous gametophytes of species from the predominantly oogamous orders Laminariales, Desmarestiales, Sporochnales, and Tilopteridales usually show significant sexual dimorphism (Sauvageau 1915; Schreiber 1932; Müller et al. 1985b). Male gametophytes are composed of small cells and produce many gametes, whereas female gametophytes are composed of large cells and produce only a single or a small number of oocytes (Table 1, Fig. 2; Destombe & Oppliger 2011). These marked morphological differences allow rapid sexing of gametophyte clones in these groups. Exceptions to this general rule of relatively clear sexual dimorphism at the level of the gametophyte include the oogamous species *Phyllariopsis brevipes* (Tilopteridales; Henry 1987a) and *Pseudochorda nagaii* (Laminariales; Kawai & Nabata 1990) and the anisogamous species *Akkesiphycus lubricus* (Laminariales; Kawai, 1986), which have dioicous but monomorphic gametophytes (Table 1). In general, these three species have retained more ancestral characters, suggesting that the dimorphism was acquired independently in the different groups. Male and female gametophytes can also exhibit differences in terms of the timing of sexual maturation. Male gametophytes of the kelp *Alaria crassifolia* exhibit *proterandry*, antheridia of male gametophytes ripen after 4 days under favourable conditions, whereas females require 10 days (Nakahara & Nakamura, 1973). Interestingly, rather than releasing their gametes during the day in response to a light signal, oogamous species in the Laminariales, Desmarestiales, Sporochnales, and Tilopteridales release their eggs at night, which in turn induce the release of spermatozoids by producing pheromones (Table 1).

There have been no reports of sexual dimorphisms between male and female thalli of dioecious brown algal species (App. 1) such as the fucoids, but it may be necessary to carry out detailed morphometric analyses to verify that there are no subtle dimorphisms in these species.

Although future work may uncover additional sexually dimorphic traits in the brown algae, it is clear that neither brown algae nor land plants exhibit the complexity of sexual dimorphisms that have been observed in many animal groups. One of the hypotheses that have been put forward to explain the low level of sexual dimorphism in flowering plants is that because most dioecious lineages are relatively young, insufficient time has elapsed in order for marked sexual dimorphisms to have evolved in this group (Barrett & Hough 2013). This hypothesis is however unlikely to explain the low level of sexual dimorphism observed in brown algae (a least in terms of morphological complexity), as dioicy appears to be a relatively ancient

Fig. 2. Male and female gametophytes of *Laminaria digitata* in a laboratory culture (micrograph courtesy of Christophe Destombe). Male and female gametophytes are indicated by male and female symbols, respectively. The spindle or barrel-shaped single cells are diatoms.

characteristic of this group (Fig. 1). An alternative explanation may be derived from differences in the reproductive biology of algae and plants compared with animals. The former are immobile and interaction between the sexes is indirect. Most brown algae, for example, use broadcast spawning and the gametes meet and fuse in the seawater medium, without any further intervention of the gametophyte from which they originate, (except in cases where gametophyte fragmentation occurs; Destombe & Oppliger 2011). Reproductive success is assured by indirect measures such as releasing gametes at the optimal phase of the tide or by equipping gametes with efficient phototactic and pheromone systems (Maier 1995; Pearson 2006). The situation is similar for land plants, except that competition can occur between male gametes in species that receive pollen on a pistil (Pannell & Labouche 2013). In neither case, however, is there scope for the strong sexual selection that results from mate choice in motile animal species. In support of this hypothesis, it has been noted that among animals, and in particular invertebrate taxa, species that copulate generally exhibit significantly more marked levels of sexual dimorphism than species that broadcast their gametes (Strathmann 1990; Levitan 1998). Note, however, that there is nonetheless scope for sexual selection in brown algae on traits of importance for mating such as increased motility of the male gametes and higher pheromone production by the female gametes, even if there is no evidence of direct interaction between gametophytes.

Sex-dependent responses to environmental factors

In some cases, sexually dimorphic traits may be detectable only under specific, usually extreme, environmental conditions. It has been reported that abiotic factors can differentially influence the survival of male and female individuals, suggesting sex-dependent susceptibilities to the environment. Sex ratios can be modified by abiotic stresses such as salinity or temperature (Oppliger et al. 2011). In kelps, egg production takes place over a narrower range of conditions than antheridium production (Harries 1932), indicating different sensitivities of male and female gametophytes. Following exposure to high temperatures in culture, *Saccharina latissima* and *Laminaria digitata* produced a higher proportion of males (Cosson 1978; Lee & Brinkhuis 1988). Norton (1977) showed that female kelp gametophytes were more sensitive to extreme temperatures than male gametophytes, and correlated this effect with the geographical extent of the region within which sexual reproduction occurred. The opposite trend was observed for *Laminaria religiosa,* extreme temperatures resulting in a decrease in the proportion of males (Funano 1983). More recently, Nelson (2005) demonstrated that high temperature and long days resulted in a sex ratio biased toward females in *Lessonia variegata*, suggesting, again, that males were less resistant to stressful conditions. Taken together, these results suggest that the effect of temperature on sex ratio in kelps is variable and species dependent. Other factors may also affect the sex ratio, for example male and female *Saccorhiza polyschides* gametophytes showed differential sensitivities to changes in salinity (Norton & South 1969).

It is also possible that males and females respond differently to biotic factors but the limited data currently available argue against such an effect. Male and female strains of *Ectocarpus* exhibit the same susceptibility to viral infections and no difference in resistance to the oomycete pathogen *Eurychasma* has been observed between the sexes (Claire Gachon, personal communication).

Ecology

In orders with equal numbers of *monoicous* and dioicous species, such as Desmarestiales and Sporochnales, species with smaller sporophytes and a shorter life span tend to be monoicous, whereas taxa with larger sporophytes and longer lifespan are dioicous (Peters et al. 1997). In these orders, monoicy, which allows selfing, is thus favoured in r-selected species, whereas K-selected environments favour dioicy and outbreeding. *Fucus* species adapted to more stressful environments high on the shore are hermaphrodites that exhibit frequent inbreeding, in contrast to dioecious species with obligate outcrossing in more benign habitats (Billard et al. 2010). In the Ectocarpales, however, where most species are small and follow the r strategy, only a minority of taxa with known sexuality are monoicous (e.g. 10% in Chordariaceae). Additional unknown factors may underlie other differences, suggested by the observation that there are no monoicous species in the order Laminariales while monoicy is common in the orders Sporochnales, Desmarestiales, and Tilopteridales, which resemble kelps in many other aspects of their reproductive biology.

Studies of sex ratios in meiotic offspring under standard culture conditions consistently indicate a similar proportion of males and females (Sauvageau 1918; Schreiber 1932; Cosson 1978), but relatively few reports are available about brown algal sex ratios in the field. In dioecious flowering plants, females usually expend more resources in reproduction than males, and a recurrent pattern observed in this group is the presence of male-biased sex ratios in marginal populations experiencing higher levels of environmental stress (Delph 1999). In *Lessonia* (Laminariales), sex ratios were found to be favoured towards females in the limits of the distribution area (Oppliger et al. 2012). This deviation from a 1:1 ratio at the margins of the species range could be due either to differential mortality/sensitivity to temperature between sexes or to geographic variations in the degree of parthenogenesis (asexual reproduction), as females are

often parthenogenetic and males are not (Oppliger et al. 2011). Female-biased sex ratios have also been reported for some natural populations of anisogamous species (Kitayama 1992; Yamagishi & Kogame 1998), and again a correlation between female-bias and parthenogenesis has been put forward as a possible explanation. Interestingly, a link between life cycle mode and sex ratio has been reported. Populations dominated by female *Cutleria cylindrica* individuals showed a direct type of life history (spores from unilocular sporangia give rise to new sporophytes, App. 1), whereas populations with a 1:1 sex ratio presented a heteromorphic, sexual life history, alternating between sporophyte and gametophyte generations (Yamagishi & Kogame 1998). There have also been occasional reports of isogamous species in which single field sporophytes had exclusively female offspring (e.g. Müller 1979; Peters & Müller 1986; Peters et al. 1987). As both male and female gametes of these species are parthenogenetic under laboratory conditions, it is unlikely that these populations result from female gamete parthenogenesis and further studies will be required to understand how such populations arise.

Molecular mechanisms underlying sexual dimorphism in the brown algae

Sex has been shown to be determined genetically in *Ectocarpus* sp. (Müller 1967b) and heteromorphic sex chromosomes have been reported in several kelp species (Evans 1963; Yasui 1992). More recently, a putative sex-determining region has been identified in a hybrid of *Laminaria japonica* and *Laminaria longissima* (Yang et al. 2009). There is therefore accumulating evidence that sex is genetically determined in brown algae and, consequently sexual dimorphism is ultimately under the control of a specific sex-determining region (SDR) of the genome (a sex locus or a sex chromosome). Note that, in plants, transitions to dioecy are correlated with the evolution of sex chromosomes that subsequently promote the appearance of sexually dimorphic traits (Rice 1984). Identification and characterisation of SDRs in brown algal species will not only provide important insights into the evolution of sexuality and sexual dimorphism in this group but will also provide much needed molecular markers to discriminate between male and female individuals.

Based on studies of sexually dimorphic animal and plant species (e.g. Zhang et al. 2004; Mank et al. 2007) it is likely that only a small set of the genes that determine the differences between sexes are located within the SDR (although these should include the master sex-determining gene), the majority of the downstream sex-related genes being scattered throughout the genome (Ellegren & Parsch 2007). Therefore, whilst it will be important to characterise brown algal SDRs, it is also necessary to compare gene expression between the two sexes to fully understand the genetic basis of sexual dimorphism in this group. Two recent studies have carried out analyses of this type, comparing male and female individuals of *Fucus* (Martins et al. 2013) and male and female gametes of *Ectocarpus* (Lipinska et al. 2013). A general trend that has been found in both land plants and animals is that male sex-biased genes tend to be expressed more strongly than female sex-biased genes (Zhang et al. 2004) and that this appears to be correlated with male sex-biased genes being under stronger selection (exhibiting higher *dN/dS* ratios across species). This effect is thought to be due, at least in part, to widespread *pleiotropy* of female sex-biased genes (Ellegren & Parsch 2007; Mank et al. 2007). In *Fucus vesiculosus,* male sex-biased genes also exhibited greater expression bias than female sex-biased genes compared with the vegetative background, suggesting that similar processes may be operating in brown algae (Martins et al. 2013).

An analysis of sex-biased gene expression in *Ectocarpus* gametes carried out by Lipinska et al. (2013) showed more than 25% of genes were differentially expressed, which is surprising considering that this species has been reported to be isogamous. This study suggests that there may be considerable differences between male and female gametes, even when the two are morphologically indistinguishable, and raises intriguing questions regarding our perception of sexual dimorphism.

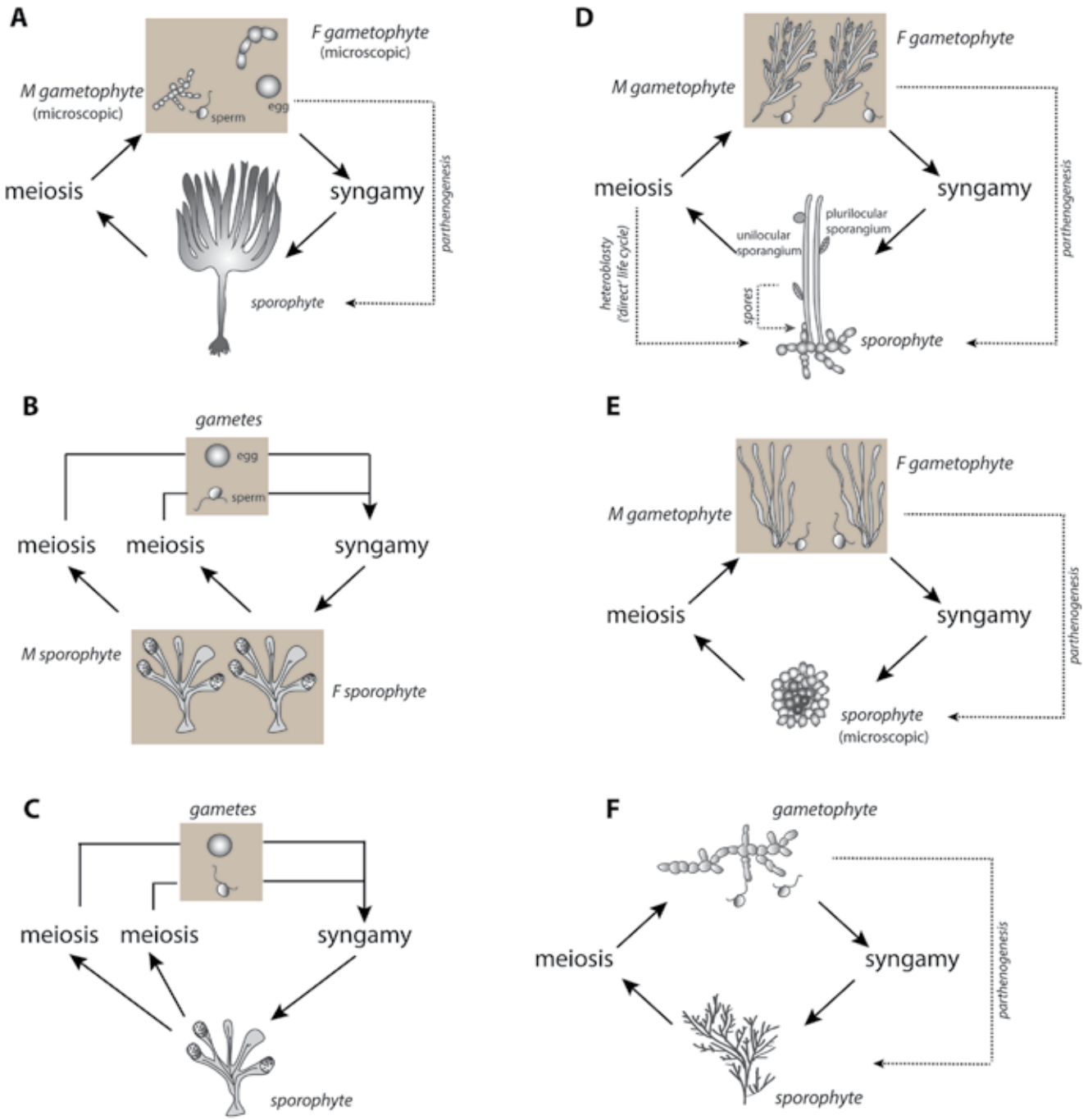
Conclusions

A number of clear sexually dimorphic traits have been described in the brown algae, observed either during the gametophyte or the gamete stage of the life cycle. In some cases these differences between male and female individuals may be important with regard to the ecology of a species, particularly at the edges of its geographical range. Despite the prevalence and probable long history of dioicy, sexual dimorphism is for most of the brown algae not as marked as in animals, possibly because the reproductive strategies of brown algae afford relatively limited scope for sexual selection. Nonetheless, the brown algae represent an interesting group for future studies of sexual dimorphism particularly with regard to gamete phenotypes as this group exhibits a broad range of gamete dimorphism from isogamous, through anisogamous, to oogamous systems. Current work aimed at identifying sex-determining regions in brown algal genomes and at comparing the transcriptomes of male and female individuals is expected to provide new insights into the molecular systems that underlie sexual dimorphisms in these seaweeds.

Appendix 1

Sexual dimorphism and brown algal life cycles

Brown algae exhibit a broad variety of life cycles, ranging from isomorphic haploid-diploid life cycles, in which both gametophyte and sporophyte generations exhibit multicellular development, to diploid life cycles, where only the diploid generation of the life cycle is multicellular (reviewed in Coelho et al. 2007; Cock et al. 2013). The ancestral brown algal sexual life cycle was presumably haploid-diploid (Silberfeld et al. 2010). In the kelps, the gametophyte generation is reduced but nonetheless develops independently of the sporophyte, and the male and female gametophytes are easily distinguishable under the microscope (**A**). In the fucoids and Ascoseira, the gametophyte generation has been lost, resulting in a diploid life cycle, with dioecious or monoecious individuals (**B** and **C**, respectively). Variations in life cycle structure occur also



within orders, for example in the Ectocarpales, which includes species with isomorphic haploid-diploid life cycles (in the Acinetosporaceae), species with slightly heteromorphic life cycles (such as *Ectocarpus*, depicted in **D**) and species with strongly heteromorphic haploid-diploid life cycles, with either the gametophyte (Chordariaceae, Adenocystaceae) or the sporophyte (Scytosiphonaceae) generation being microscopic (**E** represents an example of the latter). (**F**) Monoicous brown alga with a haploid-diploid life cycle (e.g. *Chordaria linearis*). In the figure, shaded squares represent the life cycle stages where sexual dimorphism may occur. In (**D**), *heteroblasty* refers to the development of partheno-sporophytes directly from meio-spores. M, male; F, female.

Appendix 2

Brown algae sexual systems

Brown algae exhibit a diverse range of different life cycles (Appendix 1) and this has important consequences for their sexual systems. For example, sexuality is expressed during the diploid phase in organisms with diploid life cycles such as the fucoids, whereas it is the haploid gametophyte generation that exhibits sexuality in algae such as *Ectocarpus* that have haploid-diploid life cycles (Appendix 1). Separate male and female organisms can occur in both systems but the evolutionary pathways that lead to separate sexes in each case may be very different and it is therefore important to use a nomenclature that distinguishes the two systems. The terms monoecy and dioecy are used to distinguish between species in which the diploid phase produces either both male and female gametes, on the one hand, or either male or female gametes (i.e. separate sexes), on the other. When these characteristics are observed in the haploid gametophyte generation, the terms monoicy and dioicy are used, respectively. One example of how the selection pressures that lead to the evolution of these different systems may differ is the following: whilst dioecy might evolve from monoecy to limit inbreeding (due, in the latter, to the fertilisation of female gametes by male gametes produced by the same organism), this is unlikely to be the case for dioicy because deleterious mutations should be efficiently purged during the extensive haploid phase of the life cycle. Similarly, genetic sex determination is expected to operate differently, with XX/XY or ZZ/ZW systems occurring in dioecious species but so-called U/V systems (Bachtrog et al. 2011) occurring in dioicous species.

Acknowledgements: This work was supported by the Centre National de la Recherche Scientifique, the Agence Nationale de la Recherche (Project Sexseaweed), the University Pierre and Marie Curie Emergence program, the Interreg program France (Channel)-England (project Marinexus). The funding bodies had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors wish to thank Nicolas Perrin, Myriam Valero and Christophe Destombe, for helpful comments on the manuscript. Christophe Destombe kindly provided the photograph in Figure 2.

Glossary

Dioicous: male and female sexual structures carried separately on male and female individuals during the haploid phase of the life cycle.

Dioecious: male and female sexual structures carried separately on male and female individuals during the diploid phase of the life cycle.

dN/dS: ratio of the number of non-synonymous substitutions per non-synonymous site (dN) to the number of synonymous substitutions per synonymous site (dS), which can be used as an indicator of selective pressure acting on a protein-coding gene.

Monoicous: separate male and female reproductive structures on the same individual during the haploid phase of the life cycle.

Hermaphrodite: possessing reproductive structures that contain both male and female sexual organs.

Heteroblasty: the potential of zoids to adopt different fates.

Monoecious: separate male and female reproductive structures on the same individual during the diploid phase of the life cycle.

Parthenogenesis: development of a sporophyte or gametophyte from a non-fertilized gamete. The term parthenogenesis is classically associated with female gametes, but parthenogenesis of male gametes is common in morphologically isogamous species and male gametes of anisogamous and oogamous species may also occasionally undergo parthenogenesis.

Pleiotropy: the influence that a single gene has on multiple traits.

Proterandry: release of male gametes before the release of female gametes.

References

Anderson, R.J. (1982): The life history of *Desmarestia firma* (C. Ag.) Skottsb. (Phaeophyceae, Desmarestiales). − Phycologia 21: 316–322.

Bachtrog, D., Kirkpatrick, M., Mank, J.E., McDaniel, S.F., Pires, J.C., Rice, W. & Valenzuela, N. (2011): Are all sex chromosomes created equal? − Trends Genet. 27: 350–357.

Barrett, S.C. & Hough, J. (2013): Sexual dimorphism in flowering plants. − J. Exp. Bot. 64: 67–82.

Berthold, G. (1881): Die geschlechtliche Fortpflanzung der eigentlichen Phaeosporeen. − Mitt. Zool. Stat. Neapel 2: 401–413.

Billard, E., Serrão, E., Pearson, G., Destombe, C. & Valero, M. (2010): *Fucus vesiculosus* and *spiralis* species complex: a nested model of local adaptation at the shore level. − Mar. Ecol. Prog. Series 405: 163–174.

Billard, E., Serrão, E., Pearson, G., Engel, C., Destombe, C. & Valero, M. (2005): Analysis of sexual phenotype and prezygotic fertility in natural populations of *Fucus spiralis, F. vesiculosus* (Fucaceae, Phaeophyceae) and their putative hybrids. − Eur. J. Phycol. 40: 397–407.

Boo, S.M., Lee, W.J., Yoon, H.S., Kato, A. & Kawai, H. (1999): Molecular phylogeny of Laminariales (Phaeophyceae) inferred from small subunit ribosomal DNA sequences. − Phycol. Res. 47: 109–114.

Bothwell, J.H., Marie, D., Peters, A.F., Cock, J.M. & Coelho, S.M. (2010): Role of endoreduplication and apomeiosis during parthenogenetic reproduction in the model brown alga *Ectocarpus*. − New Phytol. 188: 111–121.

Cánovas, F.G., Mota, C.F., Serrão, E.A. & Pearson, G.A. (2011): Driving south: a multi-gene phylogeny of the brown algal family Fucaceae reveals relationships and recent drivers of a marine radiation. − BMC Evol. Biol. 11: 371.

Clayton, M. (1979): The life history and sexual reproduction of *Colpomenia peregrina* (Scytosiphonaceae, Phaeophyta) in Australia. − Br. Phycol. J. 14: 1–10.

Clayton, M. (1987): Isogamy and a fucalean type of life history in the Antarctic brown alga *Ascoseira mirabilis* (Ascoseirales). − Bot. mar. 30: 447–454.

Clayton, M.N. (1986): Culture studies on the life history of *Scytothamnus australis* and *Scytothamnus fasciculatus* (Phaeophyta) with electron microscope observations on sporogenesis and gametogenesis. − Br. Phycol. J. 21: 371–386.

Clayton, M.N. (1991): Sexual reproduction and the life history of *Splachnidium rugosum* (Phaeophyceae). − Br. Phycol. J. 26: 279–294.

Clayton, M.N., Wiencke, C. (1990): The anatomy, life history and development of the Antarctic brown alga *Phaeurus antarcticus* (Desmarestiales, Phaeophyceae). − Phycologia 29: 303–315.

Cock, J.M., Godfroy, O., Macaisne, N., Peters, A.F., Coelho, S.M. (2013): Evolution and regulation of complex life cycles: a brown algal perspective. − Curr. Opin. Plant. Biol. 17: 1–6.

Coelho, S., Peters, A., Charrier, B., Roze, D., Destombe, C., Valero, M., Cock, J. (2007): Complex life cycles of multicellular eukaryotes: new approaches based on the use of model organisms. − Gene 406: 152–170.

Cosson, J. (1978): Recherches Morphogenetiques et Ecophysiologiques sur la Pheophycee *Laminaria digitata* (L.) Lamouroux. Université de Caen, Caen.

Darwin, C. (1871): The Descent of Man, and Selection in Relation to Sex. – John Murray, London.

Delph, L. (1999): Sexual dimorphism in life history. In: Geber, M.A., Dawson, T.E. & Delph, L.F. (eds), Gender and sexual dimorphism in flowering plants, pp. 149–174. − Springer Verlag, Berlin.

Delph, L.F., Arntz, A.M., Scotti-Saintagne, C. & Scotti, I. (2010): The genomic architecture of sexual dimorphism in the dioecious plant *Silene latifolia*. − Evolution 64: 2873–2886.

Derenbach, J.B., Boland, W., Fölster, E. & Müller, D.G. (1980): Interference tests with the pheromone system of the brown alga *Cutleria multifida*. − Mar. Ecol. Prog. Ser. 3: 357–361.

Destombe, C. & Oppliger, V.L. (2011): Male gametophyte fragmentation in *Laminaria digitata*: a life history strategy to enhance reproductive success. − Cahiers Biol. Mar. 52: 385–394.

Ellegren, H. & Parsch, J. (2007): The evolution of sex-biased genes and sex-biased gene expression. − Nat. Rev. Genet. 8: 689–698.

Evans, L.V. (1963): A large chromosome in the laminarian nucleus. − Nature 198: 215.

Falkenberg, P. (1879): Die Befruchtung und der Generationswechsel von *Cutleria*. − Mitt. Zool.Station Neapel 1: 420–447.

Funano, T. (1983): The ecology of *Laminaria religiosa* Miyabe. I. The life history and alternation of nuclear phases of *Laminaria religiosa*, and the physiological ecology of the gametophytes and the embryonal sporophytes. − Hokusui-Shiho 25: 61–109.

Gibson, G. & Clayton, M.N. (1987): Sexual reproduction, early development and branching in *Notheia anomala* (Phaeophyta) and its classification in the Fucales. − Phycologia 26: 363–373

Gibson, M. (1994): Reproduction in *Cladostephus spongiosus* in southern Australia (Sphacelariales, Phaeophyceae). − Phycologia 33: 378–383.

Givnish, T.J. (1980): Ecological constraints on the evolution of breeding systems in seed plants: dioecy and dispersal in gymnosperms. − Evolution 34: 959–972.

Harries, R. (1932): An investigation by cultural methods of some of the factors influencing the development of the gametophytes and early stages of the sporophytes of *Laminaria digitata, L. saccharina* and *L. cloustoni*. − Ann. Bot. 46: 893–928.

Heilbuth, J.C., Ilves, K.L. & Otto, S.P. (2001): The consequences of dioecy for seed dispersal: modeling the seed-shadow handicap. − Evolution 55: 880–888.

Henry, E.C. (1987a): The life history of *Phyllariopsis brevipes* (=*Phyllaria reniformis*) (Phyllariaceae, Laminariales, Phaeophyceae), a kelp with dioecious but sexually monomorphic gametophytes. − Phycologia 26: 17–22.

Henry, E.C. (1987b): Primitive reproductive characters and a photoperiodic response in *Saccorhiza dermatodea* (Laminariales, Phaeophyta). − Br. Phycol. J. 22: 23–113.

Henry, E.C. & Müller, D.G. (1983): Studies on the life history of *Syringoderma phinneyi* sp. nov. (Phaeophyceae). − Phycologia 22: 387–393.

Kawai, H. (1986): Life history and systematic position of *Akkesiphycus lubricus* (Phaeophyceae). − J. Phycol. 22.: 286–291.

Kawai, H., Kubota, M., Kondo, T. & Watanabe, M. (1991): Action spectra for phototaxis in zoospores of the brown alga *Pseudocorda gracilis*. − Protoplasma 161: 17–22.

Kawai, H. & Nabata, S. (1990): Life history and systematic position of *Pseudochorda gracilis* sp. nov. (Laminariales, Phaeophyceae). − J. Phycol. 26: 721–727.

Kirk, D. (2006): Oogamy: inventing the sexes. − Curr. Biol. 16: R1028–1030.

Kitayama, K. (1992): An altitudinal transect study of the vegetation on Mount Kinabalu, Borneo. − Vegetatio 102: 149–171.

Kogame, K. (1996): Morphology and life history of *Scytosiphon canaliculatus* comb. nov. (Scytosiphonales, Phaeophyceae) from Japan. − Phycol. Res. 44: 85–94.

Kogame, K. (1997): Sexual reproduction and life-history of *Petalonia fascia* (Scytosiphonales, Phaeophyceae). − Phycologia 36: 389–394.

Kogame, K. (2001): Life history of *Chnoospora implexa* (Chnoosporaceae, Phaeophyceae) in culture. − Phycol. Res. 49: 123–128.

Kuckuck, P. (1912): Beiträge zur Kenntnis der Meeresalgen. 10. Neue Untersuchungen über *Nemoderma* Schousboe. 11. Die Fortpflanzung der Phaeosporeen. − Wiss. Meeresunters. Abt. Helgoland 5: 117–154.

Kuhlenkamp, R. & Müller, D.G. (1985): Culture studies on the life history of *Haplospora globosa* and *Tilopteris mertensii* (Tilopteridales, Phaeophyceae). − British Phycol. J. 20: 301–312.

Lee, J.A. & Brinkhuis, B.H. (1988): Seasonal light and temperature interaction effects on development of *Laminaria saccharina* (Phaeophyta) gametophytes and juvenile sporophytes. − J. Phycol. 24: 181–191.

Levitan, D.R. (1998): Does Bateman's principle apply to broadcast-spawning organisms? Egg traits influence in situ fertilization rates among congeneric sea urchins. − Evolution 52: 1043–1056.

Lipinska, A.P., D'hondt, S., Van Damme, E.J.M. & De Clerck, O. (2013): Uncovering the genetic basis for early isogamete differentiation: a case study of *Ectocarpus siliculosus*. − BMC Genom 14: 909.

Maier, I. (1984): Culture studies of *Chorda tomentosa* (Phaeophyta, Laminariales). − Br. Phycol. J. 19: 95–106.

Maier, I. (1995): Brown algal pheromones. – In: Round, F.E. & Chapman, D.J. (eds), Progress in Phycological Research, pp. 51–102. − Biopress, Bristol.

Mank, J.E., Hultin-Rosenberg, L., Axelsson, E. & Ellegren, H. (2007): Rapid evolution of female-biased, but not male-biased, genes expressed in the avian brain. − Mol. Biol. Evol. 24: 2698–2706.

Martins, M.J., Mota, C.F. & Pearson, G.A. (2013): Sex-biased gene expression in the brown alga *Fucus vesiculosus*. − BMC Genom. 14: 294.

McDaniel, S.F., Atwood, J. & Burleigh, J.G. (2013): Recurrent evolution of dioecy in bryophytes. − Evolution 67: 567–572.

Motomura, T. & Sakai, Y. (1988): The occurrence of flagellated eggs in *Laminaria angustata* (Phaeophyta, Laminariales). − J. Phycol. 24: 282–285.

Müller, D. (1974): Sexual reproduction and isolation of a sex attractant in *Cutleria multifida* (Smith) Grev. (Phaeophyta). − Biochem. Physiol. Pflanz. 165: 212–215

Müller, D.G. (1967a): Ein leicht flüchtiges Gyno-Gamon der Braunalge *Ectocarpus siliculosus*. − Naturwiss. 18: 496–497.

Müller, D.G. (1967b): Generationswechsel, Kernphasenwechsel und Sexualität der Braunalge *Ectocarpus siliculosus* im Kulturversuch. − Planta 75: 39–54.

Müller, D.G. (1969): Anisogamy in *Giffordia* (Ectocarpales). − Naturwiss. 56: 220.

Müller, D.G. (1979): Genetic affinity of *Ectocarpus siliculosus* (Dillw.) Lyngb. from the Mediterranean, North Atlantic and Australia. − Phycologia 18: 312–318.

Müller, D.G. (1989): The role of pheromones in sexual reproduction of brown algae. − Algae as Experimental Systems: 201–213.

Müller, D.G., Boland, W., Marner, F.J. & Gassmann, G. (1982): Viridiene, the sexual pheromone of *Syringoderma* (Phaeophyceae). − Naturwiss. 69: 501–502.

Müller, D.G., Clayton, M.N., Gassmann, G., Boland, W., Marner, F.J., Schotten, T. & Jaenicke, L. (1985a): Cystophorene and Hormosirene, Sperm Attractants in Australian Brown-Algae. − Naturwiss. 72: 97–99.

Müller, D.G., Clayton, M.N. & Germann, I. (1985b): Sexual reproduction and life history of *Perithalia caudata* (Sporochnales, Phaeophyta). − Phycologia 24: 467–473.

Müller, D.G., Clayton, M.N., Meinderts, M., Boland, W. & Jaenicke, L. (1986): Sexual pheromone in Cladostephus (Sphacelariales, Phaeophyceae). − Naturwiss. 73: 99–100.

Müller, D.G. & Meel, H. (1982): Culture studies on the life history of *Arthrocladia villosa* (Desmarestiales, Phaeophyceae). − British Phycol. J. 17: 419–425.

Müller, D.G., Westermeier, R., Peters, A. & Boland, W. (1990): Sexual Reproduction of the Antarctic Brown Alga *Ascoseira mirabilis* (Ascoseirales, Phaeophyceae). − Bot. Mar. 33: 251–255.

Nakamura, Y. (1984): Parthenogenesis, apogamy and apospory in *Alaria crassifolia* (Laminariales). − Mar. Biol. 18: 327–332.

Nakamura, Y. & Tatewaki, M. (1975): The life history of some species of the Scytosiphonales. − Sci. Pap. Inst. Algol. Res. Hokkaido Univ. 6: 57–93.

Nelson, W.A. (2005): Life history and growth in culture of the endemic New Zealand kelp *Lessonia variegata* J. Agardh in response to differing regimes of temperature, photoperiod and light. − J. Appl. Phycol. 17: 23–28.

Nelson, W.A. & De Wreede, R.E. (1989): Reproductive phenology of *Analipus japonicus* (Harv.) Wynne (Phaeophyta) in the eastern Pacific. − Jap. J. Phycol. 37: 53–56

Norton, T.A. (1969): Growth form and environment in *Saccorhiza polyschides*. − J. Mar. Biol. Assoc.UK 49: 1025–1045.

Norton, T.A. (1977): Experiments on the factors influencing the geographical distributions of *Saccorhiza polyschides* and *Saccorhiza dermatodea*. − New Phytol. 78: 625–635.

Norton, T.A., South, G.R. (1969): The influence of salinity on the distribution of two laminarian algae. − Oikos 20: 320.

Oppliger, V.L., Correa, J.A., Faugeron, S., Beltran, J., Tellier, F., Valero, M. & Destombe, C. (2011): Sex ratio variation in the *Lessonia nigrescens* complex (Laminariales, Phaeophyceae): effect of latitude, temperature and marginality. − J. Phycol. 47: 5–12.

Oppliger, L.V., Correa, J.A., Engelen, A.H., Tellier, F., Vieira, V., Faugeron, S., Valero, M., Gomez, G. & Destombe, C. (2012): Temperature effects on gametophyte life-history traits and geographic distribution of two cryptic kelp species. − PLoS One 7: e39289.

Overton, J.B. (1913): Artificial parthenogenesis in *Fucus*. − Science 37: 841–844.

Pannell, J.R. & Labouche, A.M. (2013): The incidence and selection of multiple mating in plants. − Philos. Trans. R. Soc. Lond. B Biol. 368: 20120051.

Parker, G.A., Baker, R.R. & Smith, V.G. (1972): The origin and evolution of gamete dimorphism and the male-female phenomenon. − J. Theor. Biol. 36: 529–553.

Pearson, G.A. (2006): Revisiting synchronous gamete release by fucoid algae in the intertidal zone: fertilization success and beyond? − Integr. Comp. Biol. 46: 587–597.

Peters, A.F. (1992a): Culture studies on the life history of *Chordaria linearis* (Phaeophyceae) from Tierra del Fuego, South America. − J. Phycol. 28: 678–683.

Peters, A.F. (1992b): Culture studies on the life history of *Dictyosiphon hirsutus* (Dictyosiphonales, Phaeophyceae) from South America. − British Phycol. J. 27: 177–183.

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

Peters, A.F. & Müller, D.G. (1985): On the sexual reproduction of *Dictyosiphon foeniculaceus* (Phaeophyceae, Dictyosiphonales). − Helgoländer Meeresuntersuchungen 39: 441–447.

Peters, A.F. & Müller, D.G. (1986): Sexual reproduction of *Stilophora rhizodes* (Phaeophyceae, Chordariales) in culture. − British Phycol. J. 21: 417– 423.

Peters, A.F., Novaczek, I., Müller, D.G. & McLachlan, J.L. (1987): Culture studies on reproduction of *Sphaerotrichia divaricata* (Chordariales, Phaeophyceae). − Phycologia 26: 457–466.

Peters, A.F., Van Oppen, M.J.H., Wiencke, C., Stam, W.T. & Olsen, J.L. (1997): Phylogeny and historial ecology of the Desmarestiales (Phaeophyceae) support a southern hemisphere origin. − J. Phycol. 33: 294–309.

Phillips, J.A., Clayton, M.N., Maier, I., Boland, W. & Müller, D.G. (1990): Sexual reproduction in *Dictyota diemensis* (Dictyotales, Phaeophyta). − Phycologia 29: 367–379.

Ramírez, M.E., Müller, D.G. & Peters, A.F. (1986): Life history and taxonomy of two populations of ligulate *Desmarestia* (Phaeophyceae) from Chile. − J. Bot. 64: 2948–2954.

Rice, W.R. (1984): Sex chromosomes and the evolution of sexual dimorphism. − Evolution 38: 735–742.

Richards, A.J. (1986): Plant breeding systems. – George Allen & Unwin, London.

Sauvageau, C. (1915): Sur la sexualité hétérogamique d'une Laminaire (*Saccorhiza bulbosa*). − C.R. Acad. Sci. 161: 796–799.

Sauvageau, C. (1918): Recherches sur les Laminaires des côtes de France. − Mém. Acad. Sci. 56: 1–240.

Schreiber, E. (1932): Über die Entwicklungsgeschichte und die systematische Stellung der Desmarestiaceen. − Z. Bot. 25: 561–582.

Shan, T.F., Pang, S.J. & Gao, S.Q. (2013): Novel means for variety breeding and sporeling production in the brown seaweed *Undaria pinnatifida* (Phaeophyceae): Crossing female gametophytes from parthenosporophytes with male gametophyte clones. − Phycol. Res. 61: 154–161.

Manuscript received: 17 November 2013
Accepted: 12 March 2014

Silberfeld, T., Leigh, J.W., Verbruggen, H., Cruaud, C., De Reviers, B. & Rousseau, F. (2010): A multi-locus time-calibrated phylogeny of the brown algae (Heterokonta, Ochrophyta, Phaeophyceae): Investigating the evolutionary nature of the "brown algal crown radiation". − Mol. Phylogenet. Evol. 56: 659–674.

Stache-Crain, B., Müller, D.G. & Goff, L.J. (1997): Molecular systematics of *Ectocarpus* and *Kuckuckia* (Ectocarpales, Phaeophyceae) inferred from phylogenetic analysis of nuclear and plastid-encoded DNA sequences. − J. Phycol. 33: 152–168.

Strathmann, R.R. (1990): Why life histories evolve differently in the sea. − Am. Zool. 30: 197–207.

Thuret, G. (1854): Recherches sur la fécondation des Fucacées et les anthéridies des algues. − Ann. Sci. Nat. (Bot.) 4: 197–214.

Togashi, T., Bartelt, J.L., Yoshimura, J., Tainaka, K. & Cox, P.A. (2012): Evolutionary trajectories explain the diversified evolution of isogamy and anisogamy in marine green algae. − Proc. Natl. Acad. Sci.USA 109: 13692–13697.

Tronholm, A., Sansón, M., Afonso-Carrillo, J. & De Clerck, O. (2008): Distinctive morphological features, life-cycle phases and seasonal variations in subtropical populations of *Dictyota dichotoma* (Dictyotales, Phaeophyceae). − Bot. Mar. 51: 132–144.

van Den Hoek, C. & Flinterman, A. (1968): The life-history of *Sphacelaria furcigera* Kütz. (Phaeophyceae). − Blumea 16: 193–242.

van Den Hoek, C., Mann, D.G. & Jahns, H.M. (1995): Algae: An Introduction to Phycology. – Cambridge University Press, Cambridge.

Vernet, P. & Harper, J.L. (1980): The costs of sex in seaweeds. − Biol. J. Linn. Soc. 13: 129–138.

Wiencke, C. & Clayton, M.N. (1990): Sexual reproduction, life history, and early development in culture of the Antartic brown alga *Himantothallus grandifolius* (Desmarestiales, Phaeophyceae). − Phycologia 29: 9–18

Wyatt, R. (1982): Population ecology of bryophytes. − J. Hattori Bot. Lab. 52: 179–198.

Wyatt, R. & Anderson, L.E. (1984): Breeding systems in Bryophytes. – In: Dyer, A.F. & Duckett, J.G. (eds), The experimental Biology of Bryophytes. − Academic Press, London.

Yamagishi, Y. & Kogame, K. (1998): Female dominant population of *Colpomenia peregrina* (Scytosiphonales, Phaeophyceae). − Bot. Mar. 41: 217–222.

Yang, G.P., Sun, Y., Shi, Y.Y., Zhang, L., Guo, S.S., Li, B., Li, X.J., Li, Z.L., Cong, Y.Z., Zhao, Y.S. & Wang, W.Q. (2009): Construction and characterization of a tentative amplified fragment length polymorphism-simple sequence repeat linkage map of *Laminaria* (Laminariales, Phaeophyta). − J. Phycol. 45: 873–878.

Yasui, H. (1992): Chromosome numbers and a sex chromosome of *Laminaria yendoana* Miyabe (Phaeophyta). − Nippon Suisan Gakkaishi 58: 1385.

Zhang, Z., Hambuch, T.M. & Parsch, J. (2004): Molecular evolution of sex-biased genes in *Drosophila*. − Mol. Biol. Evol. 21: 2130–2139.

Annexe 2 : Pedigree d'une partie des souches d'*Ectocarpus*

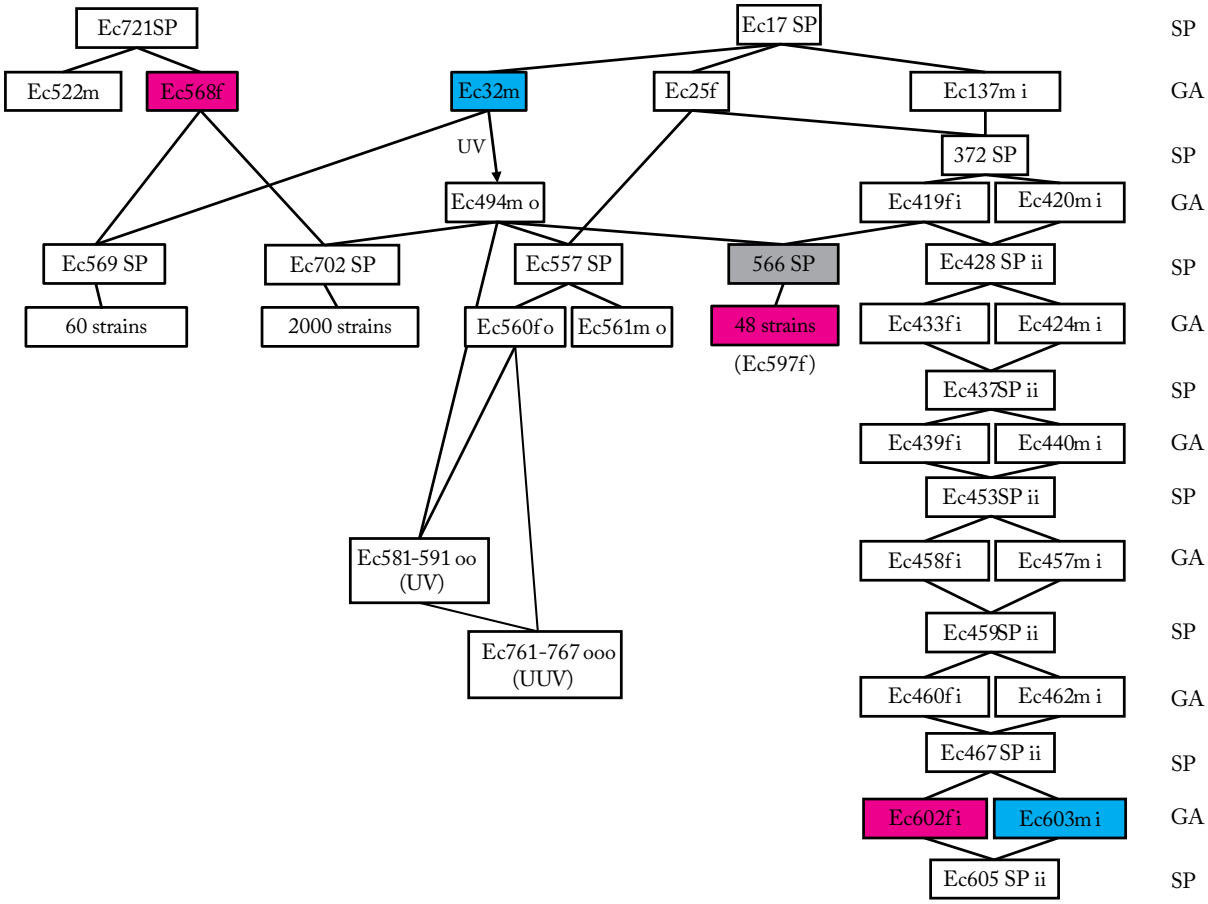


Figure S1 : Pedigree d'une partie des souches d'*Ectocarpus*. En bleu, les souches mâles et en rose, les souches femelles présentées dans ce manuscrit.

Annexe 3 : Liste des formations reçus

PERL : fundament, regular expression, references and BioPERL	Station Biologique de Roscoff	2,5 jours – 25 au 27 Juin 2013
European Course on Comparative Genomics	ENS Lyon	11 jours – 21 Janvier au 1 février 2013

Annexe 4 : Liste des formations données

Formation ABiMS - Initation cluster	Station Biologique de Roscoff	16 Juin 2015 - 6 heures
Licence 1 : Parcours Microbiologie et Sécurité des Aliments – Introduction à la Bioinformatique : cas pratique de la plateforme ABiMS.	Ecole Supérieure d'Ingénieurs en Agroalimentaire de Bretagne Atlantique, Brest	21 Janvier 2015 – 1 heure 30
Ecole de bioinformatique - Initiation au traitement des données de génomique obtenuespar séquençage à haut débit	Station Biologique de Roscoff	5 au 10 Octobre 2014 - ~ 20heures
Formation ABiMS – Galaxy RNA-seq avec référence	Station Biologique de Roscoff	12 Juin 2014 – 4 heures
Formation ABiMS - Initation cluster	Station Biologique de Roscoff	15 Mai 2014 – 6h
Formation ABiMS - Initation cluster	Station Biologique de Roscoff	8 Mars 2014 - 6h
Licence 1 : Parcours Microbiologie et Sécurité des Aliments – Introduction à la Bioinformatique : cas pratique de la plateforme ABiMS	Ecole Supérieure d'Ingénieurs en Agroalimentaire de Bretagne Atlantique, Brest	21 Janvier 2014 – 1 heure 30
Ecole de bioinformatique - Initiation au traitement des données de génomique obtenues par séquençage à haut débit	Station Biologique de Roscoff	17 au 23 Novembre 2013 - ~30 heures
Formation Licence LBM1 – Architecture des ordinateurs	Station Biologique de Roscoff	17 Octobre 2013 – 3 heures
Formation ABiMS – Galaxy RNA-seq avec référence	Station Biologique de Roscoff	19 Septembre 2013 – 4 heures
Formation ABiMS - Initation cluster	Station Biologique de Roscoff	16 Septembre 2013 – 4 heures
Formation ABiMS - Initation cluster	Station Biologique de Roscoff	9 Juillet 2013 – 4 heures
Ecole de bioinformatique - Initiation au traitement des données de génomique obtenues par séquençage à haut débit	Station Biologique de Roscoff	14 au 18 Janvier 2013 - ~10 heures
Formation Licence LBM1 – Architecture des ordinateurs	Station Biologique de Roscoff	20 Septembre 2012 – 3 heures

Annexe 5 : Liste des participations aux congrès

JOBIM 2015	Clermont-Ferrand	Communication orale	6 au 9 Juillet 2015
6 th annual meeting of the EFOR network	Paris	Communication orale	9 au 11 Mars 2015
Journées de la Société Phycologique de France	Roscoff	Poster	16 au 18 décembre 2013
JOBIM 2013	Toulouse	Poster	1 au 4 juillet 2013
JOBIM 2012	Rennes	Poster	3 au 6 juillet 2012
Esil 2012	Roscoff	Poster	23 au 25 avril 2012

Liste des figures