Etude des facteurs de transcription impliqués dans l'accumulation lipidique en condition de stress azoté chez la microalgue haptophyte Isochrysis affinis galbana

Stanislas Thiriet-Rupert

UNIVERSITE BRETAGNE LOIRE

Université du Maine

Thèse de Doctorat

Stanislas THIRIET-RUPERT

Mémoire présenté en vue de l'obtention du
grade de Docteur de l'Université du Maine
sous le sous le sceau de l'Université Bretagne Loire

École doctorale : *Végétal, Environnement, Nutrition, Agroalimentaire, Mer (VENAM)*

Discipline : *Biologie des organismes, Biochimie et biologie moléculaire*
Spécialité : *Biologie marine*
Unité de recherche *: Laboratoire PBA, Unité BRM, Centre IFREMER Atlantique*
Laboratoire Mer, Molécules, Santé (MMS), UFR Sciences et Techniques, Université du Maine

Soutenue le 10 janvier 2017
Thèse N° : 102823

Etude des facteurs de transcription impliqués dans l'accumulation lipidique en condition de stress azoté chez la microalgue haptophyte *Isochrysis* affinis *galbana*

JURY

Directeur de Thèse :	**Jean-Paul CADORET**, Directeur de recherche, Greensea
Co-directeur de Thèse :	**Benoît CHENAIS**, Professeur, Université du Maine - IUML – IFR 3473
Rapporteurs :	**Hans Johannes GEISELMANN**, Professeur, - UMR 5163 - Université Grenoble Alpes
	Jacqueline GRIMA-PETTENATI, Directeur de recherche HDR, CNRS – UMR5546 - Université Toulouse III Paul Sabatier
Examinateurs :	**Angela FALCIATORE**, Chercheur HDR, CNRS - UMR 7238 - Université Pierre et Marie Curie Paris 06
	Bruno SAINT-JEAN, Chercheur, IFREMER
	Benoît SCHOEFS, Professeur, Université du Maine - IUML – IFR 3473, Président du jury

Remerciements

Je tiens tout d'abord à remercier Jacqueline Grima-Pettenati et Hans Geiselmann d'avoir accepté d'être rapporteurs de cette thèse, et Angela Falciatore et Benoît Schoefs d'avoir accepté de l'examiner.

Je tiens, bien sûr, à remercier chaleureusement Jean-Paul Cadoret et Benoît Chénais, mes deux co-directeurs de thèse. Vous avez toujours été disponibles pour discuter tant du déroulement de la thèse que de l'après thèse ou de sujets divers. Merci pour vos conseils.

Un très grand MERCI à Bruno Saint-Jean qui m'a encadré durant ces trois années. La légende dit que tu dois ton légendaire brushing à mon encadrement… mais je préfère penser que tu les avais soigneusement conservés pour en faire cadeau à ta petite Violette ;-) Peut-être le petit Lombard (oui, j'insiste, le destin a parlé :-p) arborera-t-il fièrement la même chevelure bouclée.

Je dois également remercier Gregory Carrier, le maitre Jedi qui a initié le jeune padawan que je suis à la bio-informatique. Ce domaine que je ne connaissais que très vaguement il y a trois ans et que j'ai appris à apprécier avec toi, au point de vouloir orienter la deuxième partie de ma thèse vers la construction et l'analyse de réseaux de gènes. Si j'ai pu apprendre et appliquer autant de choses en trois ans, c'est en grande partie grâce à toi !

Si cette thèse c'est aussi bien déroulée, c'est en très grande partie grâce à l'excellente ambiance qui règne au sein du labo PBA ! Lorsqu'on travaille dans de telles conditions on ne peut qu'être heureux d'aller au labo le matin, et les petites baisses de régime ou de moral passagères s'envolent très vite grâce à vous ! Il suffit de s'assoir à son bureau et, déjà, l'ambiance vous imprègne. Une première vanne fuse… Puis une deuxième. Rapidement, un sourire chasse les dernières brumes du réveil et la journée de travail est lancée !

Merci à Loïc, l'homme des îles à la verve acérée ! Ton prochain château de sable en Nouvelle Calédonie aura peut-être la forme d'un pot de rillettes. Aurélie qui, bien que travaillant à ifremer, n'a pas dû manger assez de poisson pour entretenir sa mémoire ! Aurélie qui, bien que travaillant à ifremer, n'a pas dû manger assez de poisson pour entretenir sa mémoire ! (mince, je t'avais déjà citée, 3 ans à te côtoyer et ma mémoire en ressent déjà les effets :-p). Ewa, grâce à toi je sais maintenant qu'il y a une vie avant le bigos… pardon, avant Bigos, et une vie après Bigos ! Catherine, la maman du labo, toujours souriante, dynamique et bienveillante. JB, la preuve

vivante que les métaleux peuvent, eux aussi, être fréquentables (n'en déplaise à ton amie Christine…). Matthieu, un thésard pas comme les autres et roi de la fameuse garnierade ! Nathalie et Raymond, le MacGyver lorrain (…ou alsacien…), mes compatriotes thermophobes. Gaël, le gardien bienveillant de ce troupeau de brebis égarées. Elodie qui voulait m'enchainer à la paillasse. Et bien non mademoiselle, l'esclavage est aboli en France ! Et Isabelle, la vraie chef, tes rêves de soutane et de blanche neige seront peut-être un jour exaucés ;-)

Merci également à tous les fameux « accessoires » du labo et des alentours. Camille, Sonia, Mathilde, Aurel, Elodie et la fine équipe des thésards Damien, Zita, Xavier, Cécile, Adèle, Francesco, Nour, Jerem, Audrey et Laetitia. Un gros bisou tout particulier à Judith, amie thésarde qui m'a supporté presque trois ans dans le même bureau (même si l'inverse est plus proche de la réalité :-p). Une amie comme on en rencontre assez peu. Evoquer nos diverses difficultés et anecdotes autour d' « un verre » a toujours été un très bon moyen de se détendre, d'autant plus en aussi bonne compagnie !!

Merci à Nathalie, Aurore, Justine, Emmanuelle et Hélène que j'ai eu un grand plaisir à revoir à chacun de mes passage au Mans. Et ceux de toujours, Mathieu, Marie, Quentin, Antoine, Sylvain, Michel, Inès, Max, Jerem et Cedo, toujours là pour une bonne dose de convivialité, que ce soit au Mans, à Nantes ou aux vieilles charrues ! Sans oublier ma Juliette, mon ange gardien. Merci pour ta joie de vivre, ton soutien de chaque instant, les légumes, nos plantes, ta maitrise de l'orthographe, et tellement plus !! Je pense que ton CDD va être renouvelé !

Enfin, un grand merci à la cafetière ! Celle-ci ayant décidé qu'elle serait vide à chaque fois que l'envie d'un café se faisait sentir, refaire du café a sans doute été la manip que j'ai le plus réalisée pendant cette thèse !!

Si cette thèse m'a beaucoup (beaucoup) apporté scientifiquement, elle m'a encore plus apporté humainement. On s'épanouit à travers les personnes qu'on rencontre, avec qui on échange et qu'on apprend à connaitre et à apprécier. Passer ces trois années avec vous a été un vrai bonheur et j'espère qu'on restera en contact.

Table des matières

Liste des publications et communications

Publication publiée concernant le sujet de thèse

Thiriet-Rupert S, Carrier G, Chénais B, Trottier C, Bougaran G, Cadoret J-P, Schoefs B, Saint-Jean B. 2016. Transcription factors in microalgae: genome-wide prediction and comparative analysis. *BMC genomics* **17**: 282.

Publication en préparation concernant le sujet de thèse

Thiriet-Rupert S, Carrier G, Schoefs B, Trottier C, Bougaran G, Cadoret J-P, Chénais B, Saint-Jean B. Transcription factors involved in the phenotype of a domesticated oleaginous microalgae strain of *Tisochrysis lutea*. En preparation.

Communications effectuées lors d'un congrès national ou international durant la thèse

- Transcription factors involved in the phenotype of a domesticated oleaginous microalgae strain of *Tisochrysis lutea*. 2016. *Journées de la Société Phycologique de France*. 07-08 décembre 2016, Banyuls-sur-mer. Présentation orale 15 minutes

- Reconstruction of gene network and regulatory network of an improved microalgae strain. 2016. *aDVANCES IN SYSTEMS AND SYNTHETIC BIOLOGY: Modelling Complex Biological Systems in the Context of Genomics, Thematic Research School*. March 21st to 25th 2016, Évry. Poster

- Genome-wide prediction and comparative analysis of transcription factors in microalgae. 2015. *Journées de la Société Phycologique de France*. 24-25 septembre 2015, Vannes. Présentation orale 15 minutes

- Genome-wide prediction and comparative analysis of transcription factors in microalgae. 2015. *EPC6 - 6th European Phycological Congress*. 23-28 August 2015, London. Présentation orale de 15 minutes

- Study of transcription factors involved in lipid accumulation-induced by nitrogen starvation in the microalgae *Tisochrysis lutea*. 2013. *Alg'n'Chem 2014* March 31 - April 3 2014 Montpellier. Poster.

- Study of transcription factors involved in lipid accumulation-induced by nitrogen starvation in the microalgae *Tisochrysis lutea*. 2013. *Young Algaeneers Symposium (YAS) 2014* April 3 - April 5 2014 Montpellier/Narbonne. Poster.

Autre communication effectuée durant la thèse

- Comprendre la production de lipides chez la microalgue *Tisochrysis lutea*. 2015. *13ème Forum Jeune-Recherche* 2015, Université du Maine, Le Mans. Poster

Liste des Figures

Liste des Tableaux

Introduction générale

I. Le phytoplancton

Le phytoplancton représente les microorganismes photosynthétiques aquatiques : les cyanobactéries (procaryotes), et les microalgues (eucaryotes). Bien que ne représentant que 1% de la biomasse photosynthétique sur Terre, le phytoplancton joue un rôle clé dans le fonctionnement des écosystèmes aquatiques. Il assure notamment 45% de la production primaire de la planète et plus de 90% de la production primaire des milieux aquatiques (Geider *et al.*, 2001). Le phytoplancton est à la base de la chaine alimentaire du fait de sa fonction de producteur primaire et son caractère autotrophe lui permet également d'assurer la fixation de plus de 10 milliards de tonnes de carbone atmosphérique par an. La matière organique et les squelettes carbonatés ainsi formés migrent ensuite par sédimentation vers le fond des océans, formant un puits de carbone primordial dans la séquestration du CO_2 atmosphérique (Bowler *et al.*, 2009). Le phytoplancton joue également un rôle important dans le cycle biogéochimique d'autres éléments clés tels l'azote ou le phosphore (Falkowski *et al.*, 2004). Les communautés phytoplanctoniques océaniques sont principalement dominées par des cyanobactéries de deux genres, *Prochlorococcus* et *Synechococcus*. Concernant les microalgues, le nombre d'espèces de ces organismes sur la planète s'élèverait à plusieurs milliers, voire millions (Andersen, 1992; Guiry, 2012). Une étude récente estime plus précisément ce nombre à 150 000 unité taxonomique opérationnelle (de Vargas *et al.*, 2015). De plus, cette diversité se traduit également par la répartition des espèces de microalgues identifiées à ce jour dans de nombreux groupes taxonomiques des eucaryotes (Not *et al.*, 2012) (Figure 1). Parmi ces nombreux taxons, les diatomées sont considérées comme les plus abondantes dans le milieu océanique, principalement au niveau du littoral, et seraient responsables de 20% de la fixation de carbone par photosynthèse à l'échelle du globe. Les dinoflagellées et les haptophytes seraient, ensuite, les plus représentés (Field *et al.*, 1998). Toutefois, l'abondance et le rôle écologique des haptophytes seraient fortement sous-estimés du fait du nombre d'espèces dont la taille, inférieure à 3 µm, rend difficile l'identification (Liu *et al.*, 2009). Les haptophytes seraient donc un groupe majeur de l'écologie des milieux océaniques.

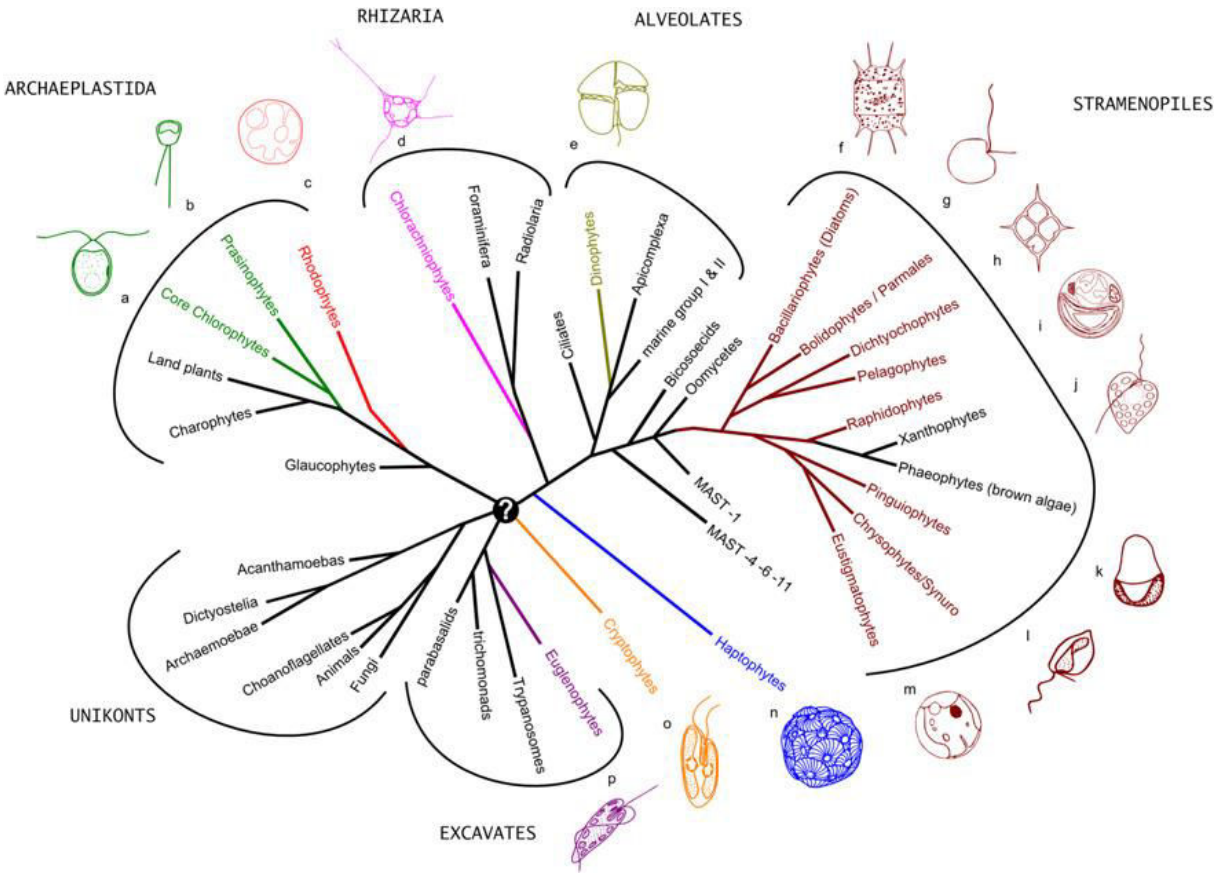

Figure 1: représentation schématique de la répartition des taxons phytoplanctoniques dans l'arbre phylogénétique des eucaryotes (Not *et al.*, 2012).

II. Les haptophytes

Les haptophytes sont des organismes unicellulaires eucaryotes majoritairement marins, photosynthétiques pour la plupart bien que certaines espèces soient mixotrophes (Tillmann, 1998). Leur taille est généralement comprise entre 2 et 30 µm, mais certaines études environnementales ont rapporté l'existence de nombreuses espèces de pico-haptophytes inférieurs à 2 µM (Moon-van der Staay *et al.*, 2000 ; Liu *et al.*, 2009). Les membres de ce groupe monophylétique très ancien (Figure 2) auraient divergés il y a approximativement un milliard d'années (Medlin *et al.*, 2008). La classification phylogénique des haptophytes a d'abord été fondée sur la présence d'un haptonème. Cet appendice impliqué dans l'adhésion au substrat, le déplacement de particules, ou la capture de proies est situé entre leur deux flagelles. La taille de ces deux flagelles distingue les deux grands groupes d'haptophytes : les Pavlovophyceae et les Coccolithophyceae. Au sein de ce dernier, se trouvent le groupe des coccolithophores, les plus connus des haptophytes, auquel appartient l'espèce modèle de ce taxon : *Emiliania huxleyi*. Celle-ci est, de fait, la plus largement étudiée et présente à sa surface des écailles de calcite appelées coccolithes caractéristiques des coccolithophores. La forme et la présence de ces coccolithes dépend des espèces et de leur cycle de vie (Brownlee *et al.*, 2015) (Figure 3). Ces organismes étant très anciens, ils ont rapidement pris part au cycle biogéochimique du carbone et leur sédimentation est aujourd'hui encore retrouvée dans le relief de nos côtes, les falaises d'Etretat en Normandie en étant un très bel exemple.

Certaines espèces d'haptophytes ont également la particularité de former des efflorescences phytoplanctoniques appelées « bloom » (Figure 4). Ces « blooms » correspondent à une très forte augmentation de la concentration d'une ou plusieurs espèces d'algues lorsque les conditions environnementales sont favorables (disponibilité en nutriments, oxygène et lumière). L'haptophyte modèle *Emiliania huxleyi* peut, par exemple, former des efflorescences pouvant dépasser les 100 000 km² de surface et atteindre des concentrations cellulaires supérieures à 10 000 cellules / mL. Leur rôle écologique est également important, puisque les coccolithophores participent au cycle du carbone et du calcium au travers de la formation de coccolithes. Ceux-ci étant formés de carbonate de calcium, la part imputée aux haptophytes dans la sédimentation de ce composé est estimée à environ 50% (Milliman, 1993; Shiraiwa, 2003).

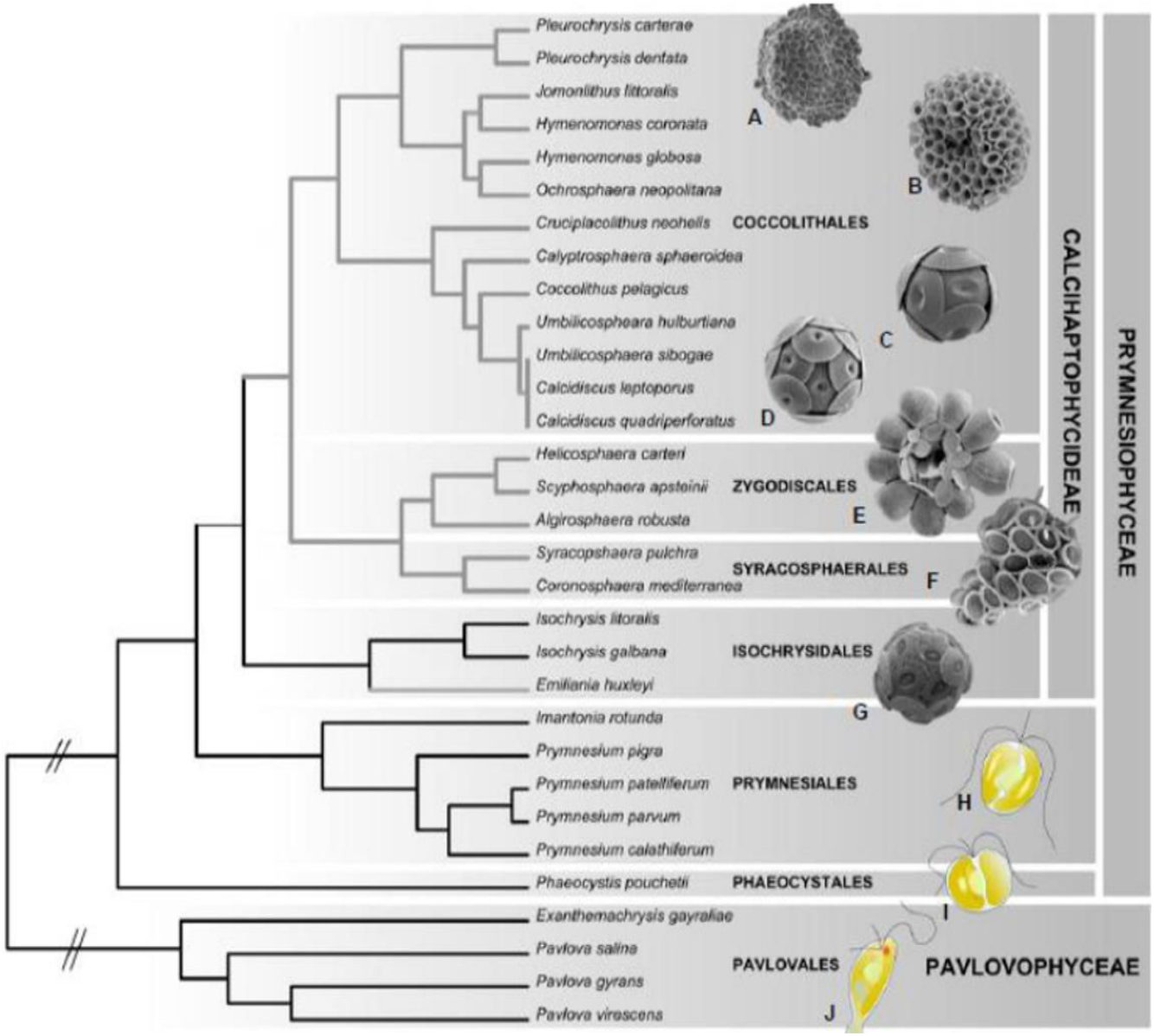

Figure 2 : arbre phylogénétique des haptophytes construit à partir de leurs séquences 18S (de Vargas *et al.*, 2007).

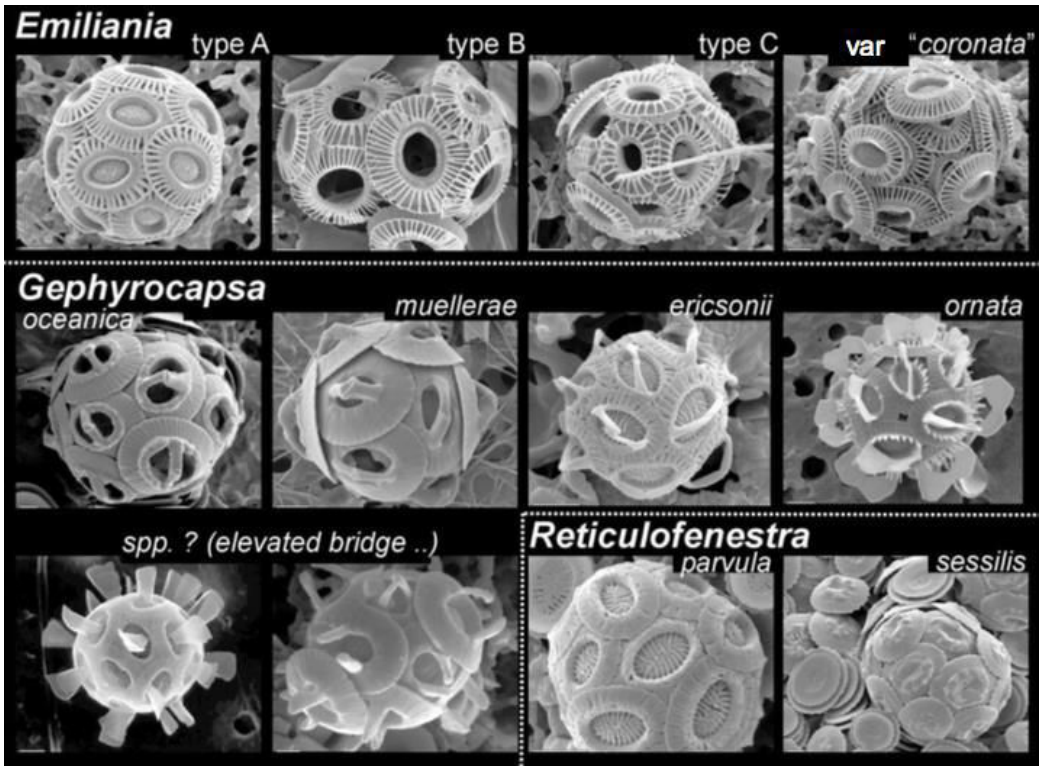

Figure 3 : exemples de la diversité morphologique des haptophytes (Young *et al.*, 2003).

Figure 4 : photo satellite représentant un « bloom » de l'haptophyte coccolitophore *Emiliania huxleyi*. Source : S. Groom, Plymouth Marine Laboratory, U.K., sbg@pml.ac.uk "Genomics:GTL Roadmap,"

De plus, la part de la production primaire océanique imputée aux haptophytes serait fortement sous-estimée (Liu *et al.*, 2009 ; Jardillier *et al.*, 2010). Les haptophytes sont les seuls organismes à produire un pigment photosynthétique particulier (le 19'-hexanoyl-oxy-fucoxanthine), lequel est omniprésent dans l'eau de mer. Un bilan quantitatif de l'importance de ce pigment à l'échelle du globe au cours de l'année 2000 montre que les haptophytes pourraient contribuer pour 30 à 50% de la production primaire océanique, et que leur biomasse serait jusqu'à deux fois plus importante que celle des diatomées ou des cyanobactéries (Figure 5) (Liu *et al.*, 2009). D'autre part, leur rôle dans le cycle du soufre serait également non négligeable, notamment via l'émission de diméthylsulfure (DMS), l'un des principaux composés organiques volatiles à la surface des océans (Li *et al.*, 2010).

Enfin, certains haptophytes (appartenant à l'ordre des isochrysidales) ont la particularité de produire des lipides neutres à très longue chaîne (C37-C39) appelés alkénones qui contribuent au stockage du carbone sédimentaire (Eltgroth et al 2005). De plus, la température de l'eau influe sur la production de ces lipides. En effet, des températures élevées favorisent la production d'alkénones di-insaturés par rapport aux alkénones tri-insaturés (Prahl & Wakeham, 1987). Du fait de cette caractéristique, les alkénones sont utilisés en paléoclimatologie comme biomarqueurs (Eglinton & Eglinton, 2008).

Malgré ce rôle primordial dans l'écologie des écosystèmes marins, la physiologie et le métabolisme de ces organismes restent encore peu connus. Les connaissances acquises à ce jour concernent essentiellement *E. huxleyi* qui est la première espèce haptophyte pour laquelle des données transcriptomiques ont été générées (von Dassow *et al.*, 2009) et dont le génome séquencé est disponible (Read *et al.*, 2013). Très récemment, un deuxième génome d'haptophyte a été publié : celui de la microalgue non coccolithophore *Chrysochromulina tobin* qui appartient à l'ordre de prymnesiales (Hovde *et al.*, 2015). Enfin, le génome d'une troisième haptophyte, *Tisochrysis lutea*, est disponible au laboratoire et va prochainement être rendu public (Carrier *et al.*, *en préparation*). Les différents taxons de microalgues étant très éloignés les uns des autres, peu d'informations peuvent être obtenues par homologie (que ce soit de séquences, de comportements ou de mécanismes). Cette grande diversité étant due à une histoire évolutive très complexe que les études phylogénétiques et phylogénomiques peinent à élucider, notamment à cause du faible nombre de génomes séquencés chez certains taxons comme les haptophytes.

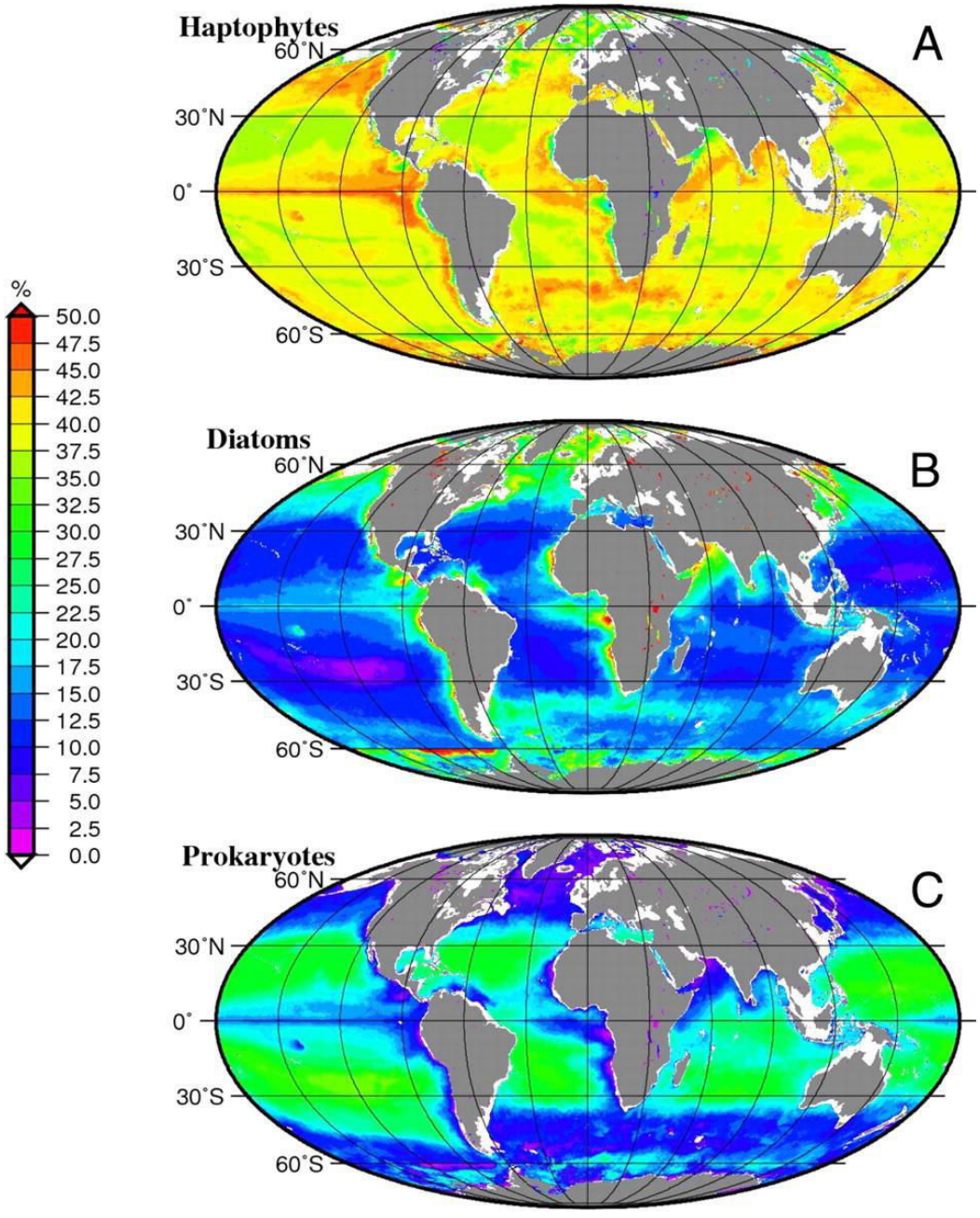

Figure 5 : contribution relative des (A) haptophytes, (B) diatomées, et (C) procaryotes photosynthétiques à la production primaire au cours de l'année 2000 (Liu *et al.*, 2009).

III. L'histoire évolutive des microalgues

L'histoire débute il y a environ 1,8 milliards d'années, lorsqu'un procaryote photo-autotrophe, proche des cyanobactéries actuelles, est phagocytée par une cellule hétérotrophe primitive. Suite à cette absorption, le procaryote photosynthétique n'a pas été entièrement digéré mais domestiqué par la cellule hétérotrophe via des phénomènes de perte et d'échange de gènes, de remaniements structuraux et de communication avec la cellule hôte, donnant naissance à un organite intracellulaire spécialisé dans la photosynthèse : le chloroplaste. Cet hypothèse d'endosymbiose communément admise abouti à l'émergence de trois lignées de microalgues appelées archaeplastidae (Delwiche, 1999) : les glaucophytes, les rhodophytes (microalgues rouges), et les chlorophytes (microalgues vertes) à partir desquelles les plantes terrestres auraient divergées il y a environ 900 millions d'années (Hedges *et al.*, 2004). La grande diversité de microalgues observée à ce jour aurait ensuite émergée d'une série d'endosymbioses secondaires (endosymbiose de microalgue vertes et rouges par un hétérotrophe) (Figure 6), tertiaires et quaternaires. Cependant, l'ordre et le nombre de ces endosymbioses fait toujours débat (Keeling, 2013).

L'hypothèse endosymbiotique a été proposée par Lynn Margulis (Sagan, 1967) suite à l'observation des multiples membranes entourant les chloroplastes. En effet, bien que les chloroplastes issus d'une endosymbiose primaire ne soient entourés que par les deux membranes de la cyanobactérie GRAM négative (la membrane phagosomale ayant été perdue lors du processus de domestication) (Cavalier-Smith, 1982), les chloroplastes issus d'une endosymbiose secondaire sont, eux, entourés par quatre membranes : les deux membranes de la cyanobactérie, la membrane plasmique de l'algue phagocytée et la membrane phagosomale (McFadden, 1999) (Figure 7). Cependant chez les euglènes et les dinophlagellées issues d'une endosymbiose secondaire, la membrane plasmique de l'algue phagocytée aurait disparue, ne laissant que trois membranes autour de leur chloroplaste (Sulli *et al.*, 1999 ; Nassoury *et al.*, 2003). Suite à cette hypothèse, la biologie évolutive a fait un bond (bien qu'il ait fallu plus de 10 ans avant que la théorie de Lynn Margulis ne soit vraiment acceptée par la communauté scientifique). La première théorie concernant l'histoire évolutive des microalgues élaborée par Cavalier-Smith (1999) est fondée sur leur composition pigmentaire. Ainsi, les Chlorarachniophytes et les Euglènes seraient issues de l'unique endosymbiose secondaire d'une microalgue verte suivie par une divergence de

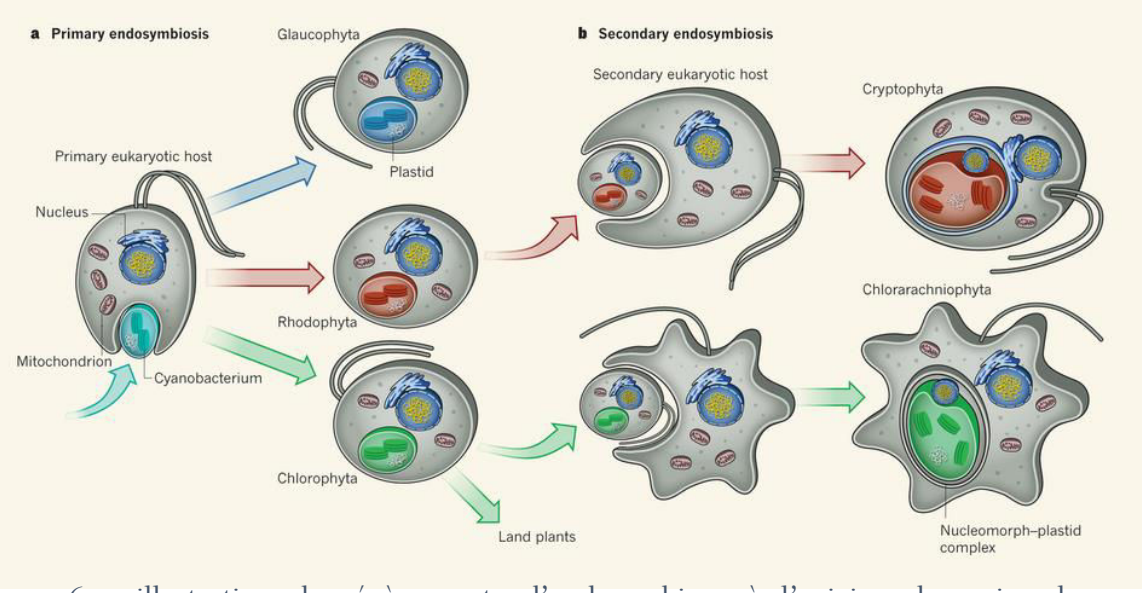

Figure 6 : illustration des évènements d'endosymbiose à l'origine des microalgues. a) endosymbiose primaire d'une cyanobactérie primitive par un hétérotrophe primitif. Trois lignées d'algues ont émergées d'une endosymbiose primaire : le glaucophytes, les rhodophytes et les chlorophytes. b) deux exemples d'endosymbioses secondaires. Les cryptophytes, résultant de l'endosymbiose secondaire d'une microalgue rouge par un hétérotrophe. Les chlorarachniophytes, résultant de l'endosymbiose secondaire d'une microalgue verte. (Gould, 2012).

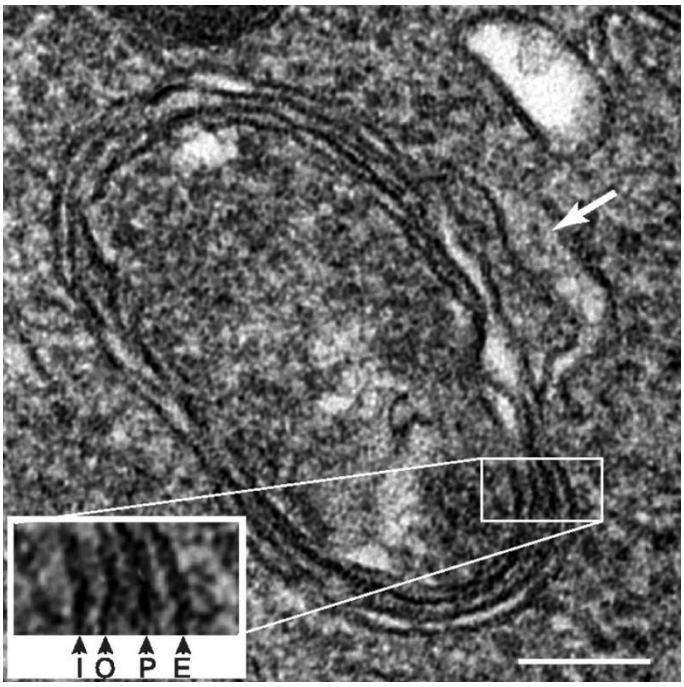

Figure 7 : observation en microscopie électronique à transmission d'un plaste d'apicomplexe (*Toxoplasma gondii*) issu d'une endosymbiose secondaire. Les quatre membranes sont bien visibles : I, membrane interne du plaste d'origine cyanobactérienne; O, membrane externe du plaste d'origine cyanobactérienne; P, membrane plasmique de l'algue internalisée lors de l'endosymbiose; et E, membrane la plus externe ayant pour origine le système endomembranaire de l'apicomplexe (Parsons *et al.*, 2007).

ces deux lignées. Cinq autres lignées seraient elles issues de l'unique endosymbiose secondaire d'une microalgue rouge suivie par la divergence de chacune d'elles : Les cryptophytes, les haptophytes, les straménopiles et les alvéolés (regroupant les dinoflagellées et le cas particulier des apicomplexes, des parasites intracellulaires obligatoires dont certaines espèces comportent un chloroplaste entouré par quatre membranes (Lang-Unnasch *et al.*, 1998) (Figure 8). Ces cinq lignées formant un super-groupe monophylétique, puisque issues d'une même endosymbiose secondaire, le super-groupe des chromalvéolés (Cavalier-Smith, 1999). Toutefois, les straménopiles et les alvéolés regroupent des organismes variés photosynthétiques et non photosynthétiques tels que les apicomplexes, les dinoflagellés et les ciliés. La présence de ces organismes hétérotrophes au sein des chromalvéolés est surprenante puisque l'ensemble de ce super-groupe serait issu d'un évènement d'endosymbiose secondaire. L'hypothèse avancée fut que le chloroplaste, acquis par l'ancêtre commun des chromalvéolés, ait été secondairement perdu chez les chromalvéolés à ce jour hétérotrophes. Cette hypothèse fut soutenue par la présence de gènes d'endosymbiontes chez les champignons straménopiles (Tyler *et al.*, 2006), les apicomplexes (Huang *et al.*, 2004), les dinoflagellée non photosynthétique (Sanchez-Puerta *et al.*, 2007) ou encore chez les ciliés (Reyes-Prieto *et al.*, 2008). Ce super-groupe des chromalvéolés apparait donc très vaste et diversifié, et comporte finalement peu d'organismes photosynthétiques (Adl *et al.*, 2005). Par la suite, de nombreuses études génomiques ont cherché à préciser cette hypothèse (Burki, 2014).

Le fait que plusieurs évènements d'endosymbiose secondaires aient eu lieu dans l'histoire évolutive des algues n'est pas remis en cause, en revanche le nombre de ces évènements fait encore débat. Concernant les deux lignées issues de l'endosymbiose secondaire d'une microalgue verte (euglènes et chlorarachniophytes), l'indépendance de ces deux évènements est maintenant établie par des études phylogénétiques tant à l'échelle de l'endosymbionte (le génome chloroplastique) (Rogers *et al.*, 2007) qu'à celle de l'hôte (le génome nucléaire) (Kumazaki *et al.*, 1983 ; Bhattacharya *et al.*, 1995 ; Keeling, 2001). De plus, le chloroplaste des euglènes est entouré de trois membranes alors que celui des chlorarachniophytes en comporte quatre. Enfin, contrairement aux euglènes, suite à l'évènement d'endosymbiose secondaire, les chlorarachriophytes ont conservé un reliquat du noyau de la microalgue verte phagocytée appelé nucléomorphe (Gilson *et al.*, 2006 ; Suzuki *et al.*, 2015). Ce génome ultra-réduit est constitué de trois chromosomes et est situé entre la deuxième et la troisième membrane du chloroplaste.

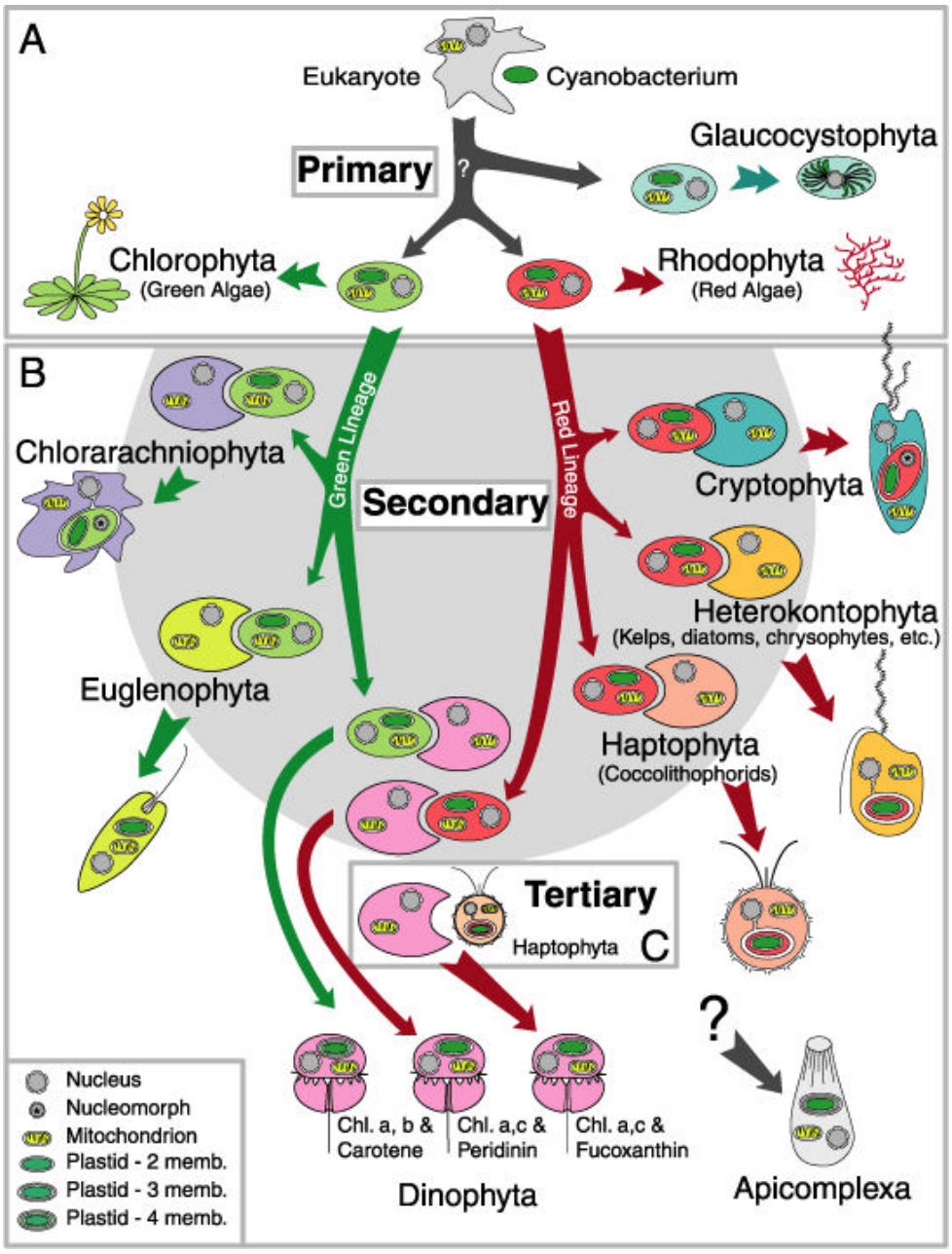

Figure 8 : représentation schématique de l'histoire évolutive des microalgues selon l'hypothèse d'une origine commune des chromalvéolés. Les endosymbioses primaires à l'origine des archaeplastidae (A), puis les endosymbioses secondaires (B) et tertiaires (C). Delwich, 1999.

Autant d'arguments plaidant en faveur de l'émergence de ces deux lignées suite à deux endosymbioses secondaires différentes d'une microalgue verte. Concernant les lignées issues de l'endosymbiose secondaire d'une microalgue rouge, la reconstitution de leur histoire évolutive est plus complexe. Cela est dû notamment au fait que ces lignées sont plus nombreuses et que la phylogénie des hôtes impliqués n'est pas encore élucidée. Toutefois, la monophylie de l'ensemble des phylums formant le super-groupe des chromalvéolés reste très controversée. Les études phylogénétiques menées sur l'ensemble du génome chloroplastique ont regroupé les cryptophytes, les straménopiles et les haptophytes (Hagopian *et al.*, 2004). Néanmoins, la structure très particulière du génome chloroplastique des dinoflagellées et des apicomplexes a longtemps rendu impossible l'incorporation de ces deux lignées aux arbres phylogénétiques générés. En effet, Les apicomplexes étant des parasites obligatoires, leur génome chloroplastique est très réduit et ne comporte pas de gène lié à la photosynthèse (Wilson *et al.*, 1996). Quant aux dinoflagellées, la plupart des gènes de leur génome chloroplastique ont été transférés vers le génome nucléaire et la séquence de chaque gène chloroplastique restant a été circularisée (Zhang *et al.*, 1999 ; Green, 2004).

Cependant, l'identification de deux nouvelles espèces d'algues proches des apicomplexes et dont le génome chloroplastique est moins réduit a permis de compléter ces études (Moore *et al.*, 2008 ; Oborník *et al.*, 2012). La proximité phylogénique des chloroplastes des alvéolés (dont font partie les apicomplexes) et des straménopiles a ainsi été mise en évidence ; toutefois la robustesse du super-groupe des chromalvéolés n'était pas concluante (Janouškovec *et al.*, 2010).

De même, les études phylogénétiques établies à partir de gènes de l'hôte (nucléaires) n'ont confirmé que la monophylie des straménopiles et des alvéolés (dinoflagellées et apicomplexes) (Van de Peer *et al.*, 1996 ; Harper *et al.*, 2005 ; Elias *et al.*, 2009). Des études plus larges de phylogénomique ont ensuite permis d'utiliser de plus grands jeux de données et ont mis en évidence le fort lien entre alvéolés et straménopiles d'une part, et le groupe des rhizaria (un très large groupe de protistes planctoniques) d'autre part, formant le clade SAR (Straménopiles, Alvéolés, Rhizaria) (Figure 10) (Hackett *et al.*, 2007 ; Burki *et al.*, 2008, 2010). Mais encore une fois, la position des haptophytes et des cryptophytes, autant l'un par rapport à l'autre que par rapport au clade SAR, n'est pas bien définie (Baurain *et al.*, 2010 ; Burki *et al.*, 2012), remettant en cause la cohérence du super-groupe des chromalvéolés (Figure 9).

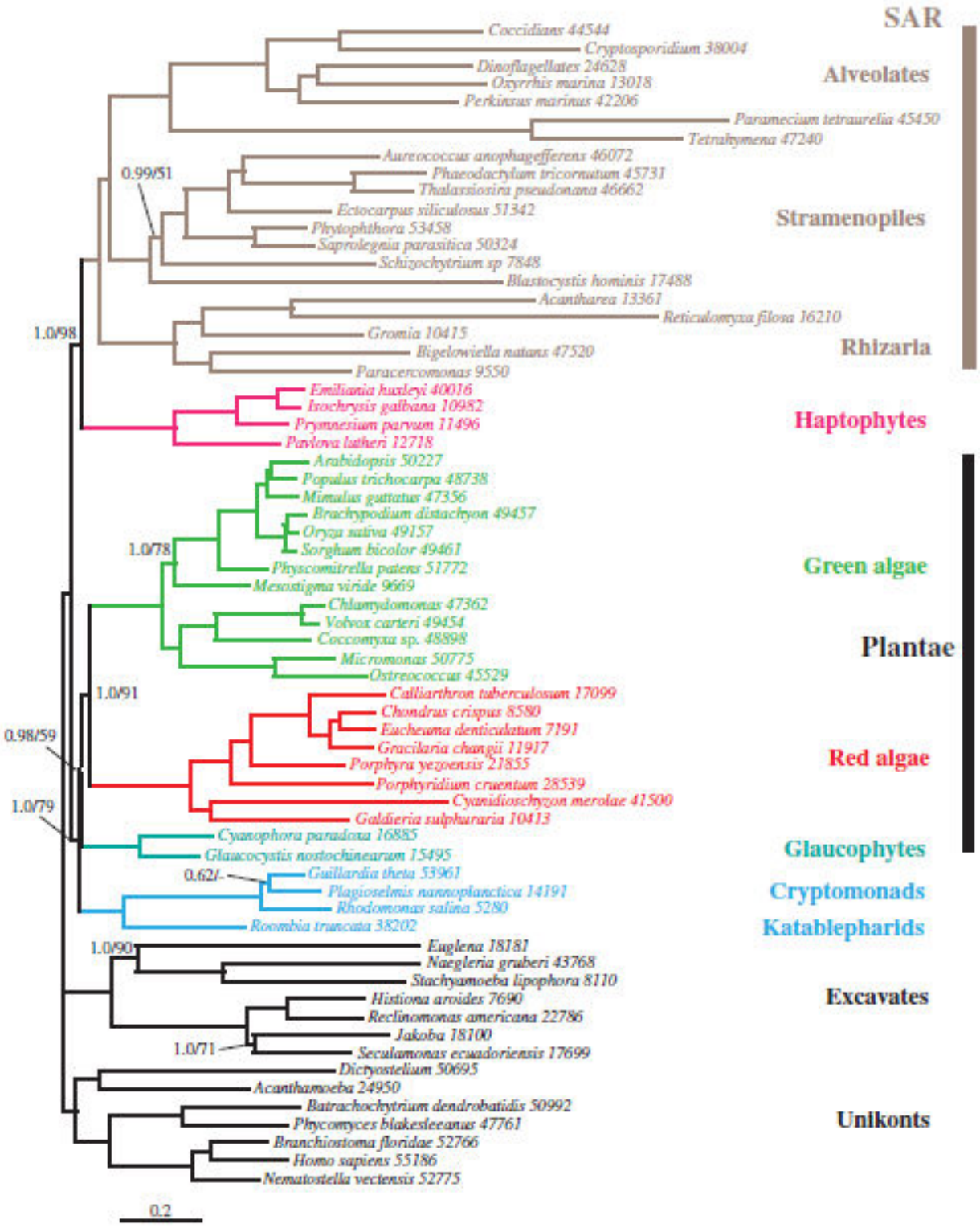

Figure 9 : arbre inféré par l'étude phylogénomique de Burki et al (2012) et représentant les différents groupes des eucaryotes. Les haptophytes sont bien embranchées au groupe SAR (Straméniopiles, Alvéolés, Rhizaria). Mais les cryptophytes (cryptomonads) sont embranchés aux archaeplastidiae (Plantae), remettant en cause la monophylie des chromalvéolés. (Burki *et al.*, 2012)

Il ressort donc, à la lumière d'études sur leur génome chloroplastique, que les lignées photosynthétiques réunies dans ce super-groupe des chromalvéolés partageraient bien le même endosymbionte. En revanche, la phylogénie des hôtes montre qu'ils sont, eux, plus éloignés phylogénétiquement. Ceci impliquerait l'acquisition d'un endosymbionte de la même espèce par différents hôtes, lesquels ayant ensuite donné naissance aux différentes lignées photosynthétiques liés à l'endosymbiose d'une microalgue rouge (Keeling, 2013) (Figure 10). Cette chronologie va à l'encontre de l'hypothèse des chromalvéolés, selon laquelle ces lignées seraient issues d'une unique endosymbiose secondaire. De plus, cette hypothèse est également fondée sur de nombreux évènements de perte de chloroplastes, notamment au sein des alvéolés. Or l'hypothèse que ces évènements puissent être si fréquents est, elle aussi, très controversée (Stiller *et al.*, 2014). En conséquence, les détracteurs de l'hypothèse des chromalvéolés font référence à ce super-groupe en parlant « des lignées CASH » (Cryptophytes, Alvéolés, Straménopiles, Haptophytes) afin de ne pas faire référence aux chromalvéolés en tant que groupe taxonomique avéré (Baurain *et al.*, 2010). Depuis, différentes études se sont attelées à l'élaboration d'hypothèses expliquant cette acquisition indépendante d'un même endosymbionte issu d'une microalgue rouge chez les lignées CASH.

Ces études avancent qu'une des lignées serait apparue par endosymbiose secondaire d'une microalgue rouge, et que ce chloroplaste se serait ensuite répandu dans les autres lignées par des évènements d'endosymbiose tertiaire voir quaternaire. Ce qui serait bien en accord avec la proximité phylogénétique de ces lignées du point de vue de l'endosymbionte (génome chloroplastique), contrastée par leur éloignement du point de vue de l'hôte (génome nucléaire). Cette hypothèse appelée « l'hypothèse rhodoplexe » (Petersen *et al.*, 2014), a, depuis, été affinée (Stiller *et al.*, 2014 ; Ševč íková *et al.*, 2015). Selon ces derniers travaux, la lignée des cryptophytes serait apparue la première par endosymbiose secondaire, puisque ces organismes ont conservé un nucléomorphe, noyau vestigial de l'algue rouge domestiquée (Archibald, 2007). Une endosymbiose tertiaire d'une cryptophyte serait ensuite à l'origine des ochrophytes (les straménopiles photosynthétiques). Et deux évènements d'endosymbioses quaternaires d'une ochrophyte serait à l'origine des haptophytes d'une part (Stiller *et al.*, 2014) (Figure 11), et des alvéolés photosynthétiques d'autre part (Ševčíková *et al.*, 2015).

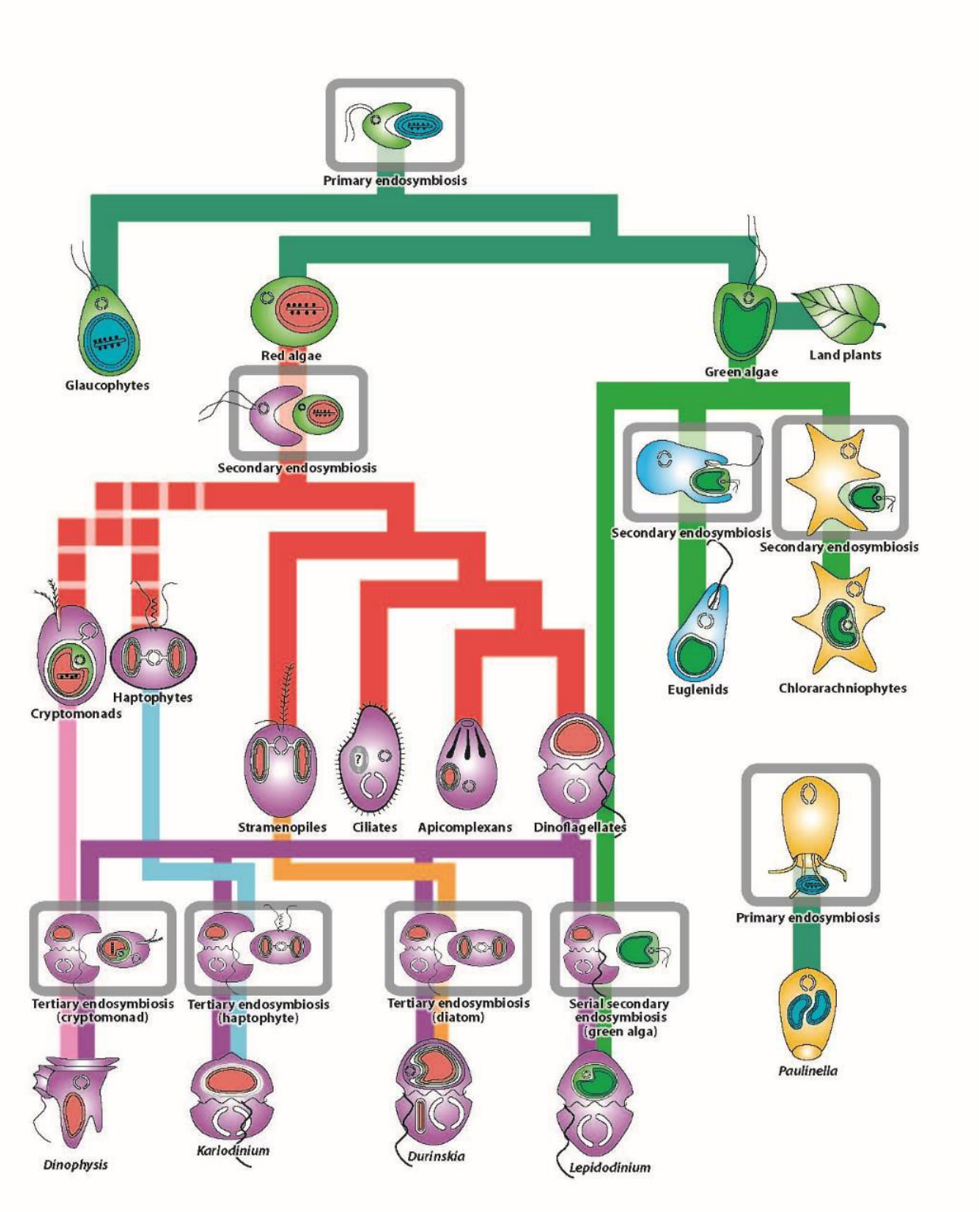

Figure 10 : représentation schématique des différents évènements d'endosymbiose retraçant l'histoire évolutive des algues (Keeling, 2013).

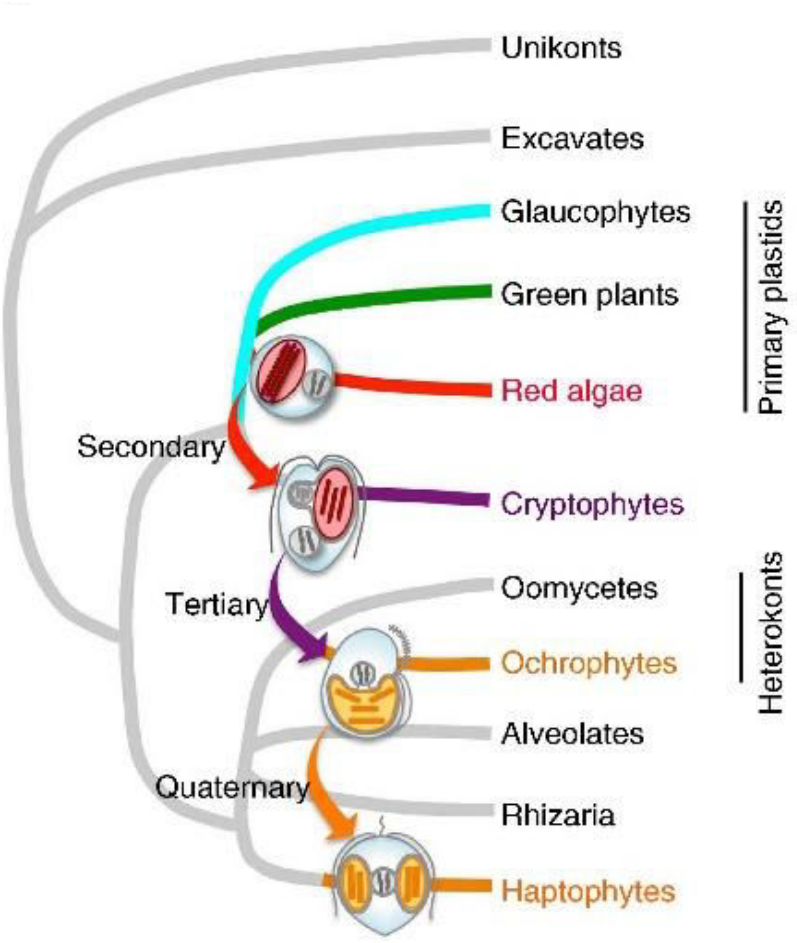

Figure 11 : représentation schématique des évènements d'endosymbioses en série à l'origine, selon l'hypothèse de Stiller et al (2014), de la répartition des chloroplastes originaires d'une microalgue rouge chez les lignées CASH. Les Cryptophytes seraient apparues par l'endosymbiose secondaire d'une microalgue rouge. Par la suite, les Ochrophytes seraient apparues par l'endosymbiose tertiaire d'une Cryptophyte, et les Haptophytes par l'endosymbiose quaternaire d'une Ochrophyte. (Stiller *et al.*, 2014)

De plus, ce schéma est encore plus compliqué par les évènements de transferts de gènes horizontaux qui ont jalonnés cette histoire déjà complexe (Figure 12, Chan & Bhattacharya, 2010). Le transfert de matériel génétique entre organismes appartenant à des taxons différents, ou transfert horizontal, participe à l'évolution des génomes tant procaryotes qu'eucaryotes (Andersson, 2005 ; Keeling & Palmer, 2008 ; Zolfaghari Emameh *et al.*, 2016). L'histoire des algues est elle aussi parsemée par ce genre d'évènements. Des cas de transferts de gènes horizontaux (TGH) ont été identifiés aussi bien entre le génome de bactéries ou d'archébactéries vers celui d'une algue (Nosenko & Bhattacharya, 2007 ; Qiu *et al.*, 2013a,b ; Schönknecht *et al.*, 2013), ainsi que les cas particuliers d'un TGH d'une bactérie vers le génome chloroplastique (Mackiewicz *et al.*, 2013), ou entre un champignon et une algue terrestre (Beck *et al.*, 2014). Chez les algues, un cas particulier de TGH d'eucaryote à eucaryote est également connu : le transfert de gènes depuis le génome de l'endosymbionte (chloroplastique aussi bien que nucléaire) vers le génome de l'hôte, ou transfert de gènes endosymbiotique (TGE). Ceux-ci ont joué un rôle primordial dans la domestication de l'endosymbionte et l'établissement des communications entre le noyau et le chloroplaste (Martin *et al.*, 2002 ; Timmis *et al.*, 2004). L'incorporation, la rétention et l'utilisation de ces gènes par l'organisme lui permettent d'enrichir son patrimoine génique, lui conférant ainsi un avantage adaptatif décisif (Chan *et al.*, 2011 ; Maruyama *et al.*, 2011 ; Schönknecht *et al.*, 2013 ; Bhattacharya *et al.*, 2013).

Ces génomes mosaïques compliquent toutefois l'élucidation de l'histoire évolutive des organismes et conduisent certaines études à envisager des phénomènes d'endosymbiose dits cryptiques. Ces études partent de l'observation que dans le génome de microalgues des lignées CASH, pourtant issues d'endosymbioses liées à une microalgue rouge, des gènes de la lignée verte ont été identifiés (Frommolt *et al.*, 2008 ; Moustafa *et al.*, 2009 ; Woehle *et al.*, 2011). Tout comme des gènes de la lignée rouge ont été identifiés dans le génome de chlorarachniophytes et d'euglènes pourtant issues d'une endosymbiose secondaire d'une microalgue verte (Maruyama *et al.*, 2011 ; Yang *et al.*, 2014). La présence de ces gènes a été expliquée de deux façons : (i) un évènement d'endosymbiose d'une microalgue verte chez l'ancêtre commun des lignées CASH, dont le chloroplaste « d'origine verte » aurait ensuite été remplacé par l'endosymbiose d'une microalgue rouge. Les gènes de la lignée verte identifiés dans le génome de l'hôte auraient été transférés par TGE lors de la domestication de l'endosymbionte avant son remplacement par l'endosymbiose d'une microalgue rouge (Figure 13).

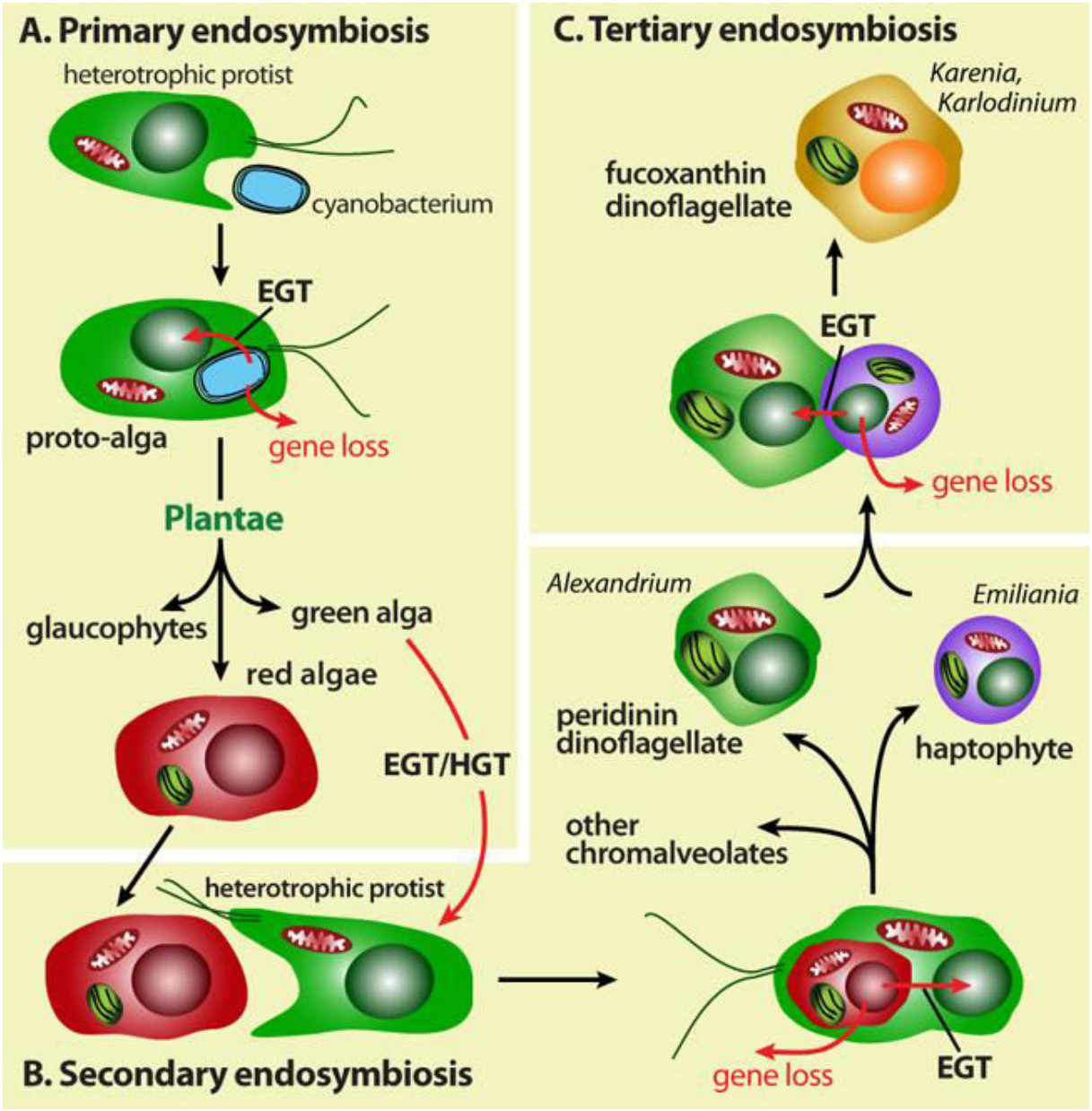

Figure 12 : représentation schématique des transferts de gènes (horizontaux et endosymbiotiques) au cours de différents évènements d'endosymbioses de l'histoire évolutive des algues (Chan & Battacharya 2010).

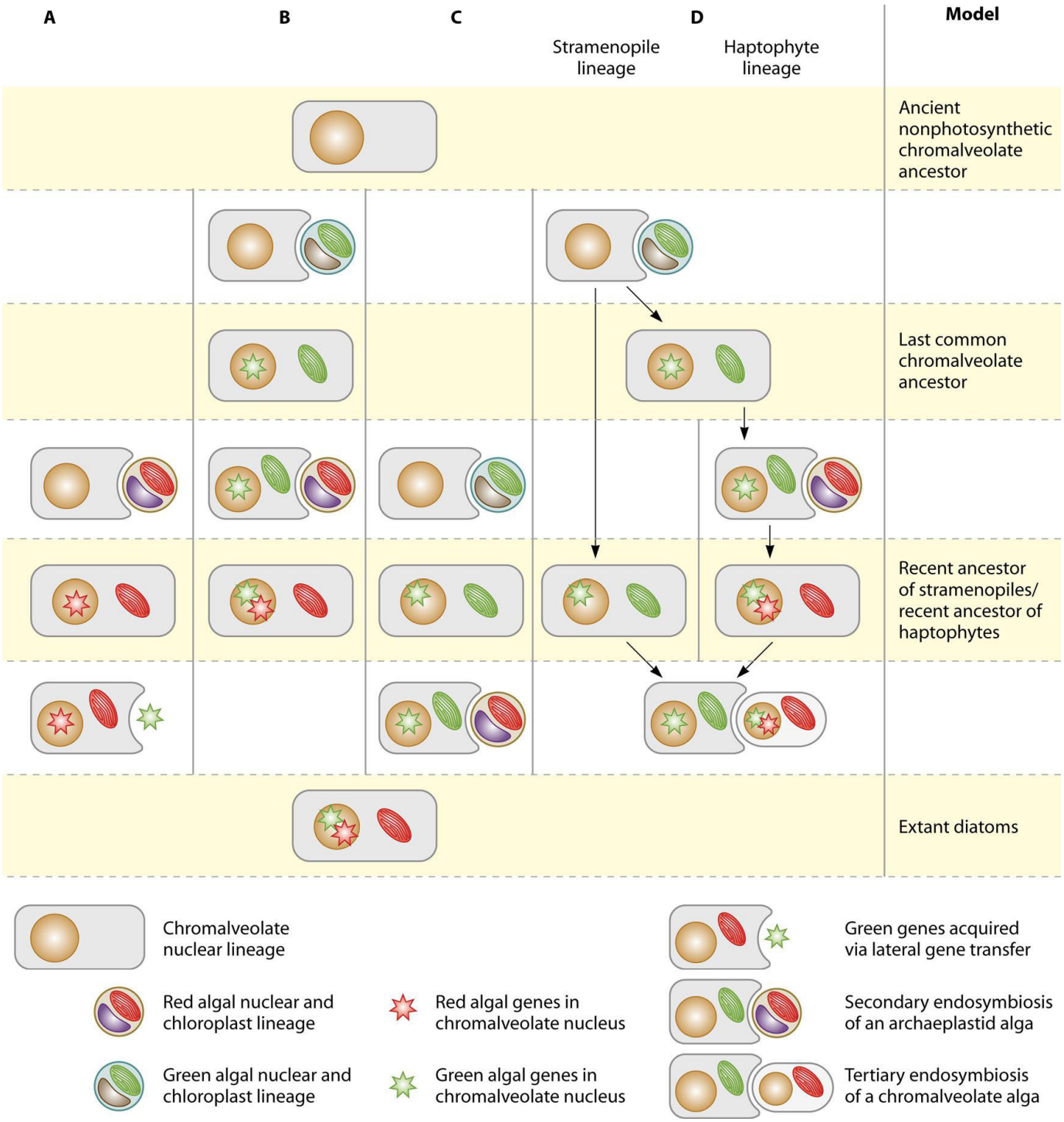

Figure 13 : les différentes hypothèses proposées par Dorrell & Smith (2011) afin d'expliquer la présence de gènes microalgues vertes dans le génome de microalgues des lignées CASH, exemple des diatomées. A, un transfert de gènes latéral récent; B, une ancienne endosymbiose secondaire; C, une récente endosymbiose secondaire; D l'endosymbiose tertiaire d'une haptophyte. (Dorrell & Smith 2011)

L'inverse étant proposé dans le cas des chlorarachniophytes et des euglènes (l'endosymbiose d'une microalgue rouge suivie par celle d'une microalgue verte). Cependant, ces hypothèses ont été critiquées et réfutées puisqu'elles sont en partie fondées sur l'hypothèse des chromalvéolés (l'existence d'un ancêtre commun aux lignées CASH). De plus, le lot de gènes « d'origine verte » transférés dans le génome de l'ancêtre commun aurait dû se retrouver chez chacune des lignées CASH, hors ce n'est pas le cas. Enfin, des analyses secondaires ont montré que le nombre de ces gènes était surestimé et/ou que leur lien à une endosymbiose dite cryptique serait dû à un biais d'analyse (Deschamps & Moreira, 2012 ; Burki *et al.*, 2012). (ii) Ces gènes résulteraient donc plus probablement d'évènements de TGH. Lesquels seraient peut être liés à la prédation de différents types de microalgues avant l'endosymbiose définitive (Keeling, 2013 ; Burki, 2014).

Les différentes hypothèses élaborées semblent donc de plus en plus à même d'expliquer les résultats obtenus par des études de plus en plus complètes du fait du nombre croissant d'organismes séquencés. De plus, les mécanismes par lesquels un endosymbionte est retenu puis domestiqué sont encore obscurs et l'étude d'organismes chez lesquels ce phénomène a eu lieu il y a plus ou moins longtemps (différents niveaux d'endosymbiose) peut aider à mieux en comprendre la nature. Mais depuis lors, ces organismes ont évolués, rendant difficile cette compréhension. Dans cette optique, une autre voie a également été explorée il y a quelques années. En effet, une telle symbiose a été recréée en laboratoire entre une cyanobactérie (*Synechocystis*) et une paramécie (*Paramecium brusaria*) (Ohkawa *et al.*, 2011). Cette approche pourrait permettre de mieux comprendre les premiers pas de cette longue histoire, comme les adaptations métaboliques nécessaires au passage de la vie libre à la vie endosymbiotique (Sørensen *et al.*, 2016).

Toutefois, l'ensemble de cette histoire est encore loin d'être élucidée. Afin de confirmer les hypothèses proposées, des études globales s'intéressant aux différents génomes (chloroplastique et nucléaire) ainsi que des études plus ciblées sur un type de gènes particuliers, tels que les facteurs de transcription, ont été menées (Rayko *et al.*, 2010 ; Buitrago-Flórez *et al.*, 2014). Ces derniers travaux permettent notamment de mettre en évidence des familles de facteurs de transcription spécifiques d'une lignée de microalgue et ainsi de retracer leur histoire évolutive. Cependant, certaines lignées comportent encore peu d'organismes dont le génome (nucléaire et/ou chloroplastique) est séquencé (rhodophytes, glaucophytes, chlorarachniophytes, euglènes,

cryptophytes et haptophytes). Au fil du temps, cet échantillonnage s'élargit et permet d'élaborer des hypothèses de plus en plus robustes et complètes, mais certaines lignées restent encore peu représentées (Kim *et al.*, 2014; Bhattacharya *et al.*, 2015).

IV. Intérêt biotechnologique et valorisation des microalgues

Cette histoire évolutive longue et complexe résulte en une très grande diversité des microalgues, dont chaque lignée a mené sa propre évolution. Cette diversité s'illustre autant d'un point de vue de leur morphologie (sphérique, ovoïde, fusiforme, cylindrique, pyramidale), que par leur plasticité métabolique qui leur permet de coloniser de nombreux milieux : marins, eaux douces, eaux saumâtres, ainsi que des milieux extrêmes en terme de salinité, température ou encore de pH. Cette plasticité métabolique est la source de nombreuses molécules d'intérêt biotechnologique dont les applications sont très variées (Figure 14) (Koller *et al.*, 2014 ; Yaakob *et al.*, 2014). De plus, leur capacité à convertir l'énergie solaire en biomasse avec un rendement supérieur à celui des plantes terrestres et des macroalgues font de ces organismes une bonne alternative dans la production de composés à haute valeur ajoutée (Cadoret & Bernard, 2008). Autant d'atouts qui font que les microalgues sont utilisées pour des applications variées allant de l'alimentation animale à la bioremédiation, en passant par la santé humaine, la production de biocarburants jusqu'à l'exploitation des mécanismes impliqués dans la formation du frustule des diatomées (Figure 15) pour des applications nanotechnologiques (Kröger & Poulsen, 2008 ; Cadoret *et al.*, 2014).

1. Alimentation animale

L'un des premiers secteurs à faire appel aux microalgues est l'aquaculture qui les utilise traditionnellement pour l'élevage des larves et juvéniles de mollusques, de crevettes et de poissons (Benemann; Hemaiswarya *et al.*, 2011). L'utilisation de microalgues permet un apport adéquat en acides aminés essentiels, vitamines et acides gras poly-insaturés à longue chaîne (Brown *et al.*, 1997). Cet apport semble même optimal dans le cas de l'élevage des mollusques, puisqu'aucun régime de substitution n'est aussi performant (Coutteau & Sorgeloos, 1992).

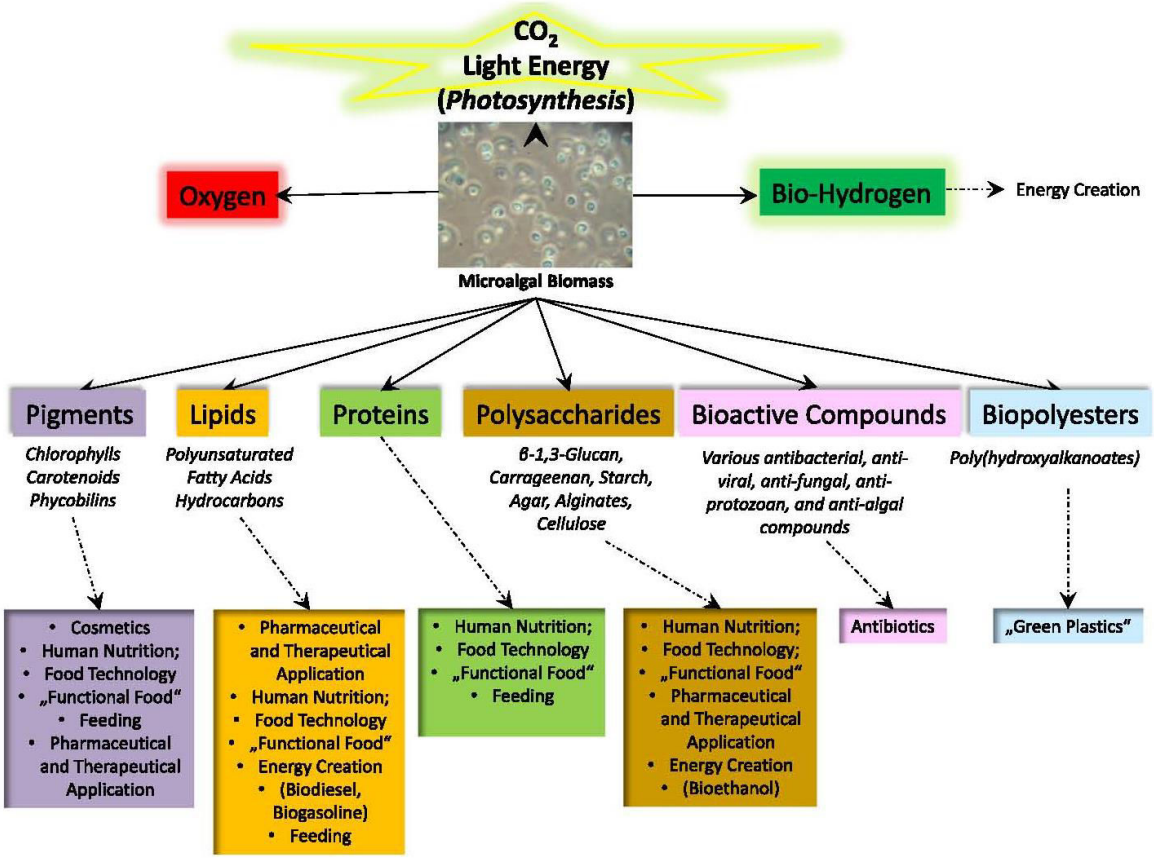

Figure 14 : les différentes voies de valorisation applicables aux microalgues via les composées à haute valeur ajoutée qu'elles synthétisent (Koller *et al.*, 2014).

Figure 15 : Diversité morphologique des frustules de diatomée. (source : http://coursbiologie.net/diatomees.html)

2. Santé humaine

Du fait de leur diversité ainsi que celle des habitats qu'elles colonisent, les microalgues sont une source très convoitée de molécules bioactives. Ainsi, des composés à valeur médicinale ont depuis longtemps été recherchés chez ces organismes (Metting & Pyne, 1986), et un grand nombre d'activités biologiques ont depuis été recensées. On retrouve des propriétés antioxydantes, anti-inflammatoires, anti-mutagènes chez les pigments (Mimouni *et al.*, 2012 ; de Morais *et al.*, 2015) et notamment l'astaxantine (Figure 16) (Shah *et al.*, 2016), des activités antimicrobiennes et antivirales (Borowitzka, 1997; de Jesus Raposo *et al.*, 2013) ou encore les activités hypocholestérolémiantes des acides gras poly-insaturés (Mimouni *et al.*, 2012). De plus, la production d'anticorps monoclonaux chez la microalgue *Chlamydomonas reinhardtii* a suscité un vif intérêt (Franklin & Mayfield, 2005). Cette alternative au système mammifère classiquement utilisé semble très avantageuse du fait des plus faibles couts de production et de la facilité de transformer le génome nucléaire et chloroplastique de cette microalgue. De manière plus générale, les avantages que présentent les microalgues en matière de rendements et de couts de production ainsi que la mise au point d'outils de transformation génétique (Figure 17) font de ces organismes des plateformes biotechnologiques prometteuses (Cadoret *et al.*, 2012 ; Scaife & Smith, 2016).

3. Alimentaire

La consommation de microalgues dans les pays occidentaux est certes peu courante, mais elle pourrait constituer un moyen de lutter contre la faim et la malnutrition dans les pays en voie de développement. La spiruline étant un organisme procaryote photosynthétique, il ne s'agit pas à proprement parlé d'une microalgue. Ces organismes font toutefois l'objet de nombreuses études dans cette optique (Habib *et al.*, 2008). Les spirulines sont d'ailleurs récoltées et consommées par les populations africaines et amérindiennes depuis des siècles. Les microalgues eucaryotes, pour leur part, sont utilisées comme compléments alimentaires du fait de leur forte teneur en protéines et lipides, notamment les acides éicosapentaénoïque (EPA) et docosahexaénoique (DHA) qui ne peuvent pas être synthétisés par notre organisme (Mendes *et al.*, 2008).

Figure 16 : Culture de la microalgue *haematococcus pluvialis* en vue de la production d'astaxanthine en chine (à gauche) et en Israël (à droite). (sources : http://bggworld.com/astaxanthin-astazinetm/ et http://www.israel21c.org/is-2016-the-year-of-the-algae/)

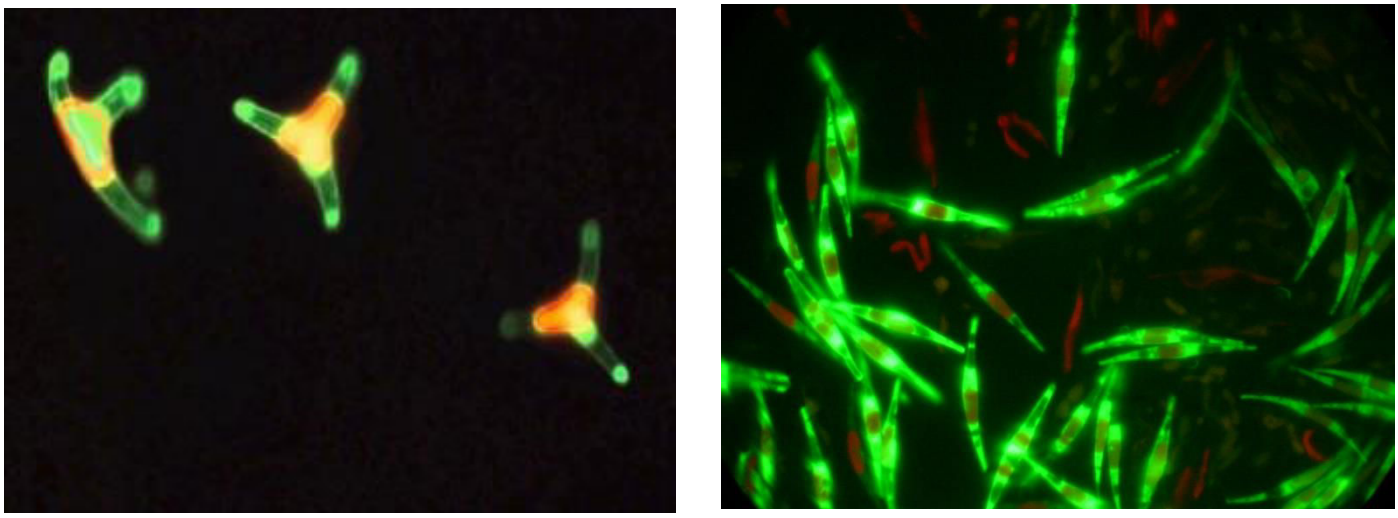

Figure 17 : visualisation par microscopie à fluorescence de la diatomée modèle *Phaeodactylum tricornutum* transformée génétiquement pour exprimer la protéine fluorescente GFP. Source : laboratoire PBA, IFREMER Nantes

Cependant, les applications dans ce domaine sont freinées par des couts de production encore élevés (Becker, 2003). Enfin, en France, seule la spiruline et les espèces *Odontella aurita* et *Chlorella* sp. bénéficient d'une autorisation pour la consommation humaine, délivrée par l'ANSES. A titre de comparaison, 19 espèces de macroalgues sont actuellement autorisées à la consommation.

4. Bioremédiation

La protection de notre environnement est une question qui suscite de plus en plus d'intérêt depuis quelques années. Dans cette optique, les microalgues peuvent être un outil particulièrement adapté. En effet, grâce à leur surface spécifique élevée (rapport entre la surface et le volume de l'organisme), aux polymères qu'elles produisent et aux mécanismes d'absorption qu'elles ont développés au cours de leur évolution, les microalgues ont des capacités intéressantes pour traiter des milieux pollués ou eutrophiques. Ces applications vont de l'épuration d'effluents d'élevage (Godos *et al.*, 2009) et d'eaux urbaines ou industrielles (Martínez *et al.*, 2000 ; Lim *et al.*, 2010) grâce à leur capacités d'absorption de l'azote et du phosphore, jusqu'au traitement d'eaux polluées par des métaux tels que le cadmium, le nickel, le mercure ou le plomb (Aksu, 1998; Chen *et al.*, 1998; Sialve *et al.*, 2009; Fouilland, 2012).

5. Biocarburants

Dans certaines conditions de culture (en limitation azotée, par exemple), certaines espèces de microalgues sont capables de produire de grandes quantités de composés tels le triacylglycérol et l'amidon. Ces composés peuvent être utilisés pour la production de biodiesel à partir de triglycérides, de bioéthanol à partir d'amidon, ou encore de biogaz à partir de la méthanisation des microalgues. Ces énergies renouvelables dérivées de microalgues sont également appelées biocarburants de troisième génération (Figure 18). Des recherches sur la production de biodiesel à partir de microalgues ont été initiées durant la crise énergétique des années 1970, et cette thématique a connu un regain d'intérêt à travers le monde depuis les années 2000.

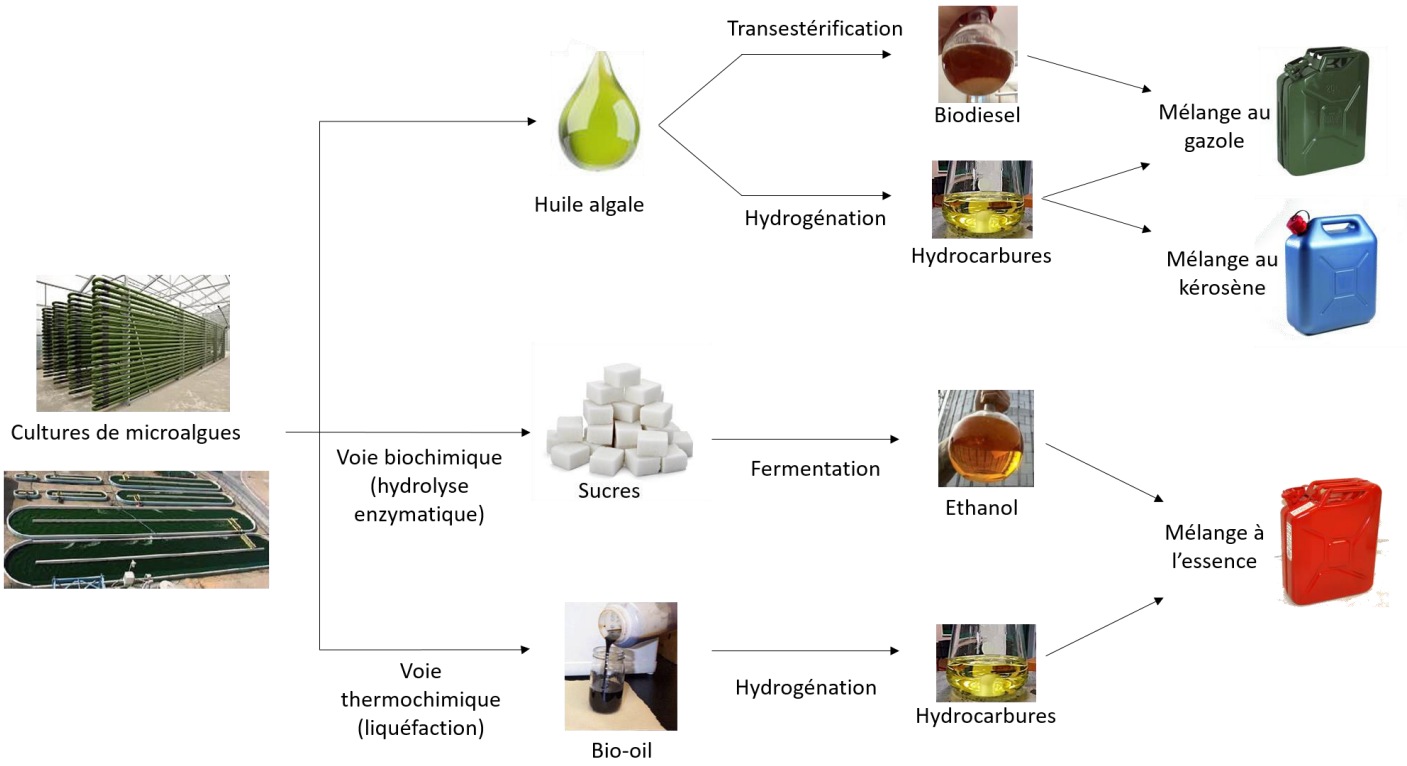

Figure 18 : production de biocarburant à partir de microalgues (adaptée de http://www.ifpenergiesnouvelles.fr/Espace-Decouverte/Tous-les-Zooms/Des-biocarburants-a-partir-de-microalgues).

Les microalgues ayant des rendements de production de biomasse et d'huile à l'hectare supérieurs à ceux des plantes terrestres, leur utilisation pour la production de biocarburants est très avantageuse (Cadoret & Bernard, 2008). L'utilisation des microalgues permet également d'éviter la compétition pour les terres arables ainsi que l'exploitation de leurs capacités à la remédiation du CO_2 et des eaux usées, lesquelles permettent la production de biomasse ensuite valorisée par la production de biocarburants (Mata *et al.*, 2010 ; Brennan & Owende, 2013). Cependant, et malgré ces avantages prometteurs, la production de biocarburants de troisième génération en est encore à ses débuts et des progrès restent à faire pour qu'elle soit économiquement viable (Chisti, 2013 ; Fields *et al.*, 2014).

V. Mieux comprendre le métabolisme et sa régulation pour optimiser la production de composés d'intérêt

La production de ces composés à haute valeur ajoutée dépend de la croissance des microalgues et des paramètres de culture. L'application de facteurs de stress environnementaux (privation nutritive, température, salinité, lumière…) aux cultures permet de favoriser la production de certains composés (Markou & Nerantzis, 2013). Toutefois, l'application de ces stress peut influencer la croissance des microalgues et, en conséquence, le rendement de production du composé ciblé (Adams *et al.*, 2013). Dans le but d'éviter cet inconvénient, des études ont cherché à mieux comprendre le métabolisme des microalgues afin de mieux prédire quel stress et quelle amplitude de ce stress permettraient d'optimiser la production du composé ciblé (Baroukh *et al.*, 2015). Ces études pouvant aboutir soit à une approche d'orientation métabolique par application des conditions optimales prédites, soit à une approche plus ciblée (Courchesne *et al.*, 2009 ; Yu *et al.*, 2011). Dans ce dernier cas, des gènes clés impliqués dans l'acclimatation métabolique de la microalgue ou de la voie de biosynthèse du composé d'intérêt sont identifiés. Des techniques de biologie moléculaire telles la transgenèse ou le RNA silencing, encore peu mises au point chez les microalgues mais de plus en plus développées (Walker *et al.*, 2005 ; Jinkerson & Jonikas, 2015), sont ensuite utilisées afin d'augmenter ou réprimer l'effet de ces gènes, dans le but d'augmenter la production du composé d'intérêt.

Compte tenu des promesses des microalgues dans la production de biodiesel dérivés de lipides de réserves, nombre d'études se sont focalisées sur ces composés. A la fois des approches de

répression (Trentacoste *et al.*, 2013 ; Ma *et al.*, 2014) et de surexpression (Xue *et al.*, 2015) de gènes codant des enzymes clé du métabolisme lipidique (biosynthèse et catabolisme) ont abouti à des résultats prometteurs. L'accumulation lipidique étant associée à une condition de culture limitée en azote, la compréhension du métabolisme de l'azote et de son lien avec l'accumulation lipidique est elle aussi primordiale. Le lien entre ces deux métabolismes a notamment été exploité chez la diatomée *P. tricornutum* pour obtenir une augmentation de la production de lipides suite à la répression du gène de la nitrate réductase, premier point de contrôle de la voie de l'assimilation de l'azote (Levitan *et al.*, 2015). Cependant, ces approches ne ciblent qu'une enzyme, alors que les processus impliqués dans le remaniement des flux de carbone dans la cellule en réponse au stress azoté sont multiples et interconnectés. Cette action combinée est bien illustrée par une étude chez la microalgue *Chlorella minutissima* (Hsieh *et al.*, 2012) chez qui l'expression de cinq enzymes exogènes a permis de doubler la production de triglycérides, alors que séparément, aucune de ces enzymes n'avait d'effet sur cette production. Du fait de cette complexité du métabolisme et des connexions entre les différentes voies qui le composent, il semblerait plus efficace de cibler plusieurs enzymes intervenant spécifiquement dans la voie de biosynthèse du composé d'intérêt plutôt que de moduler indépendamment l'expression d'une seule enzyme (Hsieh *et al.*, 2012). Cependant, cibler spécifiquement plusieurs enzymes à la fois par génie génétique peut parfois se révéler délicat pour la plupart des organismes. En conséquence, une approche ciblant des gènes clés tels que les facteurs de transcription semble particulièrement intéressante (Courchesne *et al.*, 2009 ; Rabara *et al.*, 2014). En effet, les facteurs de transcription (FTs) sont connus pour intervenir dans les étapes clés de la régulation de la transcription génique. De plus, parmi les différents remaniements métaboliques induits durant un stress biotique ou abiotique, il apparaît que l'expression de certaines enzymes métaboliques spécifiques soit sous la dépendance de quelques FTs seulement (Vom Endt *et al.*, 2002 ; Czemmel *et al.*, 2009). En conséquence, la modulation de l'expression d'un de ces FTs impliqués dans ces remaniements métaboliques pourrait avoir un effet positif sur la production d'un composé d'intérêt. Ainsi, s'intéresser aux FTs et à leur rôle dans la régulation de la transcription ouvre des perspectives de recherche intéressantes pour la production de métabolites d'intérêt mais également pour une meilleure compréhension de la biologie de la cellule en général.

VI. La transcription chez les eucaryotes

Chez tout organisme vivant, le développement, la mise en place de la morphologie, de la physiologie ainsi que des mécanismes tels que ceux permettant l'acclimatation aux changements du milieu, sont le fait de la modulation de l'expression des gènes. L'expression d'un gène consiste à décoder l'information contenue dans sa séquence génomique afin de produire une protéine fonctionnelle. La première étape de cette expression des gènes est la transcription. Elle consiste à transcrire l'information contenue dans la séquence génomique des gènes en un ARN simple brin. Contrairement aux procaryotes chez qui ce processus n'est assuré que par une seule ARN polymérase, les eucaryotes en utilisent trois principales (Haag & Pikaard, 2011). Chacune d'entre elles assurant la transcription de différentes classes de gènes. L'ARN polymérase I assure la transcription des ARN ribosomiques dans le nucléole (28S, 18S, 5.8S). L'ARN polymérase II assure la transcription des ARN messagers (Kornberg, 2007). Et l'ARN polymérase III est impliquée dans la transcription des ARN de transferts et autres petits ARN (5S, ARN non codants, Short Interspersed Nuclear Elements (SINEs)…). De plus, deux ARN polymérases supplémentaires (ARN polymérases IV et V), spécifiques aux plantes terrestres, ont été identifiées. Celles-ci sont impliquées dans la synthèse de petits ARN interférents, jouant un rôle dans la régulation de la transcription de certains gènes (Herr, 2005 ; Landick, 2009).

Parmi les différentes ARN polymérases, la plus étudiée est l'ARN polymérase II puisqu'elle intervient dans la transcription des ARN messagers qui sont destinés à produire les protéines fonctionnelles permettant le fonctionnement d'une cellule. Dans le cas de cette ARN polymérase II, l'activation de la transcription dépend d'un complexe multi-protéique : le complexe de pré-initialisation de la transcription (CPI). Contrairement à l'ARN polymérase procaryote, l'ARN polymérase II ne reconnait pas elle-même le promoteur des gènes, c'est le CPI qui joue ce rôle. Celui-ci est constitué des facteurs de transcription (FTs) dit généraux puisqu'ils interviennent dans l'initialisation de la transcription de tous les gènes ciblés par l'ARN polymérase II. Le premier FT général impliqué est le TFIID. Celui-ci est un complexe protéique composé d'une protéine appelée TBP (TATA-box binding protein) et de TAFs (TBP associated factors). Sa composition peut varier d'un gène à un autre ou, par exemple, en fonction du type de tissu (Verrijzer, 2001). La protéine TBP cible une séquence particulière au sein des promoteurs des gènes, la boîte TATA,

permettant la fixation du TFIID. Sont ensuite recrutés par interactions protéine-protéine les facteurs généraux TFIIA, TFIIB, TFIIF (associé à l'ARN polymérase II), TFIIE et TFIIH (Orphanides *et al.*, 1996) (Figure 19). Cependant, ce modèle général suppose que tous les gènes transcrits par l'ARN polymérase II présentent une boîte TATA dans leur promoteur proximal. Or, ce n'est pas toujours le cas. En effet, la boîte TATA n'est retrouvée que dans 13 % des promoteurs de la levure *Saccharomyces cerevisiae* (Basehoar *et al.*, 2004) et 10 % chez l'Homme (Yang *et al.*, 2007). Néanmoins, des séquences de fixation alternatives à la boîte TATA ont été identifiées (Seizl *et al.*, 2011), permettant d'expliquer, en partie, que même des gènes n'ayant pas de boîte TATA au sein de leur promoteur requièrent l'intervention de TBP pour leur expression (Pugh & Tjian, 1991 ; Kim & Iyer, 2004). Il en découle donc l'existence de différents modes de reconnaissance des promoteurs ainsi qu'une plasticité de la composition du CPI et des protéines associées à celui-ci (Chang & Jaehning, 1997 ; Sikorski & Buratowski, 2009). Cette plasticité permet également une certaine spécificité de l'activation des gènes en fonction du type de tissu ou du stade de développement de l'organisme (Deato & Tjian, 2008 ; Kazantseva & Palm, 2014).

VII. La régulation de la transcription

1. Les facteurs de transcription : structure et fonctionnement

Une régulation plus fine et plus spécifique est nécessaire lorsqu'un organisme doit faire face aux modifications de son environnement (disponibilité en nutriments, changements de température …) ou au cours de son développement. Ces fines régulations sont notamment assurées par l'action des facteurs de transcription (FTs). Ces acteurs clés sont caractérisés par la présence dans leur séquence protéique d'un domaine de liaison à l'ADN et d'un domaine effecteur permettant des interactions protéine-protéine (Figure 20). Les domaines de liaison à l'ADN permettent une interaction avec la double hélice de différentes manières en fonction de leur structure tridimensionnelle (Luscombe *et al.*, 2000). La majorité des FTs (aussi appelées facteur *trans*) ne possèdent qu'un seul type de domaine de liaison à l'ADN qui peut aussi bien être présent en une seule qu'en plusieurs copies au sein de la séquence (Charoensawan *et al.*, 2010b). Il existe une grande diversité de domaines de liaison à l'ADN et, de ce fait, de FTs. Ces derniers sont donc

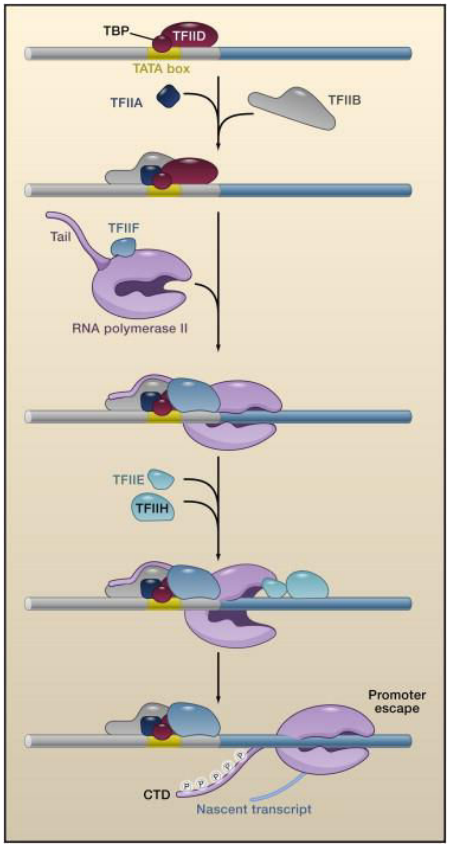

Figure 19 : schématisation des différentes étapes de l'assemblage du complexe de pré-initialisation de la transcription au niveau d'un promoteur eucaryote (Levine, 2011).

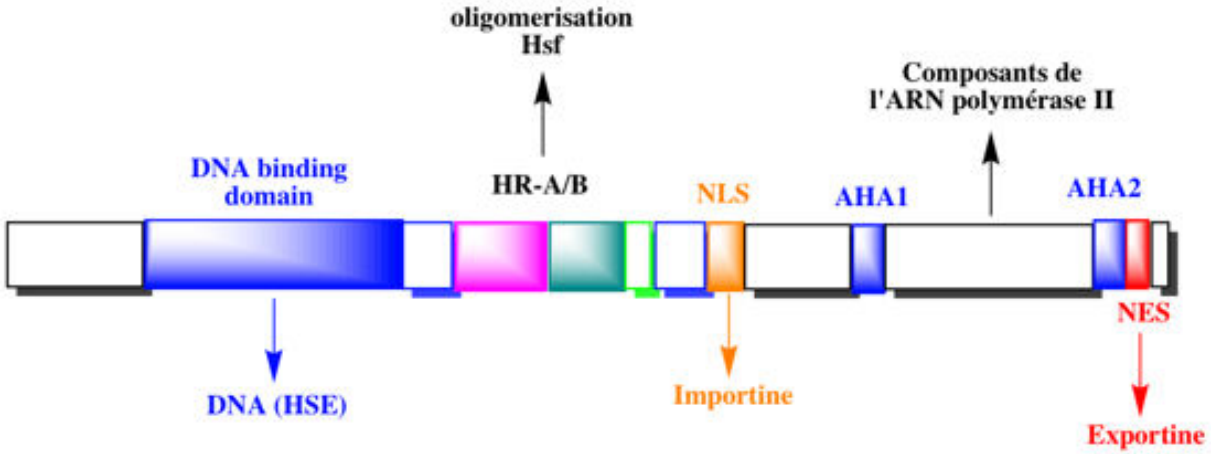

Figure 20 : structure d'un facteur de transcription (FT), exemple d'un Heat Shock Factor (HSF). Le domaine de liaison à l'ADN permet la fixation du HSF sur son site spécifique appelé Heat Shock Element (HSE). Les domaines HR-A et HR-B permettent l'oligomérisation des HSFs, puisqu'ils sont actifs sous forme de trimères. Le NLS est un signal de localisation cellulaire adressant le FT au noyau suite à sa production dans le cytoplasme. Les domaines AHA1 et AHA2 sont les domaines effecteurs des HSFs, leur permettant notamment d'interagir avec les composants du CPI. Le NES est une séquence d'exportation nucléaire. Source : http://biochimej.univ-angers.fr/Page2/COURS/Zsuite/4StressLeaHsp/1StressLeaHsp.htm

classés en familles, chacune d'elles étant caractérisée par la présence d'un domaine ou d'une combinaison de domaines de liaison à l'ADN. Ces domaines de liaison à l'ADN sont caractérisés par une séquence protéique conservée (Luscombe *et al.*, 2000 ; Riechmann et al 2000). Par exemple, les FTs de la famille ERF (Ethylene Response Factors) sont caractérisés par la présence du seul domaine de liaison à l'ADN de type APETALA2 (AP2) (Jin *et al.*, 2014) (Figure 21).

Les facteurs de transcription en se fixant à l'ADN via les domaines de liaison, modifient l'expression des gènes. Les domaines de liaison à l'ADN reconnaissent et se fixent à des séquences spécifiques (6 à 12 paires de bases en moyenne), appelées éléments *cis*, situées au niveau des séquences régulatrices des gènes. Ces séquences régulatrices peuvent aussi bien faire partie du promoteur proximal (jusqu'à 1 000 ou 1 500 paires de bases en amont du site d'initiation de la transcription), que des régions activatrices ou inhibitrices distales situées jusqu'à plusieurs dizaines de milliers de paires de bases en amont ou en aval du gène et même au niveau des introns (Blackwood & Kadonaga, 1998), ou encore inter-chromosomiques (Spilianakis *et al.*, 2005). Les séquences distales sont appelées « enhancer » quand elles ont une action activatrice sur la transcription, et « silencer » quand leur action provoque une inhibition de la transcription (Figure 22).

Quel que soit le type de séquence promotrice impliquée, le mode de régulation de la transcription fait intervenir des FTs. Une fois liés à leur site de fixation au sein de ces séquences promotrices, les FTs jouent leur rôle de régulateur de la transcription en recrutant des protéines partenaires grâce à leur domaine effecteur. Ces domaines effecteurs permettent la régulation de la transcription via des interactions protéine-protéine avec des protéines composant le CPI ou des enzymes modifiant la structure de la chromatine (Figure 23). Les domaines effecteurs permettant de réguler l'expression des gènes via un contact direct avec les composants du CPI ont pour effet de renforcer le recrutement et la stabilisation du CPI. Contrairement aux domaines de liaison à l'ADN, les domaines trans-activateurs ne sont pas caractérisés par une forte homologie de séquence, mais par la proportion d'un certain type d'acide aminé dans leur séquence protéique (Johnson et al 1993). Une de ces classes de domaines est caractérisée par une proportion d'acides-aminés de nature acide produisant une forte charge nette négative (Hope & Struhl, 1986 ; Hahn, 1993).

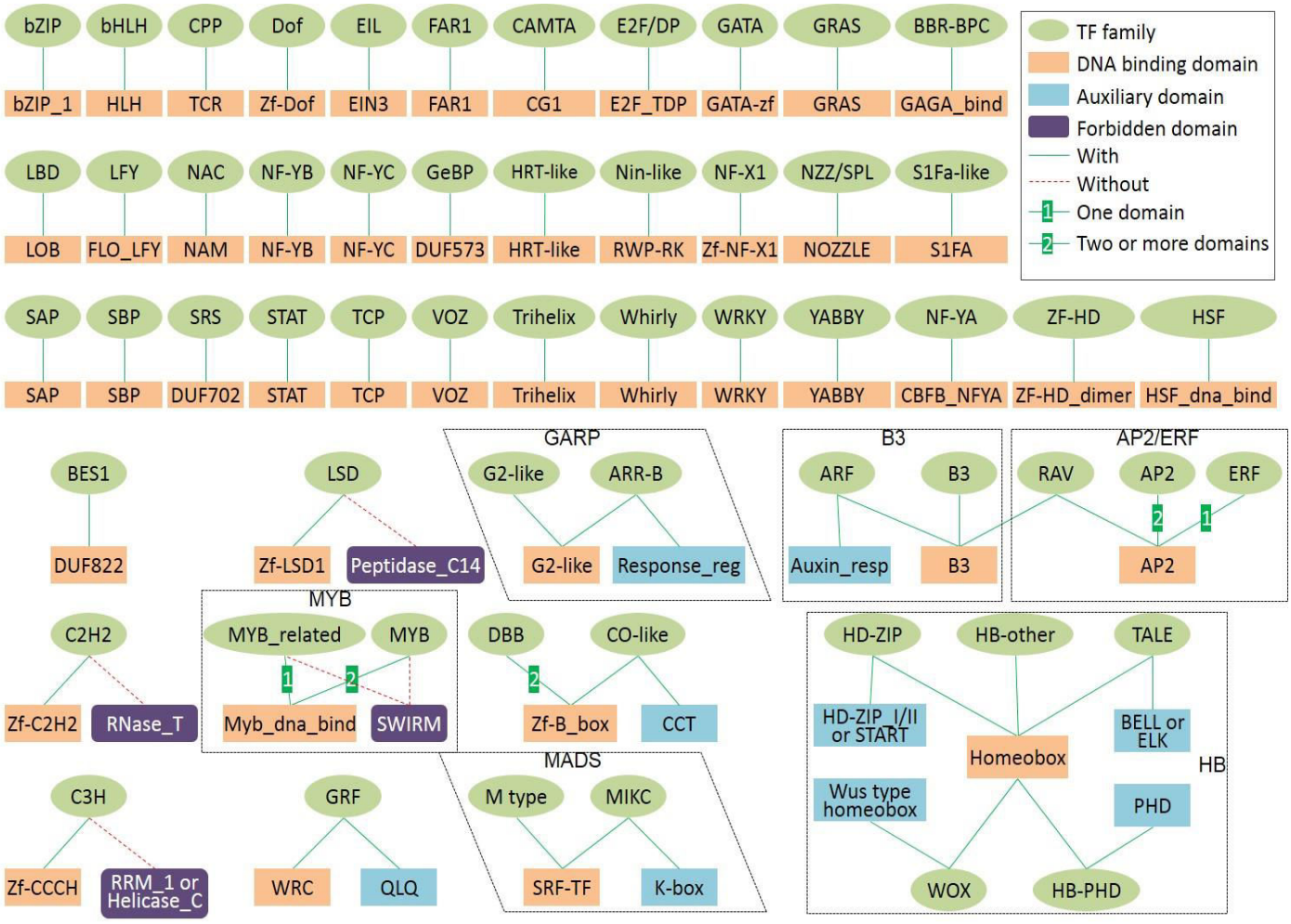

Figure 21 : familles de FTs connues chez les plantes et les règles d'assignement correspondant à chacune d'elles. Source : PlantTFDB.com

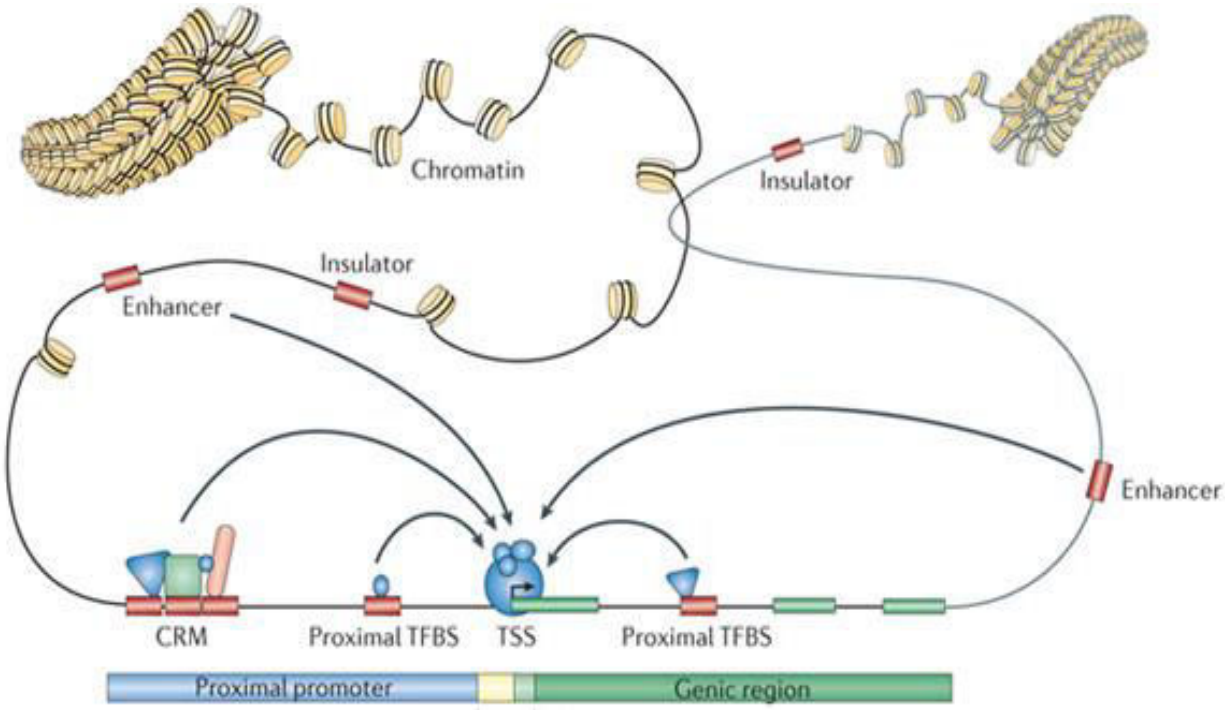

Figure 22 : schéma illustrant les différents types de séquences promotrices intervenant dans la régulation de la transcription, et leur distance par rapport au site d'initiation de la transcription (TSS). Un promoteur proximal est situé immédiatement en amont du TSS. En son sein, se trouvent des sites de fixation des FTs (TFBS), parfois regroupés en clusters (CRM). Des TFBS peuvent également être situés plus loin (enhancers/silencers) en amont, en aval ou dans la séquence du gène (TFBS dans la région génique, en vert) (Lenhard *et al.*, 2012).

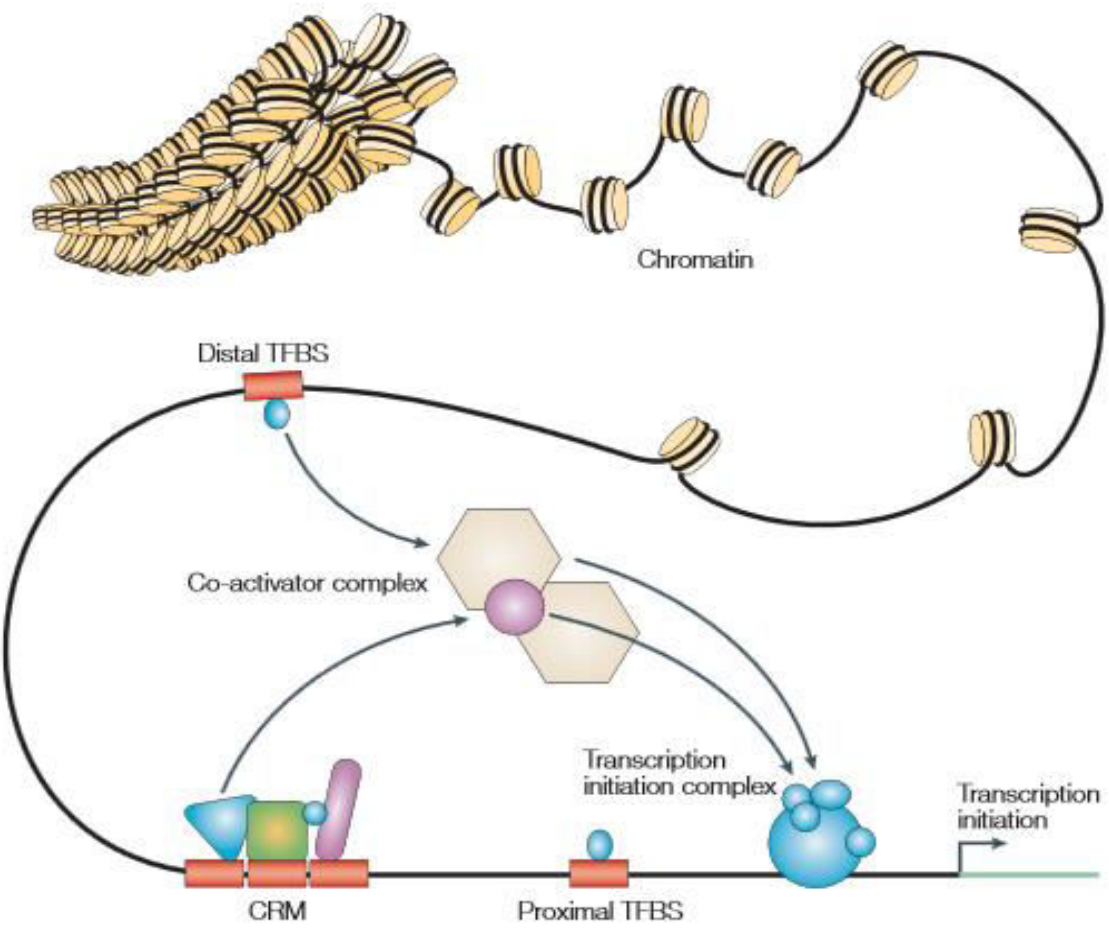

Figure 23 : schéma représentant l'action des différents acteurs de la régulation de la transcription. Les FTs liés à leur site de fixation (TFBS) agissent de concert via des interactions protéine-protéine afin de recruter des protéines partenaires (co-activateurs ou co-répresseurs ; ici Co-activator complex). Ces interactions ont pour effet de stabiliser et favoriser la formation du CPI (ici Transcription initiation complex) ou de modifier la structure de la chromatine (Wasserman & Sandelin, 2004).

Cette nature acide semble primordiale à la fonction du domaine, comme le suggère l'augmentation de l'activation de la transcription provoquée par un ajout de résidus acides au sein du domaine trans-activateur du FT GAL4 (Gill & Ptashne, 1987). Il en va de même pour les domaines riches en glutamine (Courey & Tjian, 1988) et en proline (Gerber *et al.*, 1994) dont au moins 25% de leur composition en acides-aminés correspondant à ces résidus, alors que les domaines riches en sérines et thréonines sont moins communs et pourraient faire intervenir un mécanisme de phosphorylation (Johnson *et al.*, 1993 ; DeFalco & Childs, 1996). Quant aux domaines HOB1/HOB2 ils ne sont retrouvés que chez les FTs proto-oncogènes Jun et Fos (Sutherland *et al.*, 1992). Ces différents domaines permettent aux FTs d'interagir avec différents types de protéines, aboutissant à l'implication de nombreux acteurs dans la régulation de l'expression de leurs gènes cibles.

2. Les FTs dans la régulation de la transcription : action combinée de nombreux acteurs

Grâce à ces domaines trans-activateurs, les FTs interagissent avec des éléments du CPI tels le TFIIB ou encore certains TAFs entrant dans la composition du TFIID (Choy & Green, 1993 ; Kim & Roeder, 1994 ; Chiang & Roeder, 1995), entrainant un changement de conformation de ces FTs généraux. Ce changement de conformation a pour conséquence d'augmenter leur activité, c'est-à-dire leur capacité à recruter les autres composantes du CPI (Horikoshi *et al.*, 1988 ; Roberts & Green, 1994), renforçant ainsi la formation du CPI au niveau des promoteurs des gènes. Le recrutement et la stabilisation du CPI a également lieu via l'intervention d'un complexe protéique appelé Mediator. D'abord identifié chez la levure *Saccharomyces cerevisiae*, ce complexe semble très répandu chez les eucaryotes (Bourbon, 2008). Il interagit en solution (non lié à l'ADN) directement avec l'ARN polymérase II, le TFIIB, le TFIIF et le TFIIH, formant un complexe appelé holoenzyme (Kim *et al.*, 1994). Le Mediator agit comme une sorte de pont entre le FT et le CPI puisqu'il interagit directement à la fois avec le domaine trans-activateur du FT fixé sur le promoteur du gène, et avec l'ARN polymérase II via la formation de l'holoenzyme (Figure 24) (Koleske & Young, 1994 ; Hengartner *et al.*, 1995).

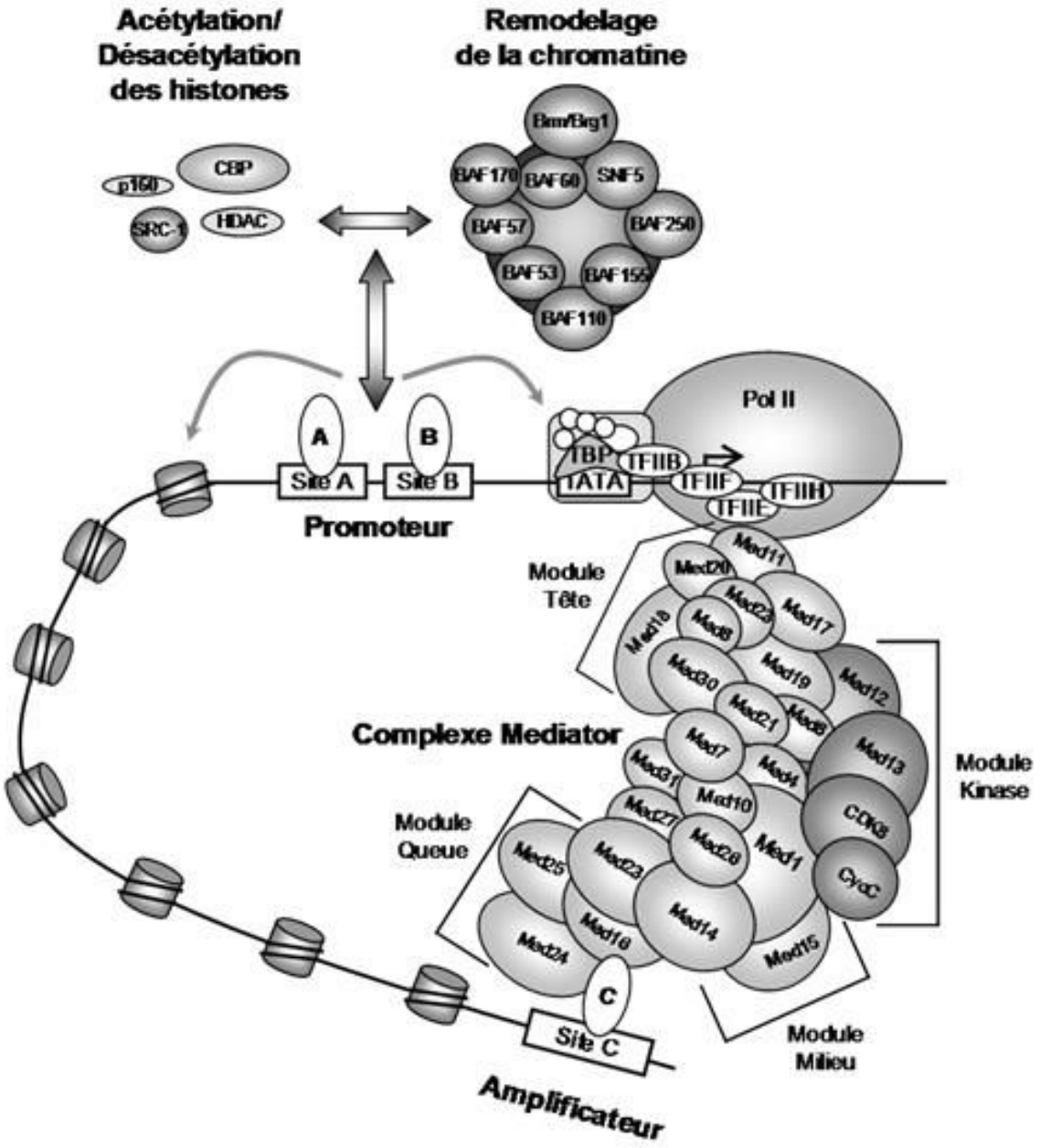

Figure 24 : schéma de l'intervention du complexe médiator dans la régulation de la transcription. Les FTs liés à leur site de fixation dans le promoteur du gène (A et B) recrutent des protéines partenaires modifiant la structure de la chromatine (acétylation/désacétylation des histones et remodelage de la chromatine). De plus, le recrutement du complexe médiator via un FT fixé au niveau d'une séquence *enhancer* permet de faciliter le recrutement de l'ARN polymérase II sur le site d'initiation de la transcription et de stimuler l'activité du CPI.(St-Pierre, 2009)

Outre cette facilitation et stabilisation de l'assemblage du CPI, la régulation de l'expression des gènes par les FTs se fait également via le recrutement de protéines agissant sur la structure de la chromatine. Dans le noyau d'une cellule eucaryote, l'ADN comporte plusieurs niveaux de compaction. L'unité de base de cette compaction est le nucléosome. Celui-ci est formé par un octamère d'histones (deux histones H2A, deux histones H2B, deux histones H3 et deux histones H4) (Richmond *et al.*, 1984) autour duquel une longueur de 147 paires de bases d'ADN est enroulée. Chacun de ces nucléosomes est séparé par un court segment d'ADN dont la taille peut varier (10 à 80 paires de bases) (Felsenfeld & Groudine, 2003). Un niveau de compaction supérieur est ensuite atteint avec la formation d'une fibre compacte qui est stabilisée par l'intervention d'une cinquième histone, l'histone H1 (ou l'histone H5), se liant au nucléosome ainsi qu'au segment d'ADN le séparant du nucléosome suivant (Figure 25). Ainsi compactée, la chromatine, appelée hétérochromatine, offre peu de possibilités aux FTs et au CPI d'accéder au promoteur des gènes. Ceux-ci ne sont donc pas exprimés. En réponse à un stimulus, cette compaction peut être modifiée, rendant accessible des portions régulatrices du génome qui étaient compactées et inversement (Agalioti *et al.*, 2000) (Figure 26). Cette chromatine relâchée est dite euchromatine. En conséquence, l'état de compaction de la chromatine fait partie intégrante de la régulation de l'expression des gènes. Cette compaction est régulée par des mécanismes épigénétiques permettant une régulation transcriptionnelle en fonction de l'environnement ou du type de tissu (Racanelli *et al.*, 2008 ; Jung *et al.*, 2015). Les gènes situés dans les régions d'hétérochromatine sont donc difficilement accessibles. Toutefois, des études de cartographie des nucléosomes le long de génomes eucaryotes montrent que certaines régions promotrices proximales ont tendance à être libres de nucléosomes et donc accessibles aux FTs (Yuan *et al.*, 2005 ; Mavrich *et al.*, 2008 ; Zhang *et al.*, 2015). Cette caractéristique n'étant pas commune à tous les promoteurs, d'autres mécanismes doivent être mis en place afin de permettre aux FTs et au CPI d'accéder à ces promoteurs et de réguler l'expression des gènes qu'ils contrôlent.

Certains FTs ont la capacité de se lier à leur site de fixation lorsqu'il est situé sur une portion d'ADN enroulé autour d'un nucléosome. Ces FTs sont appelés FTs pionniers (Zaret & Carroll, 2011 ; Zaret & Mango, 2016) (Figure 27). Cette capacité est due au fait que la structure de leur domaine de liaison à l'ADN est particulière, comme celle des FTs FoxA dont la structure ressemble à celle des histones H5, tout en permettant la reconnaissance d'un site de fixation spécifique (Clark *et al.*, 1993). Les FTs FoxA ont également un domaine permettant une fixation

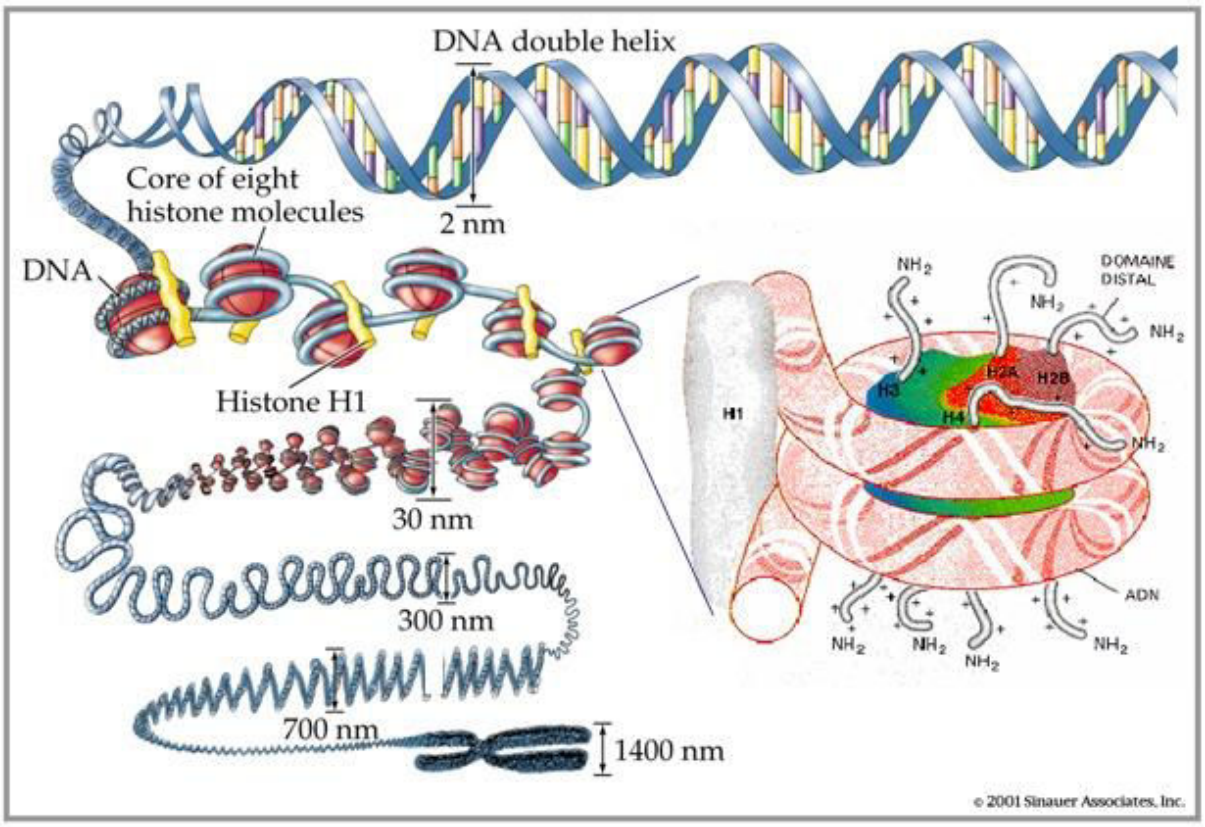

Figure 25 : schéma des différents niveaux de compaction de l'ADN dans le noyau d'une cellule eucaryote. La partie droite représente le schéma d'un nucléosome. *(*Fortin, 2005*)*

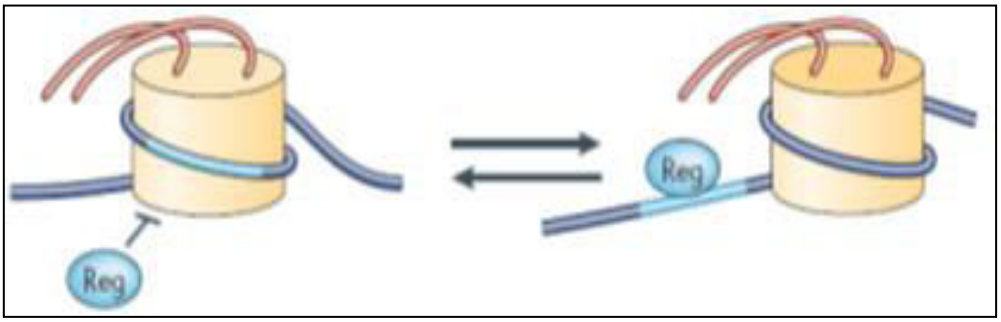

Figure 26 : rôle de la compaction de l'ADN dans la régulation de la transcription. Les modifications de la chromatine peuvent permettre la fixation de FTs (sphère bleue notée « Reg ») en rendant accessible leur site de fixation (portion bleue claire).

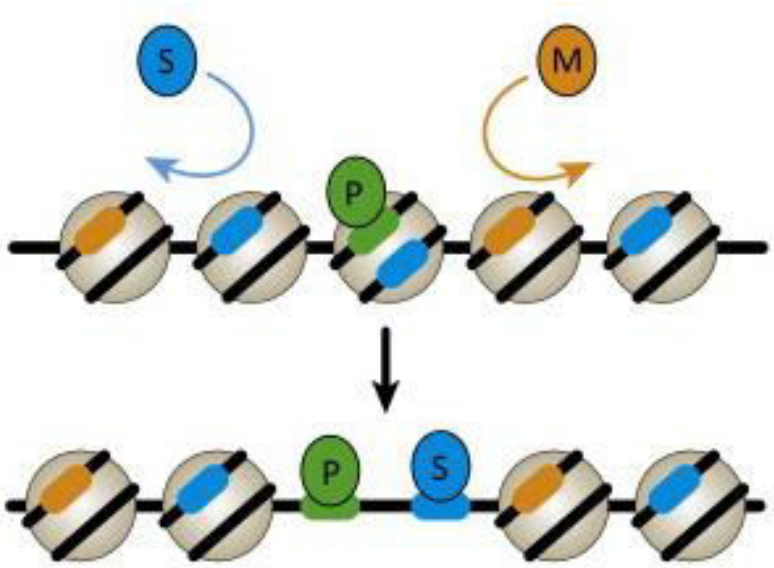

Figure 27 : représentation schématique du rôle d'un FT pionnier. Un FT pionnier (sphère verte notée P) a la capacité de se lier à son site de fixation lorsque celui-ci est situé dans une portion compactée de la chromatine. Sa fixation permet ensuite celle d'autres FTs (Slattery *et al.*, 2014).

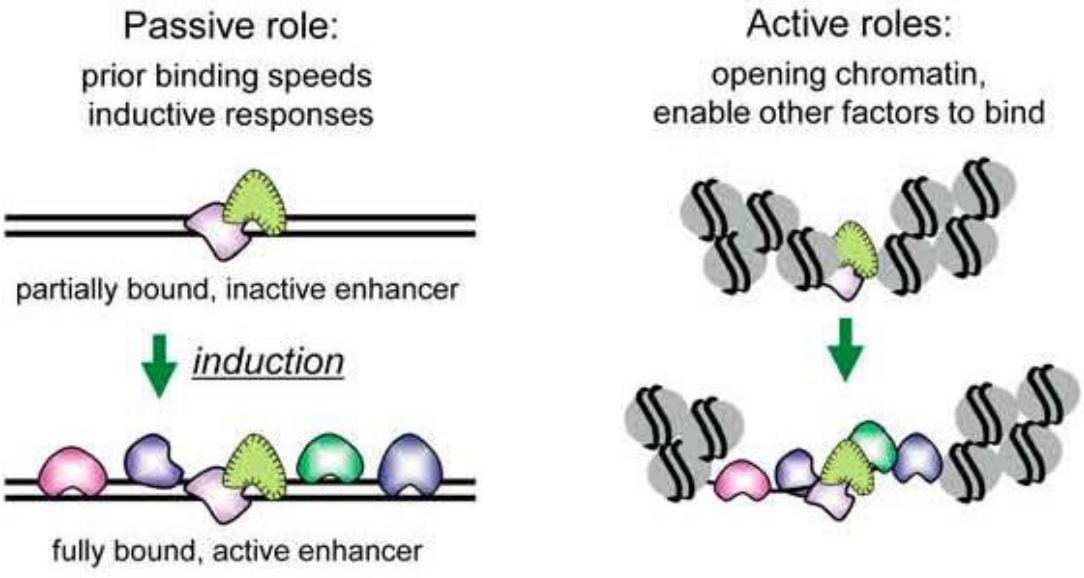

Figure 28 : les deux modes d'action des FTs pionniers. Un mode passif dans lequel le FT pionnier est préalablement lié à son site de fixation. Il constituera un ancrage pour la fixation d'autres FTs lorsque la régulation des gènes concernés sera requise. Et un mode actif selon lequel la fixation du FT pionnier facilite directement la relaxation de la chromatine ou permettre à d'autres FTs de se lier à leur site de fixation (Zaret & Carroll, 2011).

aux histones, stabilisant la fixation du FT (Cirillo *et al.*, 2002), ainsi qu'un domaine d'activation de la transcription (Pani *et al.*, 1992). Cette reconnaissance du site de fixation situé sur une portion de la chromatine, pourtant inaccessible pour la majorité des FTs, est également facilitée par la capacité des FTs pionniers à reconnaitre une séquence réduite de la séquence caractérisant leur site de fixation (Soufi *et al.*, 2015). Une fois fixés, le rôle de ces FTs pionniers peut revêtir deux aspects (Figure 28). Un mode passif dans lequel leur fixation préalable à toute régulation de la transcription constituera un ancrage pour la fixation d'autres FTs lorsque la régulation des gènes concernés sera requise. Cet ancrage préalable permet une réponse plus rapide de la cellule, notamment pendant le développement et dans la régulation hormonale chez les animaux (Gualdi *et al.*, 1996 ; Carroll *et al.*, 2005).

Mais les FTs pionniers régulent également l'expression des gènes de manière active en permettant la relaxation de la chromatine par leur seule liaison à leur site de fixation (Cirillo *et al.*, 2002). Ils peuvent également permettre à d'autres FTs de se lier à l'ADN (Cirillo & Zaret, 1999 ; Watts *et al.*, 2011), recruter directement l'ARN polymérase II (Hsu *et al.*, 2015), ou encore interagir avec des protéines intervenant dans la modification de la structure de la chromatine (Li *et al.*, 2012). Ce dernier exemple fait intervenir des protéines jouant un rôle crucial dans la régulation de l'expression des gènes grâce aux réactions enzymatiques qu'elles catalysent, lesquelles modifient la compaction de la chromatine. Parmi celles-ci, on retrouve deux sortes de mécanismes. (i) Un remodelage de la chromatine ATP-dépendant provoqué par des complexes protéiques déplaçant les nucléosomes et les histones au sein de la chromatine (Ho & Crabtree, 2010). (ii) Des modifications post-traductionnelles des histones modulant la compaction (Kouzarides, 2007) (Figure 29). La queue N-terminale des histones est en effet sujette à de nombreuses modifications post-traductionnelles, telles des acétylations, méthylations, phosphorylations, sumoylation et ubiquitinations (Roh *et al.*, 2005 ; Zhu *et al.*, 2005 ; Ge *et al.*, 2006 ; Nathan *et al.*, 2006 ; Wang *et al.*, 2006 ; Li *et al.*, 2007). Mais ces enzymes peuvent également intervenir dans la régulation de l'activité des FTs eux-mêmes via une acétylation (Kouzarides, 2000), une phosphorylation (Brivanlou & Darnell, 2002), ou encore une sumoylation ou une ubiquitination (Conaway *et al.*, 2002 ; Gill, 2005). Il s'agit donc d'une action de ces enzymes combinée à celle des FTs au niveau des séquences promotrices des gènes qui permet le recrutement de l'ARN polymérase II et l'initiation de la transcription.

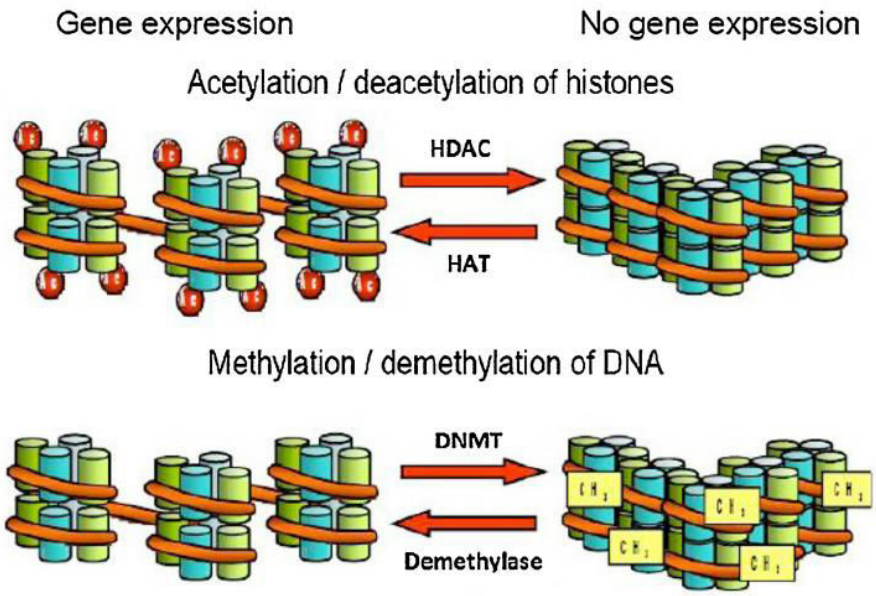

Figure 29 : illustration représentant les modifications post-traductionnelles des histones impactant la condensation de la chromatine. Une chromatine active, permettant l'expression des gènes via la fixation des différents acteurs de la régulation de la transcription, est caractérisée par la présence d'acétylations. Au contraire, une chromatine inactive est condensée et caractérisée par la présence de méthylations. HDAC : Histone Deacetylase, HAT : Histone acetyl-transferase, DNMT : DNA Methyl-transferase. (Vandermeers *et al.*, 2013)

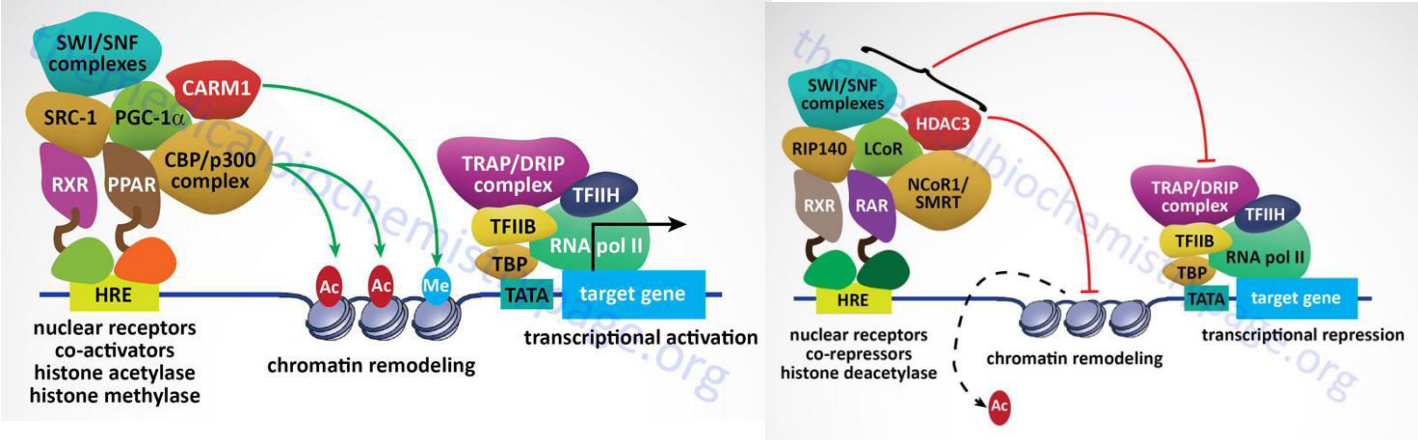

Figure 30 : exemple de la régulation de l'expression d'un gène par des récepteur nucléaires non stéroïdiens. Ces FTs sont actifs sous forme d'hétérodimères et se fixent sur leur site de fixation (HRE). Ils recrutent ensuite différentes protéines partenaires en fonction de la nature du dimère formé. A) les protéines partenaires recrutées ont une action activatrice de la transcription, modifiant la structure de la chromatine et permettant la formation du CPI. B) les protéines partenaires recrutées ont une action inhibitrice de la transcription. La chromatine est maintenue compactée et la formation du CPI est inhibée.
Source : http://themedicalbiochemistrypage.org/gene-regulation.php

Au niveau des promoteurs, les sites de fixation des FTs sont regroupés en modules permettant une action combinée des FTs dans la régulation de l'expression des gènes (Reményi *et al.*, 2004) que ce soit depuis un promoteur proximal ou distal (Lemon & Tjian, 2000). L'espacement des sites de fixations au sein des promoteurs peut également être un paramètre crucial dans l'expression des gènes afin que le recrutement des protéines partenaires soit possible (Ng *et al.*, 2014). Le fait que les FTs puissent reconnaître une séquence dégénérée de leur site de fixation (Wunderlich & Mirny, 2009) permet également une variation de l'association de FTs au niveau d'un promoteur, puisque plusieurs FTs peuvent entrer en compétition pour un même site de fixation (Zabet & Adryan, 2013). Cette complexité est encore accrue par la capacité de certains FTs à former des homo- ou hétéro-multimères (Amoutzias *et al.*, 2008), tels les bHLH (Taelman *et al.*, 2004), les bZIP (Amoutzias *et al.*, 2007), les HSF (Yamamoto *et al.*, 2009b) ou encore les MADS-box (Puranik *et al.*, 2014). Chaque hétéro-dimère ciblant un site de fixation différent. Enfin, la petite taille des sites de fixation des FTs ainsi que la reconnaissance de séquences dégénérées implique une occurrence élevée de sites de fixation potentiels au sein du génome. Cependant, la plupart de ces séquences ne sont pas biologiquement fonctionnelles puisque peu sont liées à un FT *in vivo*, la compaction de la chromatine jouant un rôle régulateur quant à l'accession de ces sites potentiels (Guertin & Lis, 2010). Il apparait donc de plus en plus clair que la régulation de la transcription fait intervenir une combinaison de plusieurs FTs agissant de manière synergique (Lemon & Tjian, 2000). De plus, cette combinaison de FTs étant propre à chaque promoteur, et chaque combinaison de FTs permettant de recruter une grande diversité de protéines partenaires (co-activateurs ou co-répresseurs), un même FT peut activer la transcription d'un gène et réprimer celle d'un autre (Figure 30) (Ma, 2005). Cette action combinée d'une poignée de FTs pouvant réguler l'expression d'un grand nombre de gènes en recrutant diverses protéines partenaires permet une fine régulation de la réponse cellulaire (Brkljacic & Grotewold, 2016).

Outre l'état de la chromatine, le contexte spatial au sein du noyau est également à prendre en compte. En effet, le noyau est organisé, chaque chromosome occupant un espace particulier appelé territoire (Figure 31). Cette conformation permet à des gènes éloignés au sein et entre chromosomes de se retrouver spatialement proches dans le noyau via la formation de boucles (Fraser & Bickmore, 2007). Il existe également au sein du noyau des régions présentant une activité transcriptionnelle élevée (Iborra *et al.*, 1996). Ces régions regroupent une concentration

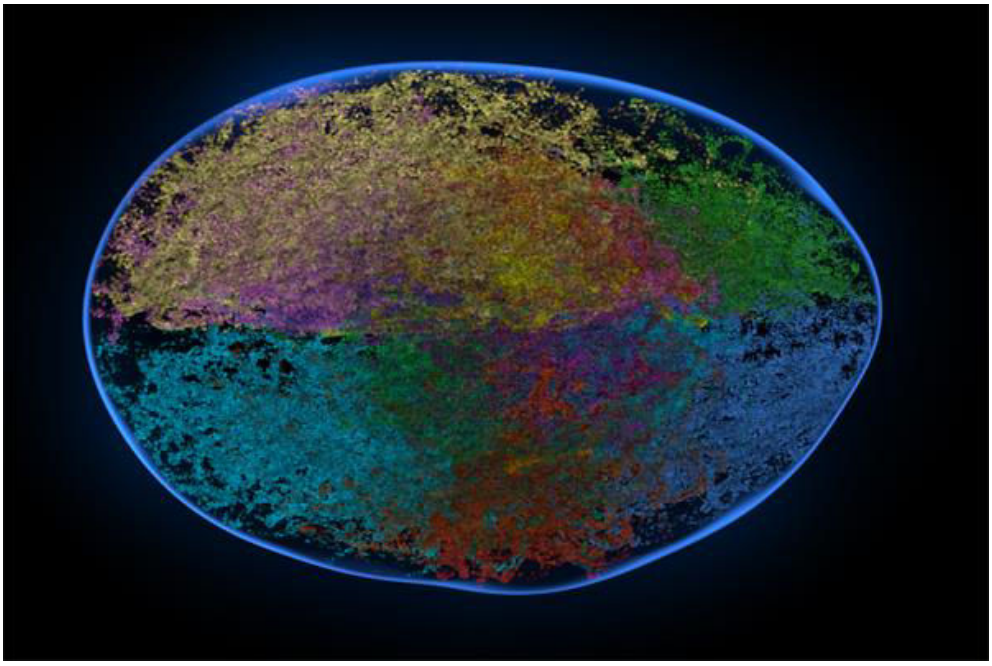

Figure 31 : représentation des territoires chromosomiques. Illustration 3D de la chromatine à l'interphase. Chaque chromosome est représenté d'une couleur différente, mettant en évidence leur territoire chromosomique. Source : http://medical-render.com/content/chromosome-territories

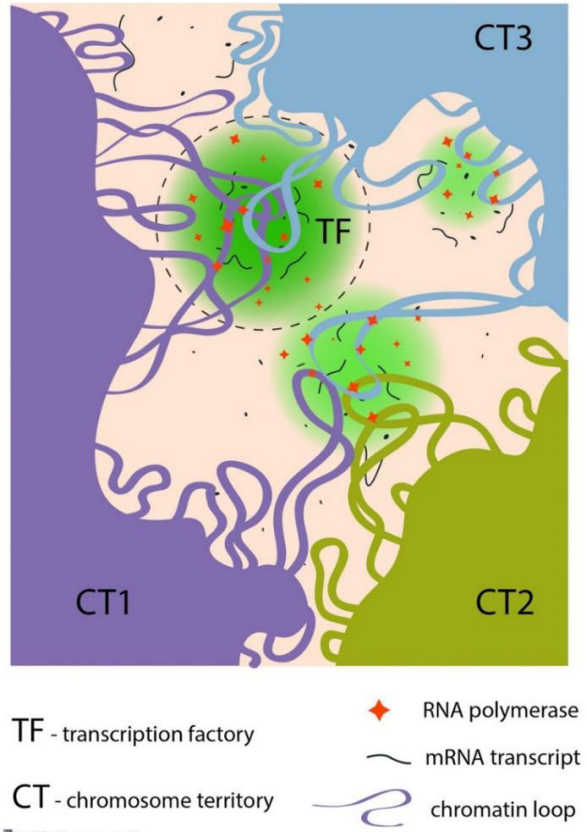

Figure 32 : schéma d'une « transcription factory » située entre trois territoires chromosomiques. Les boucles chromosomiques formées à la périphérie de chaque territoire chromosomique partagent la même « transcription factory ».
Source : https://www.mechanobio.info/topics/genome-regulation/dna-templated-transcription/

d'ARN polymérases et de facteurs associés suffisants pour permettre la transcription de plusieurs gènes (Cook, 1999) et sont appelées « transcription factories » (Figure 32). L'organisation du noyau en territoires chromosomiques permet donc à des gènes éloignés au sein du génome de se retrouver proches dans le noyau et d'être sous l'influence d'une même usine à transcription, expliquant ainsi leur co-régulation.

VIII. L'importance des FTs dans l'histoire évolutive des organismes

De par leur fonction clé dans l'expression des gènes et leur régulation, les FTs jouent un rôle déterminant dans l'évolution des organismes. L'histoire évolutive des eucaryotes est en effet ponctuée de processus tels des évènements de duplications (Wendel, 2000 ; Paterson *et al.*, 2006 ; Edger & Pires, 2009) ou encore le brassage et l'accrétion de domaines permettant des modifications entrainant l'apparition de nouvelles familles de FTs (Riechmann *et al.*, 2000 ; Lang *et al.*, 2010).

Le mouvement de domaines protéiques au sein d'une protéine et entre protéines (de Château & Björck, 1994 ; Patthy, 1996) ainsi que leur influence sur l'évolution des protéines, et notamment des FTs, n'est plus à démontrer, aussi bien chez les animaux, les plantes que les virus (Morgenstern & Atchley, 1999 ; Iyer *et al.*, 2002 ; Kawashima *et al.*, 2009 ; Carretero-Paulet *et al.*, 2010). De tels phénomènes de brassage de domaines de liaison à l'ADN peuvent provoquer l'émergence de nouveaux FTs comme, par exemple, l'association d'un homéo-domaine et d'un leucine zipper caractéristique de la famille des HD-ZIP retrouvée uniquement chez les plantes terrestres (Riechmann *et al.*, 2000 ; Lang *et al.*, 2010 ; Sharma *et al.*, 2013). Un autre exemple est celui des aureochromes que l'on retrouve uniquement chez les straménopiles photosynthétiques (Takahashi *et al.*, 2007 ; Ishikawa *et al.*, 2009 ; Vieler *et al.*, 2012 ; Schellenberger Costa *et al.*, 2013). Ces récepteurs de la lumière bleue sont caractérisés par l'association d'un domaine basique-leucine-Zipper (bZIP) responsable de la liaison à l'ADN et d'un domaine senseur LOV (Light, Oxygene, Voltage) appartenant à la superfamille de domaines PAS (Per-ARNT-Sim). Lors d'une irradiation par de la lumière bleue, le domaine LOV se lie à un FAD (flavin adenine dinucleotide) ou un FMN (flavin mononucleotide) qui joue le rôle de chromophore (Salomon *et al.*, 2000). Dans l'environnement marin, les longueurs d'onde autres

que celles de la lumière bleue sont absorbées, ces dernières étant les seules à parcourir de longues distances dans la colonne d'eau (Austin & Petzold, 1986). La lumière bleue est donc supposée jouer un rôle important chez les microalgues, comme le suggère l'implication des auréochromes dans des mécanismes clés tels le cycle cellulaire (Huysman *et al.*, 2013). Ces FTs confèrent donc aux straménopiles photosynthétiques un avantage adaptatif important pour survivre dans un environnement marin. Ces brassages de domaines aboutissent donc à de nouvelles combinaisons de domaines, lesquelles permettent une innovation en terme d'interactions, tant avec l'ADN (élements *cis*) qu'avec des protéines et notamment des régulateurs de la transcription. Ces nouveaux réseaux d'interactions permettent l'émergence de nouveaux mécanismes de régulation (Koonin & Galperin, 2003) et participent donc à l'évolution des organismes (Nowick & Stubbs, 2010). Cette complexification des mécanismes de régulation est également influencée par des évènements de duplication de gènes (Arsovski *et al.*, 2015). Dans le cas d'une duplication de gènes, les deux duplicats ont 3 devenirs possibles (Figure 33) : (1) Les deux duplicats conservent la même fonction (sub-fonctionnalisation) ; (2) l'un des deux duplicats acquiert une nouvelle fonction tandis que l'autre conserve la fonction ancestrale (neo-fonctionnalisation) ; (3) l'un des deux duplicats devient non fonctionnel et dégénère en pseudogène (Lynch & Conery, 2000). Peu de cas de néo-fonctionnalisation sont prouvés puisque dans le cas d'anciens paralogues, le gène ancestral est difficile à mettre en évidence. Cependant, certaines hypothèses ont été émises, notamment concernant les FTs de la famille des MYB impliqués dans le développement des trichomes chez *A. thaliana*. En effet, ceux-ci seraient issus d'une duplication des gènes MYB impliqués dans la biosynthèse des anthocyanes, suivie d'une divergence évolutive (Serna & Martin, 2006). Un tel exemple a également été mis en évidence chez le champignon *Candida albicans* chez qui des FTs apparus par duplication ont ensuite divergés d'un point de vue de leur spécificité de liaison à l'ADN et/ou d'interactions avec des régulateurs de la transcription (Perez *et al.*, 2014). Ces modifications aboutissant à une régulation de l'expression de groupes de gènes différents permettant d'améliorer la prolifération de *C. albicans* chez son hôte. Les changements suivant une duplication peuvent également provoquer une modification du niveau d'expression d'un duplicat. Cette différence d'expression, tant au niveau spatial qu'au niveau quantitatif, peut contribuer à une néo-fonctionnalisation (Kim *et al.*, 2012 ; Sakuma *et al.*, 2013).

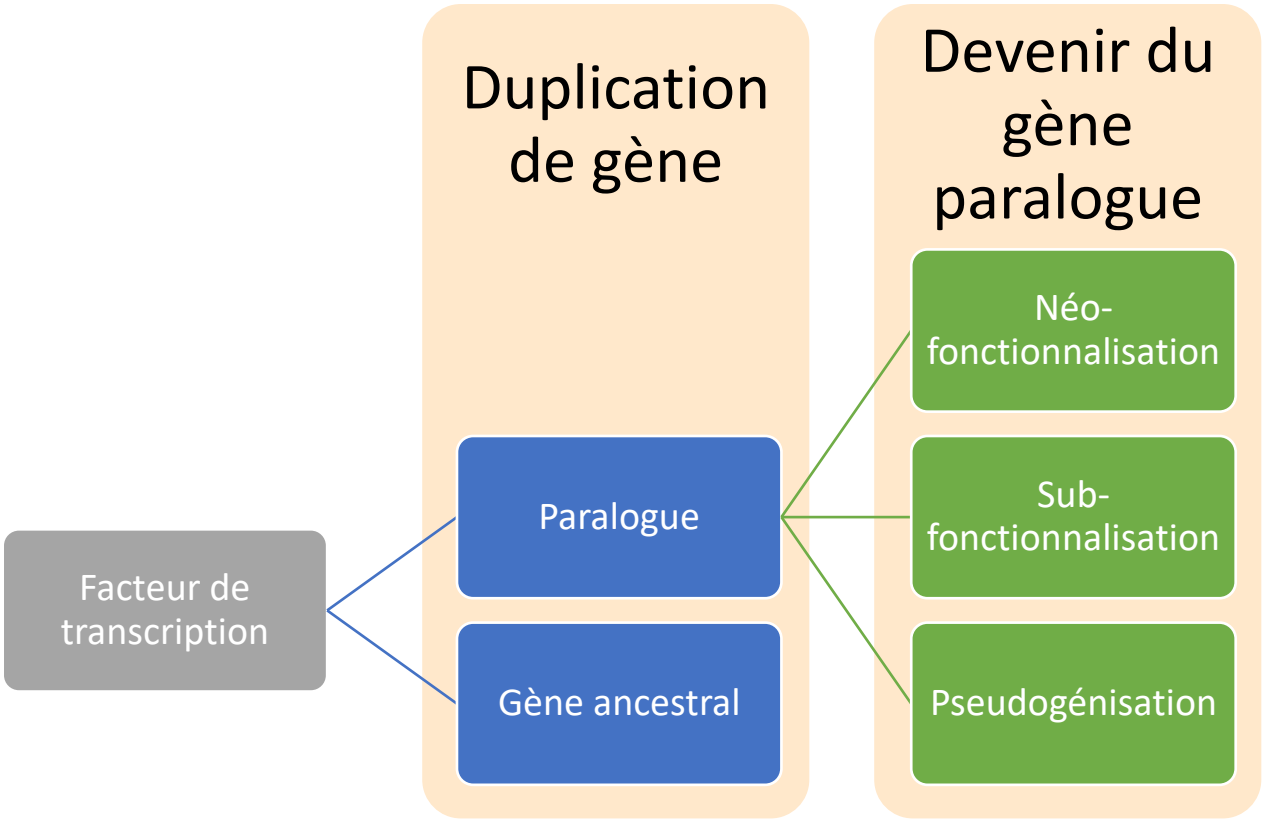

Figure 33 : les trois devenirs possibles des paralogues suite à une duplication de gène. Néo-fonctionnalisation : le gène paralogue acquiert une nouvelle fonction bénéfique à l'organisme et est retenue par sélection naturelle alors que le gène ancestral conserve la fonction originale. Sub-fonctionnalisation : les deux copies du gène conserve la fonction d'origine. Pseudogénisation : accumulation de mutations dégénératives, le gène paralogue perd toute fonction

Des duplications de portions de génome ou de génomes entiers ont également abouti à une expansion de certaines familles de FTs tels les MADS-box, les MYB ou encore les bHLH chez les plantes (Becker *et al.*, 2000 ; Feller *et al.*, 2011). Pertes et expansions de familles de FTs jalonnent l'histoire évolutive des plantes et des algues (Lang *et al.*, 2010). De telles expansions sont toutefois plus fréquentes dans le génome des plantes que dans celui des animaux ou des champignons (Shiu *et al.*, 2005). Ces expansions sont non seulement dues à un taux de duplications plus élevés chez les plantes (Wendel, 2000), mais également à un degré d'expansion plus élevé des familles de FTs par rapport aux autres gènes. C'est-à-dire qu'un FT dupliqué sera plus facilement conservé qu'un autre gène. De plus, des groupes de gènes orthologues entre *A. thaliana* et *O. sativa* codant des FTs présentent des expansions parallèles (une fois encore, plus que les autres gènes) (Shiu *et al.*, 2005). Ces conservations de FTs dupliqués pourraient s'expliquer par les pressions environnementales que subissent les plantes et auxquelles elles doivent faire face du fait de leur mode de vie sessile. Une favorisation de la duplication ou de la rétention des FTs après duplication (Shiu *et al.*, 2005 ; Hanada *et al.*, 2008), ajoutée aux mécanismes évoqués précédemment (néo-fonctionnalisation, brassage et accrétion de domaines) entrainent une complexification des réseaux de régulation des gènes, permettant ainsi une meilleure réponse aux stress subis par les plantes (Blanc & Wolfe, 2004). Ce rôle crucial des FTs chez les plantes peut aussi expliquer que 13% des protéines spécifiques des plantes identifiées dans le génome de *A. thaliana* sont des FTs, constituant ainsi le groupe fonctionnel le plus représenté (Gutiérrez *et al.*, 2004). De plus, ces FTs appartiennent à des familles liées à des fonctionnements particuliers tels les NAC, WRKY ou AP2/ERF (réponse aux stress, développement …), permettant aux plantes de mieux faire face et s'adapter aux changements de leur environnement.

De nombreuses études réalisent donc des analyses comparatives portant sur différentes espèces et/ou groupes taxonomiques et appréhendent ainsi un aspect évolutif par l'intermédiaire des FTs (Lang *et al.*, 2010 ; Charoensawan *et al.*, 2010a ; Sharma *et al.*, 2013). Ce type d'étude permet de mettre en évidence certaines particularités spécifiques à une espèce ou un taxon. Chez *A. thaliana*, 45% des FTs identifiés appartiennent à des familles identifiées uniquement chez les plantes (Riechmann *et al.*, 2000). Enfin, une caractéristique supplémentaire des plantes terrestres concerne l'expansion de la famille des MYB. En effet, 190 séquences sont observées chez *A. thaliana* contre 6 et 10 séquences chez *Drosophilia melanogaster* et *Saccharomyces cerevisiae* respectivement (Feller *et al.*, 2011). Un autre exemple est celui des Heat Shock Factors (HSFs)

qui sont en proportion plus importante parmi les FTs des deux diatomées *P. tricornutum* et *Thalassiosira pseudonana* (33% et 36% respectivement) par rapport aux autres stramenopiles (de 14% chez *Aureococcus anophagefferens* à 5% chez *Phytophtora ramorum*) (Rayko *et al.*, 2010). Dans le cas de *Caenorhabditis elegans*, les FTs appartenant à la famille des récepteurs à l'acide rétinoïque sont au nombre de 239 alors que chez les autres animaux il n'en est dénombré au maximum que 19 chez *Tetraodon nigroviridis* (Zhang *et al.*, 2012).

Dans le monde des microalgues, de telles études comparatives portant sur les FTs ont été réalisées chez les straménopiles (Rayko *et al.*, 2010) ainsi que chez les algues vertes et rouges dans le but de déterminer leur place dans l'histoire évolutive des organismes photosynthétiques (Lang *et al.*, 2010 ; Sharma *et al.*, 2013). Bien que les haptophytes soient un groupe majeur dans l'histoire évolutive des organismes photosynthétiques (Burki *et al.*, 2012), leur intégration dans de telles études est limitée du fait du manque de données génomiques disponibles. En effet, seuls deux génomes d'haptophyte ont été séquencés, celui *d'E. huxleyi* (Read *et al.*, 2013) et, très récemment celui de *Chrysochromulina tobin* (Hovde *et al.*, 2015). Combler ce manque ainsi que ceux concernant d'autres lignées de microalgues (chlorarachniophytes, cryptophytes, glaucophytes, euglènes, rhodophytes) permettrait de mieux comprendre l'histoire évolutive de ces organismes au travers des gains et pertes de familles de FTs ou de la présence de séquences spécifiques de lignées. Cette histoire évolutive est en effet ponctuée de processus tels que des évènements de duplication (Wendel, 2000 ; Paterson *et al.*, 2006 ; Edger & Pires, 2009) ou encore le brassage et l'accrétion de domaines permettant des modifications entrainant l'apparition de nouvelles familles de FTs (Riechmann *et al.*, 2000 ; Lang *et al.*, 2010).

IX. Les FTs comme cibles moléculaires pour l'orientation métabolique

Les FTs ayant un rôle déterminant dans le développement et l'acclimatation métabolique des organismes, ils constituent des cibles de choix en vue d'applications biotechnologiques. Dans une optique de production d'un composé d'intérêt, des approches d'orientation métabolique et de modulation de l'expression des gènes par génie génétique ciblant une ou plusieurs enzymes sont abordées (voir plus haut, partie « mieux comprendre le métabolisme et sa régulation pour optimiser la production de composés d'intérêt »). Toutefois, du fait de la complexité de

l'acclimatation métabolique induite par le stress et des connexions entre les différentes voies qui la composent, ces approches sont encore complexes à maitriser. Une alternative intéressante, et proposée depuis une dizaine d'années chez les plantes (Capell & Christou, 2004), consiste à cibler, non pas les enzymes impliquées qui seraient trop nombreuses, mais des facteurs de transcription (FTs). Certaines études s'attachent donc à ajouter des FTs, ou à sur- ou sous-exprimer des FTs déjà présents dans le génome de l'organisme étudié (Ibáñez-Salazar *et al.*, 2014 ; Kang *et al.*, 2014 ; Zhang *et al.*, 2014b). La modification de l'expression d'un FT influence celle de nombreux gènes. De fait, l'orientation du métabolisme vers la production d'un composé d'intérêt peut être atteinte lorsque ce FT est choisi avec soin. De plus, cette approche peut permettre d'induire les effets recherchés qui seraient produits par l'application du stress, tout en en évitant les effets négatifs sur la croissance de la microalgue.

Par exemple, un FT de soja de la famille des Dof impliqué dans la production de lipides, a été transféré dans le génome de la microalgue *Chlorella ellipsoidea* (Zhang *et al.*, 2014a). Cette transformation a abouti à la modification de l'expression de 1046 gènes et à une augmentation de 52 % de la production de lipides par la microalgue. Un FT de soja de cette même famille des Dof a également été transféré dans le génome de *C. reinhardtii*, aboutissant à un doublement de la production de lipides totaux (Ibanez-Salazar *et al.*, 2014). Enfin, très récemment, la surexpression d'un des FTs de la famille bHLH de la microlague *Nannochloropsis salina* a permis d'augmenter le taux de croissance, l'assimilation des nutriments et la production d'esters méthyliques d'acides gras (Kang *et al.*, 2015).

Néanmoins, choisir judicieusement le(s) FT(s) ciblé(s) requiert une connaissance suffisante des mécanismes impliqués dans la production de lipides en condition de stress azoté, depuis la régulation de la transcription par les FTs jusqu'aux réorientations métaboliques impliquées. Un certain nombre d'études moléculaires a été mené dans cette optique chez les microalgues et différentes approches ont été abordées au fil du temps. La plupart de ces études concernent les microalgues vertes qui restent à ce jour les mieux connues et les plus étudiées (Sanz-Luque *et al.*, 2015 ; Goncalves *et al.*, 2016). Concernant le métabolisme de l'azote, une étude s'est attachée à identifier de potentiels sites de fixation de FTs dans le promoteur de la nitrate réductase de la microalgue verte *Chlorella vulgaris* (Cannons & Shiflett, 2001). L'implication potentielle de FTs de la famille GATA a ainsi été mise en évidence. Une autre étude, menée chez la microalgue rouge

Cyanidioschizon merolae, a identifié un FT de la famille des MYB dont l'induction était spécifiquement liée à une déplétion azotée. De plus, cette expression était en corrélation avec celle de certains gènes clés de l'assimilation de l'azote (Imamura *et al.*, 2009).

Ensuite, avec l'avènement des techniques dites « –omiques », des études à l'échelle d'un transcriptome, d'un protéome ou d'un génome entier ont vu le jour dans le but d'analyser la régulation de la production lipidique des microalgues en conditions de stress azoté, et ainsi identifier les réseaux de régulation qui contrôlent cette production. Des approches génomiques ont été abordées. Bénéficiant du génome séquencé de l'eustigmatophycée *Nannochloropsis oceanica*, Hu *et al.*, (2014) ont ainsi identifié des motifs conservés dans les promoteurs de gènes impliqués dans le métabolisme lipidique chez six souches de *Nannochloropsis sp.*, et ont ainsi construit un réseau de régulation de cette voie de biosynthèse. Quant aux études fondées sur l'analyse du transcriptome ou du protéome, elles identifient principalement des FTs sur- ou sous-exprimés (souvent nombreux) en fonction des différentes conditions physiologiques, mais sans identifier plus précisément le lien de ces FTs avec les autres gènes ou les phénotypes observés (Miller *et al.*, 2010 ; Lv *et al.*, 2013 ; Guarnieri *et al.*, 2013). Bien qu'étant des candidats intéressants, ces FTs requièrent une caractérisation ultérieure afin de mieux comprendre leur rôle. Certaines études s'attachent d'ailleurs à identifier un FT candidat particulier et à caractériser plus précisément sa fonction. Ainsi, Matthijs *et al.*, (2016) ont identifié un FT appelé RGQ1 chez la diatomée modèle *P. tricornutum*, appartenant à une famille de FTs non décrite jusque-là. L'analyse de données transcriptomiques a permis d'identifier des gènes surexprimés en réponse à un stress azoté chez cette diatomée. Deux motifs conservés correspondant à des sites de fixation de FTs potentiels ont été identifiés dans les séquences promotrices de ces gènes. L'analyse simple hybride chez la levure a finalement permis de montrer que l'un de ces deux motifs était spécifiquement reconnu par ce nouveau FT, RGQ1, démontrant ainsi son implication dans la réponse à une privation d'azote. Une autre étude, cette fois chez la microalgue verte *Chlorella* (Goncalves *et al.*, 2016), a montré, par une analyse protéomique confirmée par une approche de complémentation de mutant, que le FT ROC40 de la famille des MYB-related (connu pour son implication dans le cycle cellulaire) était impliqué dans la production de lipides en condition de privation d'azote.

Toutefois, ces études ne sélectionnent leurs candidats qu'à partir de données transcriptomiques ou protéomiques, et les mécanismes liés à ces FTs ne sont pas précisément identifiés. Or,

l'acclimatation d'une cellule en réponse à un stimulus perçu requiert l'intervention de nombreux acteurs agissant de concert (Brivanlou & Darnell, 2002). La compréhension des mécanismes sous-jacents passe donc par l'intégration de différents types de données (Kitano, 2002 ; Nurse, 2003) prenant notamment en compte les FTs, leurs interactions entre eux et avec les autres régulateurs de la transcription, et l'identification de leurs gènes cibles. L'identification de ces réseaux de gènes connectés les uns aux autres dans un contexte physiologique et cellulaire particulier est, en effet, une question de plus en plus étudiée par les scientifiques (Schadt, 2009 ; Serin *et al.*, 2016). Cette question peut être abordée par l'étude du profil d'expression de l'ensemble des gènes d'une cellule dans différentes conditions, groupant entre eux les profils les plus proches. Les réseaux ainsi construits sont appelés réseaux de co-expression des gènes et comportent une structure modulaire (Carter *et al.*, 2004a). En effet, les gènes sont liés les uns aux autres en fonction de la significativité de leur co-expression, les plus proches étant groupés en modules de gènes co-exprimés (Figure 34). L'avantage de ces modules est qu'ils groupent entre eux des gènes utilisés par la cellule selon une même dynamique en fonction des différentes conditions étudiées. Ces gènes sont souvent impliqués dans un même type de réponse (Aoki *et al.*, 2007) et permettent de mieux comprendre les mécanismes impliqués dans un processus ou une réponse particulière de la cellule (Kogelman *et al.*, 2014 ; Hollender *et al.*, 2014 ; El-Sharkawy *et al.*, 2015). De plus, chaque module étant composé de gènes plus ou moins liés entre eux en fonction de la significativité statistique de leur co-expression, certains sont d'avantage liés aux autres gènes du module que les autres. Ces gènes sont appelés gènes « hub » (Figure 35). Cette position centrale joue un rôle déterminant dans la fonction cellulaire du module (Jeong *et al.*, 2001 ; Carter *et al.*, 2004a ; Cooper *et al.*, 2006), faisant de ces gènes des candidats particulièrement intéressants. La construction et l'étude de ces réseaux de co-expression permettent donc d'identifier les gènes et groupes de gènes impliqués dans la réponse d'une cellule à un stimulus particulier.

Afin de comprendre plus pleinement les mécanismes impliqués dans cette réponse, un autre type de réseau peut être construit. Il s'agit de réseaux de régulation des gènes qui, contrairement aux réseaux de co-expression, ne lient que les FTs à leurs gènes cibles (Figure 36). La construction de ces réseaux requiert donc l'identification des gènes dont l'expression est régulée par les FTs de l'organisme étudié (Franco-Zorrilla & Solano, 2017).

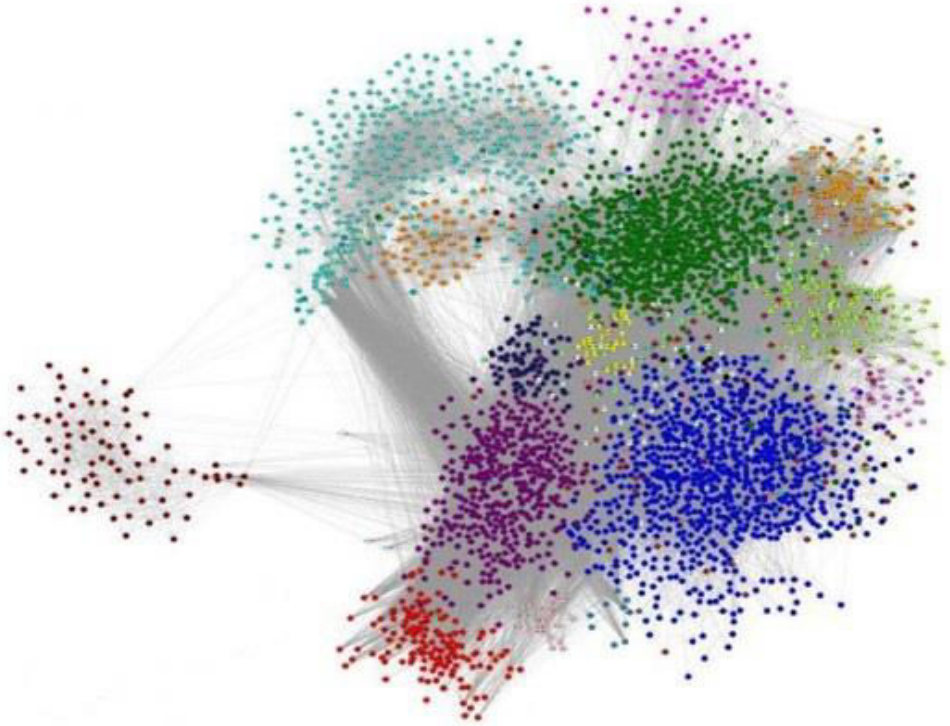

Figure 34 : représentation d'un réseau de co-expression des gènes mettant en évidence sa structure modulaire. Chaque module est représenté par une couleur (Bunyavanich *et al.*, 2014).

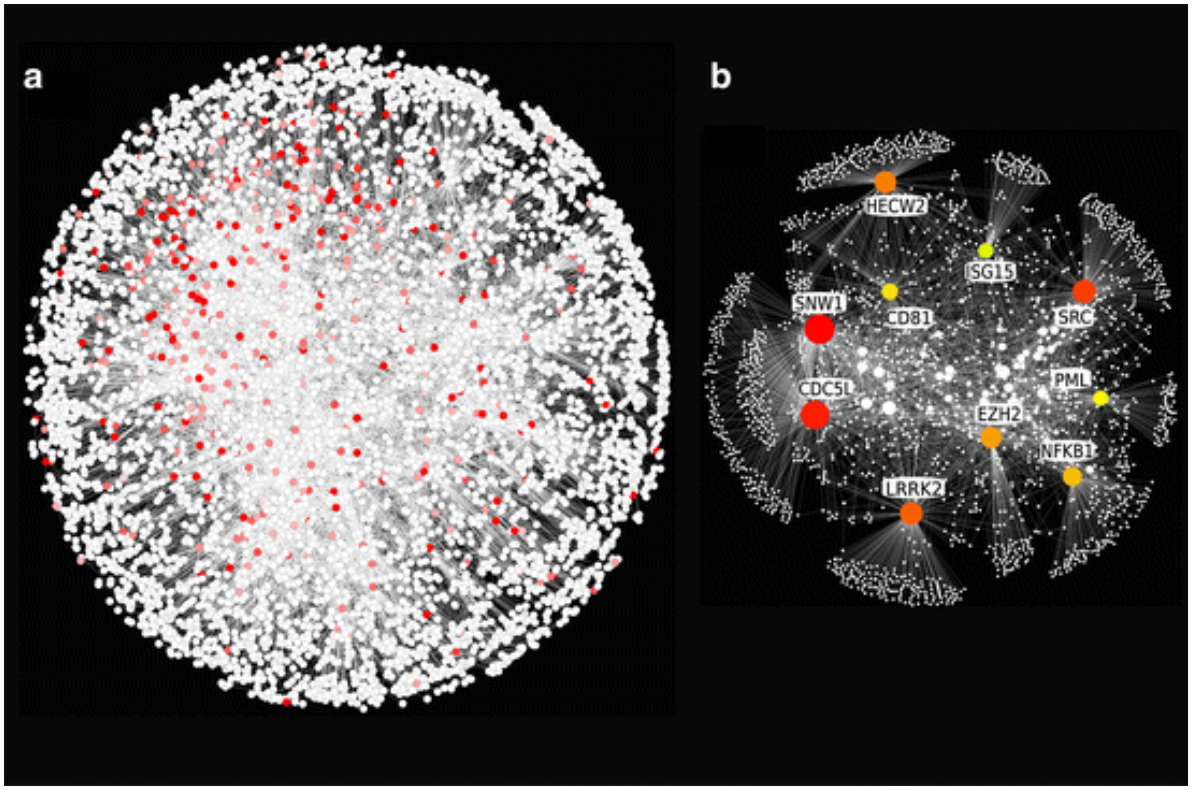

Figure 35 : illustration de la notion de gène hub. Chaque cercle représente un gène et l'ensemble du réseau de gène est représenté en a). Au sein de ce réseau, certains gènes comportent beaucoup plus de connections que les autres et son appelés gènes hub. En b), l'ensemble du réseau a été réduit aux dix gènes hub et aux gènes auxquels ils sont connectés. (Charitou *et al.*, 2016)

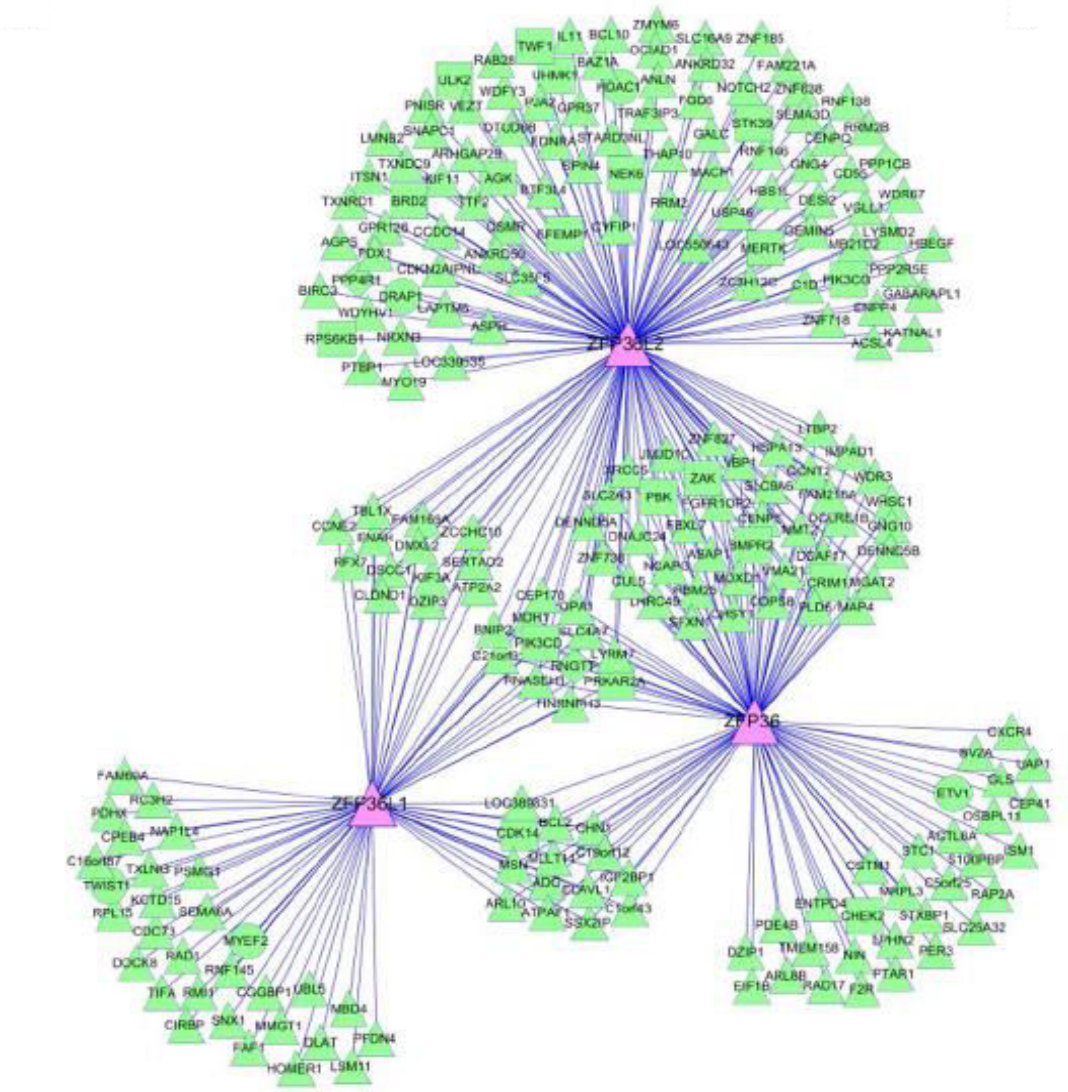

Figure 36 : représentation d'un réseau de régulation des gènes. Chaque lien représente une interaction de régulation entre le FT (en rose) et sa cible (en vert). (Zekavati *et al.*, 2014)

L'identification de ces gènes cibles peut se faire de deux manières : (i) à partir de données d'expression de gènes, lesquelles sont devenues de plus en plus accessibles avec l'avènement des technologies dites « omiques ». Pour cela, on part du postulat qu'un gène ayant les mêmes variations d'expression qu'un FT est régulé par celui-ci. Ou (ii) à partir de données d'interactions physiques grâce à des techniques de retard de migration sur gel, de levure ou bactérie simple hybride, d'immunoprécipitation de la chromatine (Taylor-Teeples *et al.*, 2015; Liu *et al.*, 2015), d'analyse de mutant provenant de banque de mutant ou généré par transformation génétique (Yang & Wu, 2012) ou encore par prédiction bioinformatique de site de fixation de FTs dans le promoteur des gènes quand le génome de l'organisme est disponible (Hu *et al.*, 2014).

Coupler ces deux approches permet d'identifier les gènes impliqués dans les remaniements métaboliques constituant la réponse de l'organisme, et les réseaux de régulation qui les orchestrent. Dans le cadre plus précis d'une identification des FTs et des réseaux de régulations associés à une production de lipides provoquée par un stress azoté chez une microalgue, seules quatre études ont été menées à ma connaissance (et moins de dix en élargissant la recherche à d'autres applications que la production de lipides), toutes les quatre chez la microalgue verte *C. reinhardtii*. En 2014, deux études (Valledor *et al.*, 2014 ; Schmollinger *et al.*, 2014) ont intégré des données transcriptomiques, protéomiques et métabolomiques mais se sont plutôt attachées à identifier (de manière très complète) les mécanismes métaboliques et de signalisation impliqués dans la réponse de l'algue à un stress azoté. Quelques FTs candidats ont également été identifiés à partir de leur expression différentielle. Deux études publiées l'année suivante ont, pour leur part, identifié des réseaux de régulation orchestrant la réponse métabolique de *C. reinhardtii* face à une carence azotée. Gargouri *et al.* (2015), en utilisant des données transcriptomiques et métaboliques, ont construit deux réseaux de régulations : (i) le premier à partir de corrélation entre l'expression des régulateurs de la transcription (RTs : FTs et régulateurs de la structure de la chromatine) et celle des gènes du métabolisme de l'algue, (ii) le deuxième à partir de corrélations entre l'expression des RTs et la quantification des métabolites primaires identifiés. En recoupant ces deux réseaux, ils ont identifié un réseau cœur pour lequel 70 RTs étaient corrélés à la fois à l'expression des gènes codant les enzymes du métabolisme et à la quantité des métabolites correspondants aux réactions catalysées par ces enzymes. A partir des différentes corrélations (RTs *vs* métabolites et RTs *vs* enzymes du métabolisme) ils ont identifié des RTs impliqués dans des mécanismes clés de la réponse de l'algue tels la production de lipides, de

carbohydrates, la photosynthèse, le métabolisme central ou encore l'assimilation de l'azote. Ils ont ensuite identifié des FTs hub (FTs centraux, ayant le plus de cibles parmi les autres gènes au sein du réseau et jouant ainsi un rôle clé dans l'orchestration de la réponse de l'organisme) assurant le développement du programme de régulation de la réponse de *C. reinhardtii* face à une carence azotée.

Enfin, López García de Lomana *et al.* (2015) ont construit un réseau de régulation des gènes de *C. reinhardtii* en condition de carence azotée, à partir des données transcriptomiques publiées par Boyle *et al.* (2012). Ce réseau de co-expression a mis en évidence au moins 17 RTs coordonnants 815 gènes, et a identifié la cinétique selon laquelle ces RTs intervenaient dans la réponse de l'algue. Puis, afin d'identifier plus précisément l'impact de ce réseau de régulation sur la production de triglycérides et de biomasse en réponse à la carence azotée, celui-ci a été intégré à une analyse métabolique en utilisant le réseau métabolique de *C. reinhardtii* publié en 2011 (Chang *et al.*, 2011). L'annotation fonctionnelle des gènes avec lesquels ces FTs étaient corrélés a permis de les lier à certaines fonctions clés telles la croissance, la photosynthèse, la production de lipides ou la réponse au stress oxydatif.

Ces études constituent une avancée importante vers la compréhension des mécanismes mis en place par les microalgues pour faire face au stress azoté, ainsi que leur orchestration par les réseaux de régulation sous-jacents. Cependant, les connaissances accumulées concernent principalement la microalgue verte *C. reinhardtii*. Or, les microalgues sont des organismes très diversifiés. Par conséquent les mécanismes de régulation mis en place par une microalgue verte seront difficilement transposables à une microalgue haptophyte ou straménopile qui sont très éloignées phylogénétiquement, contrairement aux plantes terrestres qui descendent des microalgues vertes. De ce fait, ces dernières tirent un grand profit des connaissances accumulées depuis des décennies chez les plantes terrestres. Comme l'illustrent les études de Ibanez-Salazar *et al.* (2014) et Zhang *et al.* (2014) qui ont transféré un FT de soja de la famille des Dof chez une microalgue verte (respectivement *C. reinhardtii* et *Chlorella ellipsoidea*) aboutissant à une augmentation de la production de lipides. De plus, *C. reinhardtii* étant un organisme modèle, un grand nombre de techniques et de données sont à ce jour disponibles (Jinkerson & Jonikas, 2015). Comme l'illustre l'étude de Lopez Garcia de Lomana *et al.* (2015) qui ont utilisé des données transcriptomiques publiées par Boyle *et al.* (2012) ainsi que le réseau métabolique de *C. reinhardtii* publié par Chang

et al. (2011). Une telle quantité de données et de connaissances, disponible chez les organismes modèles comme *Arabidopsis thaliana* (Provart *et al.*, 2016), reste encore à générer dans le cas des organismes non-modèles. De plus, des outils indispensables à une étude approfondie tels la transformation génétique ou la production d'anticorps spécifiques restent à développer. En conséquence, chez ces organismes moins étudiés, l'utilisation de données d'expression de gènes est souvent une première alternative intéressante dans le but de comprendre les mécanismes sous-jacents à un phénotype particulier.

X. Contexte et présentation de l'étude

L'organisme étudié lors de cette thèse est la microalgue haptophyte *Tisochrysis lutea* qui appartient à la famille des Isochrysidaceae et à l'ordre des isochrysidales. Précédemment identifiée comme *Isochrysis* affinis *galbana* clone Tahiti en se fondant sur des critères morphologiques, une étude phylogénétique a récemment renommé cette espèce *Tisochrysis lutea* (Bendif *et al.*, 2013). Elle est traditionnellement cultivée pour nourrir les larves et juvéniles de mollusques et de crustacés ou de proies vivantes pour la pisciculture et aquariologie (Okauchi, 1990 ; Hemaiswarya *et al.*, 2011). Cette utilisation est due à une composition intéressante en acides gras polyinsaturés à longue chaine tels que les acides docosahexaènoique (22:6 (n-3), DHA), stéaridonique (18:4 (n-3)) et alpha-linolénique (18:3 (n-3)) (Renaud *et al.,* 1995). Cette production d'acides gras polyinsaturés à longue chaine, et notamment le DHA, en fait également un candidat intéressant pour la nutrition et la santé (Meireles *et al.*, 2003 ; Ryckebosch *et al.*, 2014). D'autre part, *T. lutea* est une espèce couramment utilisée pour des études écophysiologiques (Saoudi-Helis *et al.*, 1994; Bougaran *et al.*, 2003, 2010; Marchetti *et al.*, 2012; Lacour *et al.*, 2012). Enfin, sa teneur en acides gras totaux (supérieure à 250 mg/g C) et une répartition des acides gras dans les lipides neutres supérieure à 70% des acides gras totaux sous carence azotée font de *T. lutea* une algue oléagineuse (Bougaran *et al.*, 2012). Cette caractéristique métabolique en fait donc un bon candidat pour la production de biodiesel (Sánchez *et al.*, 2013).

Par conséquent *T. lutea* a fait l'objet d'un programme d'amélioration d'espèce visant à augmenter sa production de lipides via une approche de mutagénèse aléatoire (Bougaran *et al.*, 2012). Ce projet a abouti, en 2012, à l'isolation d'une souche de *T. lutea* accumulant deux fois plus de lipides neutres (souche 2Xc1) que la souche sauvage (WT) en condition de carence azotée (Figure 37).

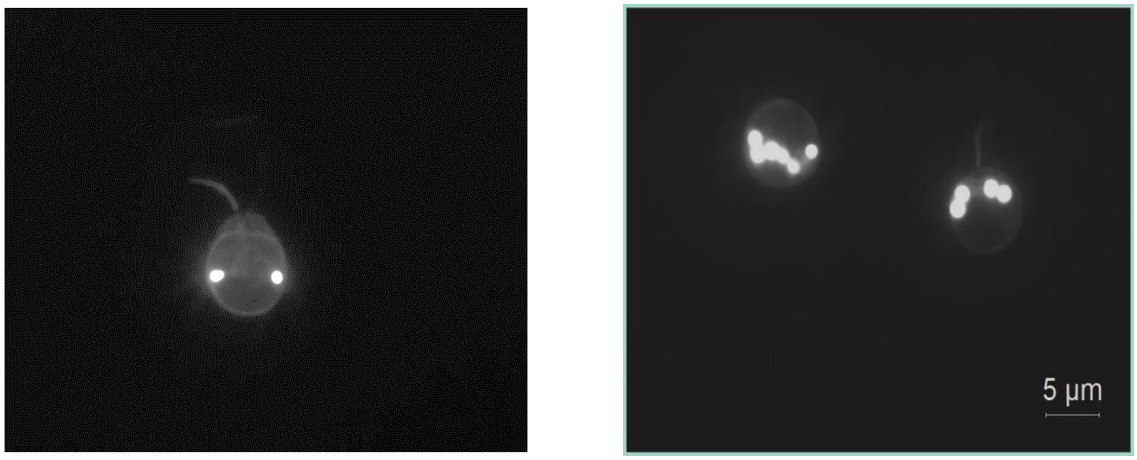

Figure 37 : observation microscopique des deux souches de *Tisochrysis lutea* (souche sauvage WT à gauche et souche mutante 2Xc1 à droite). Les vésicules lipidiques sont colorées au Nile Red.

Cette souche a été obtenue via un procédé de mutation-sélection consistant en une irradiation aux UVc suivie par une sélection par cytométrie en flux des cellules ayant le contenu lipidique le plus élevé (marquage au Nile-red). Ce phénotype particulier étant le résultat de mutations aléatoires du génome consécutives à l'exposition aux UVc, le laboratoire s'est attaché à identifier les gènes et mécanismes impliqués dans son l'établissement. Pour cela, le génome de *T. lutea* a été séquencé et une étude transcriptomique comparant les deux souches (WT et 2Xc1) cultivées en mode batch a été menée (Carrier *et al.*, 2014). Celle-ci a mis en évidence un certain nombre de gènes exprimés de façon différentielle entre les deux souches, parmi lesquels 4 FTs putatifs. Etant donné le rôle clé de ces protéines dans le bon fonctionnement d'une cellule et l'établissement du phénotype, mon travail de thèse a pour objectif de répondre aux questions suivantes : quel est l'impact des FTs dans l'établissement du phénotype de la souche mutante de *T. lutea* comparativement à la souche sauvage ? Et quels sont les mécanismes métaboliques affectés par l'action de ces FTs ?

Chapitre 1 : Identification *in silico* des facteurs de transcription (FTs) dans le génome de microalgues : vers une meilleure compréhension de l'histoire évolutive des microalgues

Tout d'abord la mise au point d'un pipeline bio-informatique a été nécessaire dans le but d'identifier et classer les FTs présents dans le génome de *T. lutea*. Compte tenu du manque de données chez les haptophytes et du vide qu'il constitue dans les études visant à comprendre l'histoire évolutive des algues, la disponibilité au laboratoire d'un nouveau génome d'haptophyte a été un avantage certain. Celui-ci a été mis à profit pour la réalisation d'une étude comparative fondée sur l'identification des FTs dans le génome d'algues de différentes lignées. Cette étude comparative a pour but d'identifier des caractéristiques de plusieurs taxons quant à la présence/absence de familles de FTs ainsi qu'au niveau des proportions de chaque famille chez les espèces étudiées.

Chapitre 2 : Identification de FTs impliqués dans l'orchestration des remaniements métaboliques constituant la réponse spécifique de la souche mutante de *T. lutea* à un stress azoté

L'ensemble des FTs étant identifiés, de nouvelles données transcriptomiques provenant de cultures des deux souches (sauvage et mutante) dans différentes conditions physiologiques ont été analysées dans le but d'identifier les FTs potentiellement impliqués dans l'établissement du phénotype mutant. L'objectif étant d'identifier les spécificités de régulation de l'expression des gènes de la souche mutante et de les lier au phénotype, une approche intégrative sera utilisée. *T. lutea* est un organisme non-modèle appartenant aux haptophytes, une lignée de microalgue très éloignée de la lignée verte, la plus étudiée, pour laquelle une grande quantité de données est disponible. Du fait du manque de données moléculaires, de connaissances du métabolisme, des mécanismes de régulation des gènes, et de techniques développées chez *T. lutea* (transformation génétique, caractérisation d'anticorps…), je me suis attaché à analyser des réseaux de co-expression des gènes et de régulation des gènes construit à partir de données transcriptomiques.

Chapitre 1 : Identification *in silico* des facteurs de transcription (FTs) dans le génome de microalgues : vers une meilleure compréhension de l'histoire évolutive des microalgues.

Comme précisé plus haut (Contexte et présentation de l'étude), le principal objectif de cette thèse consiste à identifier les FTs potentiellement impliqués dans l'établissement du phénotype de la souche mutante (2Xc1) de l'haptophyte *T. lutea*. Dans cette optique, la première étape a consisté à identifier et classer les FTs présents dans le génome de la microalgue à l'aide d'un pipeline bio-informatique. L'élaboration du pipeline, son application au protéome prédit de *T. lutea* et la présentation des résultats obtenus feront l'objet de la première partie de ce chapitre. Dans une deuxième partie, l'histoire évolutive des microalgues sera abordée par le biais des FTs. Dans cette optique, la présence et l'abondance des familles de FTs identifiées chez des microalgues de différentes lignées seront comparées. Cette partie a fait l'objet d'un article publié dans BMC Genomic en avril 2016 (Thiriet-Rupert *et al.*, 2016) et dont les principaux résultats seront présentés. Enfin, une dernière partie présentera des résultats complémentaires aux données de cette publication avec l'apport d'un nouveau génome d'haptophyte, ainsi qu'une étude des FTs de la famille des bHLH chez les haptophytes.

I. Développement d'un pipeline bio-informatique pour l'identification *in silico* de FTs dans le génome de *T. lutea*

1. Introduction

Un certain nombre d'études se sont attachées à la mise au point de différents pipelines d'identification et de classification des FTs dans le génome d'organismes représentant l'ensemble de l'arbre de la vie. La première étude portait sur les premiers organismes dont le génome a été séquencé, des archées, afin d'étudier la régulation transcriptionnelle chez ce groupe d'organismes peu connus (Aravind & Koonin, 1999). Depuis lors, l'avènement des techniques « omiques » et le séquençage d'un grand nombre de génomes, a permis d'appliquer ce type d'approche à des organismes variés, archées, bactéries (Martínez-Bueno *et al.*, 2004), cyanobactéries (Wu *et al.*, 2007), plantes (Pérez-Rodríguez *et al.*, 2010 ; Jin *et al.*, 2014), champignons (Park *et al.*, 2008) et animaux (Zhang *et al.*, 2012). Concernant les microalgues, des études comparatives portant sur les FTs ont été réalisées chez les straménopiles (Rayko *et al.*, 2010) ainsi que chez les microalgues

vertes et rouges dans le but de déterminer leur place dans l'histoire évolutive des organismes photosynthétiques (Lang *et al.*, 2010 ; Sharma *et al.*, 2013).

La première étape d'annotation des facteurs de transcription consiste à les identifier dans le génome de l'organisme étudié. Cependant, bien que cette approche soit de plus en plus utilisée, aucun pipeline bio-informatique universel n'existe encore. Par conséquent, chaque étude développe sa propre méthodologie qui, généralement, met en jeu une combinaison d'outils d'identification puisés parmi les nombreux logiciels d'analyse bioinformatique existants. Dans le cas des études appliquées aux génomes de microalgues, les deux stratégies suivantes ont été utilisées : une recherche de similarité de séquence par BLAST contre une base de données de FTs spécifique aux plantes (Rayko *et al.*, 2010 ; Sharma *et al.*, 2013) ou une annotation des domaines de liaison à l'ADN de FTs de plantes grâce à l'utilisation du logiciel HMMER (Lang *et al.*, 2010 ; Pérez-Rodríguez *et al.*, 2010 ; Jin *et al.*, 2014). Ces deux approches peuvent éventuellement être combinées afin d'être plus exhaustif dans l'identification des FTs (Wu *et al.*, 2007 ; Richardt *et al.*, 2007).

Du fait de la diversité des méthodologies utilisées pour identifier les FTs, je me suis attaché à mettre au point un pipeline le plus complet possible et le mieux adapté à l'identification de FTs chez un organisme photosynthétique. Dans ce but, les deux approches de recherche de similarité de séquence par BLAST et d'annotation des domaines de liaison à l'ADN ont été couplées. Afin d'être plus exhaustif, la recherche de FTs n'a pas été restreinte aux seuls FTs de plantes mais élargie aux familles de FTs présentes chez les algues, les champignons ou les cyanobactéries. L'application de ce pipeline au protéome de *T. lutea*, prédit à partir des données génomiques obtenues au laboratoire, a permis d'identifier les FTs appartenant aux familles connues à ce jour.

Suite à cette identification *in silico*, les FTs sont classés en familles en fonction des domaines de liaison à l'ADN présents dans leur séquence. Des règles d'assignement sont ainsi établies à partir de domaines qui doivent être présents dans une séquence pour qu'elle soit assignée à une famille donnée, et de domaines qui ne doivent pas y être associés. Par exemple, une séquence comportant un domaine homeobox et un domaine PHD (Plant Homeodomain) appartient à la famille HB-PHD, mais une séquence comportant un domaine homeobox seul appartient à la famille HB-other. Ainsi, une séquence appartenant à la famille HB-other comporte uniquement un domaine homeobox.

Le pipeline conçu dans le cadre de cette thèse combine les deux approches utilisées dans la littérature : une recherche de similarité de séquence par BLAST contre une base de données de FTs, et une annotation des domaines de liaison à l'ADN. Afin d'être le plus exhaustif possible, la recherche de similarité de séquence par BLAST a été réalisée contre une base de données construite à partir de séquences de FTs identifiés chez des microalgues (les microalgues vertes *Bathycoccus prasinos*, *Chlorella* sp, *Coccomyxa* sp, *Micromonas pusilla*, *Micromonas* sp, *Ostreococcus lucimarinus*, *Ostreococcus* sp, *Ostreococcus tauri* et *Volvox carteri* ; les microalgues rouges *Cyanidioschyzon merolae* et *Galdieria sulfuraria* ; la diatomée *Thalassiosira pseudonana*), la plante modèle *Arabidopsis thaliana*, la levure *Saccharomyces cerevisiae* et 31 souches de cyanobactéries, puisque les microalgues sont originaires de l'endosymbiose d'une cyanobactérie primitive. Quant à l'annotation des domaines de liaison à l'ADN, les logiciels InterProScan et HMMER ont été utilisés. L'utilisation d'InterProScan offre l'avantage d'annoter les domaines fonctionnels (parmi lesquels les domaines de liaison à l'ADN) de protéines à partir des 11 bases de données de domaines du consortium InterProScan (Jones *et al.*, 2014). Cette recherche est bien plus exhaustive que les études précédentes qui utilisaient seulement une ou deux bases de données (Wu *et al.*, 2007 ; Richardt *et al.*, 2007 ; Pérez-Rodríguez *et al.*, 2010). Cette méthodologie a donc le double avantage de ne pas se restreindre à une recherche de FTs connus uniquement chez les plantes, tout en étant plus exhaustive en annotant les domaines fonctionnels à partir d'une recherche complémentaire dans 11 bases de données différentes. Après exclusion des faux positifs, les FTs candidats sont classés en familles suivant des règles d'assignement définies à partir de la littérature (Rayko *et al.*, 2010 ; Pérez-Rodríguez *et al.*, 2010 ; Wu *et al.*, 2007 ; Lang *et al.*, 2010 ; Jin *et al.*, 2014).

Afin de vérifier sa fiabilité, ce pipeline a été appliqué au protéome prédit à partir du génome séquencé de la plante modèle *A. thaliana* (TAIR 10) et des trois cyanobactéries : *Synechocystis* sp. PCC 6803, *Synechococcus* sp. CC9605 et *Nostoc punctiforme* PCC73102. Les résultats obtenus ont démontré l'efficacité du pipeline pour identifier les FTs connus chez les organismes photosynthétiques tels que les plantes et les cyanobactéries (Thiriet-Rupert *et al.*, 2016).

2. Principaux résultats

Le pipeline, ainsi validé, a été appliqué au protéome prédit de l'haptophyte *T. lutea*, permettant d'identifier 155 FTs répartis en 27 familles. Ces 155 FTs représentent l'équivalent de 0,8% de l'ensemble du protéome prédit. Une telle proportion du protéome dédiée aux FTs est semblable aux proportions identifiées chez d'autres microalgues (Rayko *et al.*, 2010 ; Pérez-Rodríguez *et al.*, 2010 ; Sharma *et al.*, 2013 ; Jin *et al.*, 2014). De manière surprenante, aucun membre de la famille des bHLH (Basic Helix-Loop-Helix) n'a été identifié parmi ces 155 FTs. Cette famille est pourtant très répandue chez les eucaryotes et est la deuxième famille de FTs la plus représentée chez les plantes (Jin *et al.*, 2014). Toutefois, l'absence de cette famille bHLH a également été remarquée précédemment chez une autre haptophyte, *E. huxleyi* (Rayko *et al.*, 2010). Faire une généralité de l'absence des bHLH chez les haptophytes fondée uniquement sur deux membres de ce groupe serait une conclusion hâtive. La question de l'absence des bHLH chez les haptophytes sera abordée dans de plus amples détails dans la troisième partie de ce chapitre (Résultats complémentaires). Chez *T. lutea*, une autre absence surprenante est celle des NF-YA (Nuclear Factor Y subunit A). Ces FTs appartiennent à la famille des NF-Y, présente chez tous les eucaryotes. Les FTs NF-Y sont composés de trois sous-unités : NF-YA, NF-YB et NF-YC. Les sous-unités B et C dimérisent dans le cytoplasme puis sont transloquées dans le noyau où elles recrutent la sous-unité A, formant un trimère actif (Frontini *et al.*, 2004 ; Kahle *et al.*, 2005). Cependant d'autres FTs peuvent interagir avec les sous-unités B et C comme des FTs appartenant aux familles C2C2-CO-like et bZIP chez *Arabidopsis*, permettant l'activation des gènes cibles (Wenkel *et al.*, 2006 ; Yamamoto *et al.*, 2009a). L'absence de la sous-unité A chez l'haptophyte *T. lutea* suggère soit une perte de la fonctionnalité des NF-Y, soit la présence d'interactions avec d'autres FTs tel que montré chez les plantes. Enfin, on notera également l'absence de la sous-unité A dans le génome de certaines chlorophytes (*C. reinhardtii*, *Vovlox carteri* et *Ostreococcus tauri*) (Jin *et al.*, 2014).

Des études comparatives fondées sur la présence et l'absence de familles de FTs identifiés dans le génome d'organismes photosynthétiques ont mis en évidence des familles spécifiques de la lignée verte (Sharma *et al.*, 2013 ; Lang *et al.*, 2010). Parmi elles, cinq familles (ABI3/VP1, AP2/ERF, C2C2-LSD, CSD et TUB) ont été identifiées dans le génome de *T. lutea*, pourtant dérivée de l'endosymbiose d'une microalgue rouge (Keeling, 2013). Les deux familles AP2/ERF

(APETALA2/Ethylene Responsive Factor) et CSD (Cold Shock Domain) avaient précédemment été identifiées chez l'haptophyte *E. huxleyi* (Rayko *et al.*, 2010) mais les trois autres n'avaient jusqu'alors été identifiées que chez des organismes de la lignée verte. Leur présence sera davantage discutée dans le chapitre « Résultats complémentaires : Apport du génome de l'haptophyte *Chrysochromulina tobin* à cette étude comparative ».

Parmi les 155 FTs de *T. lutea*, des FTs de type fongique ont également été identifiés par cette approche *in silico*. Ces « Fungal TRF », très abondants et bien décrits chez les champignons (MacPherson *et al.*, 2006), ont été identifiés précédemment chez l'haptophyte *E. huxleyi* ainsi que chez des straménopiles et une microalgue rouge (Rayko *et al.*, 2010 ; Pérez-Rodríguez *et al.*, 2010). En revanche, aucun membre de cette famille n'a été identifié chez les microalgues vertes. Une telle distribution est bien en accord avec les connaissances actuelles de l'histoire évolutive des algues, les haptophytes et straménopiles étant dérivées de l'endosymbiose d'une microalgue rouge.

Enfin, cette approche a permis l'identification, pour la première fois chez une microalgue, de FTs correspondant à des familles de FTs de cyanobactéries dans le génome de *T. lutea*. Une inspection particulière a confirmé la localisation nucléaire de ces séquences en excluant une éventuelle contamination bactérienne ainsi qu'une localisation dans le génome mitochondrial ou chloroplastique. Toutefois la présence de ces séquences dans le génome nucléaire n'implique pas qu'elles soient exprimées, ni fonctionnelles si elles sont exprimées. Quant à leur origine, ces familles de FTs étant présentes à la fois chez les bactéries et les cyanobactéries, elles peuvent provenir d'événements de transferts de gènes horizontaux depuis une bactérie et/ou d'un transfert de gène endosymbiotique.

Ce pipeline est donc efficace et adapté à l'identification de FTs dans le génome d'organismes photosynthétiques dont les microalgues. De plus, son exhaustivité relative permet d'identifier un plus large spectre de familles de FTs. Ceci pouvant mener à une meilleure compréhension des mécanismes de régulation chez les microalgues, ainsi qu'à une meilleure appréhension de leur évolution par le prisme des FTs.

II. L'identification de familles de FTs pour la compréhension de l'histoire évolutive des microalgues

1. Introduction

L'histoire évolutive des microalgues est très complexe et encore très controversée. Afin de mieux la comprendre, des études génomiques sont menées en fondant leur analyse sur différentes familles de gènes, et notamment les FTs (Sharma *et al.*, 2013 ; Lang *et al.*, 2010). Ces familles de gènes sont étroitement liées au développement et à l'évolution des organismes (Gutiérrez *et al.*, 2004), notamment leur complexité (Levin & Tjian, 2003) et la diversification de leur morphologie (Lespinet *et al.*, 2002 ; Richardt *et al.*, 2007). Chez les microalgues, peu d'études ont été menées. Richardt *et al.*, (2007) ont identifié des FTs chez des organismes représentant tous les domaines du vivant, et se sont ensuite concentrés sur les organismes photosynthétiques eucaryotes. Ils ont ainsi identifié des expansions de certaines familles de FTs (proportion plus élevée d'une famille chez une espèce par rapport aux autres) liées à la complexification morphologique de ces organismes. Rayko *et al.*, (2010) ont identifié et classé les FTs dans le génome de six straménopiles (2 diatomées, 1 pelagophycée, 1 phaeophycée et 2 champignons) et de l'haptophyte *E. huxleyi*. Leur étude comparative a permis de mettre en évidence des spécificités de lignée telles l'expansion des HSFs chez les diatomées et la présence des auréochromes chez les straménopiles photosynthétiques. Lang *et al*, (2010) ont identifié les FTs présents dans le génome de sept microalgues vertes et des plantes terrestres. Ils ont ainsi identifié des événements de gain et d'expansion majeurs de FTs à des moments clés de l'évolution de ces organismes : la sortie de l'eau et l'expansion des plantes à fleurs. Sharma *et al*, (2013) ont identifié les FTs présents dans le transcriptome de l'hépatique *Marchantia polymorpha*. Les hépatiques sont des bryophytes positionnés à la limite des groupes des straménopiles et des plantes terrestres. L'étude comparative des FTs de *M. polymorpha*, de microalgues (sept microalgues vertes, deux microalgues rouges) et des plantes terrestre a permis d'évaluer l'évolution des familles de FTs depuis les microalgues vers les plantes terrestres. Enfin, Buitrago-Florez *et al.*, (2014) se sont attachés à identifier les FTs dans le génome de straménopiles (diatomées et champignons) afin de lier des familles de FTs à des caractères morphologiques illustrant la diversité de ce groupe d'organismes très diversifié.

Toutefois le manque de données génomiques constitue l'un des principaux verrous de ce type d'études. En effet, certains taxons ne comptent encore que peu de membres dont le génome est séquencé et disponible. Les cryptophytes et les chlorarachniophytes ne comptent qu'un seul de leurs membres dont le génome est disponible. Trois génomes de microalgues rouges sont actuellement disponibles dont deux appartiennent à des extrêmophiles, difficilement utilisables comme modèle représentatif. Les microalgues vertes et les straménopiles comptent respectivement dix et treize génomes disponibles et, au début de cette thèse, le génome d'*E. huxleyi* était le seul génome d'haptophyte disponible (Tableau 1).

Afin d'appréhender l'histoire évolutive des organismes par l'intermédiaire des FTs, une analyse robuste et exhaustive est nécessaire. De plus, avoir un maximum de génomes disponibles pour chaque taxon existant est primordial. Dans cette optique, avoir à disposition un deuxième génome d'haptophyte ainsi qu'un pipeline d'identification et de classification des FTs efficace chez les organismes photosynthétiques constitue une avancée certaine.

2. Principaux résultats

Cet avantage a donc été mis à profit en menant une étude comparative fondée sur l'identification des FTs présents dans le génome de sept microalgues : trois haptophytes (*T. lutea*, *E. huxleyi* et *Pavlova* sp. pour laquelle nous disposons uniquement du transcriptome) et deux straménopiles (la diatomée modèle *P. tricornutum* et l'eustigmatophycée *Nannochloropsis gaditana*) qui sont deux lignées proches (Andersen, 2004 ; Moustafa *et al.*, 2009). La microalgue verte *C. reinhardtii* et la microalgue rouge *Porphyridium purpureum*, qui résultent d'un événement d'endosymbiose primaire, complètent cette étude. Leur protéome prédit a été soumis à notre pipeline d'identification et les différentes familles de FTs identifiées ont été comparées, ainsi que leurs abondances respectives. Cette étude a fait l'objet d'une publication (Thiriet-Rupert et al 2016) présentée à la fin de ce chapitre.

Cette étude comparative a permis de mettre en évidence une répartition particulière de certaines familles. Comme l'illustre la « heatmap » générée à partir des proportions de chaque famille chez les sept microalgues (Figure 38), quatre clusters sont particulièrement intéressants.

Tableau 1 : tableau répertoriant les espèces de microalgues dont le génome est disponible

Espèce	Souche	Phylum	Taille du genome (Mega Bases)	Nombre de genes	Publication	Liens vers les données disponibles
Cyanophora paradoxa	CCMP329	Glaucophyte	70,2	27 921	Price et al., 2012	http://cyanophora.rutgers.edu/cyanophora/blast.php
Auxenochlorella protothecoides	710	Chlorophytes	22,9	7014	Gao et al., 2014	https://www.ncbi.nlm.nih.gov/genome/?term=Auxenochlorella+protothecoides
Bathycoccus prasinos	RCC1105		151	7847	Moreau et al., 2012	https://bioinformatics.psb.ugent.be/gdb/bathycoccus/RELEASE_15jul2011/
Chlamydomonas reinhardtii	CC-503 cw92 mt+		121	15143	Merchant et al., 2007	http://genome.jgi.doe.gov/Chlre4/Chlre4.home.html
Chlorella variabilis	NC64A		46	9791	Blanc et al., 2010	http://genome.jgi.doe.gov/ChlNC64A_1/ChlNC64A_1.home.html
Coccomyxa subellipsoidea	C-169		48,8	9851	Blanc et al., 2012	http://www.ncbi.nlm.nih.gov/genome/2692
Dunaliela salina	325		na	18801	Dunaliella salina Genome Sequencing Project, http://phytozome.jgi.doe.gov/	http://genome.jgi.doe.gov/pages/dynamicOrganismDownload.jsf?organism=PhytozomeV11
Gonium pectoral	K3-F3-4 (mating type minus, NIES-2863)		148,8	16290	Hanschen et al., 2016	https://www.ncbi.nlm.nih.gov/genome/?term=Gonium+pectorale
Helicosporidium	ATCC50920		12,4	6 033	Pombert et al., 2014	https://www.ncbi.nlm.nih.gov/genome/?term=Helicosporidium
Micromonas pusilla	CCMP1545		21,9	10 575	Worden et al., 2009	http://genome.jgi.doe.gov/MicpuC3/MicpuC3.home.html
Micromonas sp.	RCC299		20,9	10 056	Worden et al., 2009	http://genome.jgi.doe.gov/MicpuN3/MicpuN3.home.html
Monoraphidium neglectum	SAG 48.87		69,7	16 755	Bogen et al., 2013	https://www.ncbi.nlm.nih.gov/genome/?term=Monoraphidium+neglectum
Ostreococcus lucimarinus	CCE9901		13,2	7 651	Palenik et al., 2007	http://genome.jgi.doe.gov/Ost9901_3/Ost9901_3.home.html
Ostreococcus sp. RCC809	RCC809		13,3	7 492	Palenik et al., 2007	http://genome.jgi.doe.gov/OstRCC809_2/OstRCC809_2.home.html
Ostreococcus tauri	OTH95		12,6	7 892	Palenik et al., 2007	http://genome.jgi.doe.gov/Ostta4/Ostta4.home.html
Picochlorum sp	SENEW3 (SE3)		13,5	7367	Foflonker et al., 2015	http://cyanophora.rutgers.edu/picochlorum/
Volvox carteri f. nagariensis	EVE		137,68	14 437	Prochnik et al., 2010	http://genome.jgi.doe.gov/Volca1/Volca1.home.html
Cyanidioschyzon merolae	10D	Rhodophytes	16,55	6 170	Matsuzaki et al., 2004	http://merolae.biol.s.u-tokyo.ac.jp/
Galdieria sulphuraria	074W		13,71	6 723	Schönknecht et al., 2013	http://www.ncbi.nlm.nih.gov/genome/405
Porphyridium purpureum	DBLAB2		19,7	8 355	Bhattacharya et al., 2013	http://cyanophora.rutgers.edu/porphyridium/
Bigelowiella natans	CCMP2755	Chlorarachiophytes	94,7	21 708	Curtis et al., 2012	http://genome.jgi.doe.gov/Bigna1/Bigna1.home.html
Guillardia theta	CCMP2712	Cryptophyte	87,2	24 840	Curtis et al., 2012	http://www.ncbi.nlm.nih.gov/genome/57
Fragilariopsis cylindrus	*CCMP 1102*	Stramenopile (diatomée)	Données préliminaires		non publié	http://genome.jgi.doe.gov/Fracy1/Fracy1.home.html
Phaeodactylum tricornutum	CCAP1055/1		27,45	10 398	Bowler et al., 2008	http://genome.jgi.doe.gov/Phatr2/Phatr2.home.html
Pseudo-nitzschia multiseries	CLN-47		Données préliminaires		non publié	http://genome.jgi.doe.gov/Psemu1/Psemu1.home.html
Thalassiosira Oceanica	CCMP1005		92,04	34 684	Lommer et al., 2012	https://www.ncbi.nlm.nih.gov/nuccore/AGNL00000000
Thalassiosira pseudonana	CCMP 1335		32,44	13 025	Ambrust et al., 2004	http://genome.jgi.doe.gov/Thaps3/Thaps3.download.html
Nannochloropsis gaditana	CCMP526	Stramenopile (Eustogmatophycée)	29	8 892	Radakovits et al., 2012	http://nannochloropsis.genomeprojectsolutions-databases.com/
Nannochloropsis oceanica	CCMP1779		28,7	11 973	Vieler et al., 2012	https://bmb.natsci.msu.edu/faculty/christoph-benning/nannochloropsis-oceanica-ccmp1779/
Nannochloropsis granulata	CCMP529		na	na	données utilisé par HU et al., 2014	http://www.ncbi.nlm.nih.gov/bioproject/PRJNA65111/
Nannochloropsis oceanica	IMET1		na	9 915		http://www.ncbi.nlm.nih.gov/bioproject/PRJNA202418/
Nannochloropsis oceanica	CCMP531		na	na		http://www.ncbi.nlm.nih.gov/bioproject/PRJNA65113/
Nannochloropsis oculata	CCMP525		na	na		http://www.ncbi.nlm.nih.gov/bioproject/PRJNA65107/
Nannochloropsis salina	CCMP537		na	na		http://www.ncbi.nlm.nih.gov/bioproject/PRJNA62503/
Aureococcus anophagefferens	CCMP1984	Stramenopiles (Pelagophycée)	56,66	11 522	Gobler et al., 2011	http://genome.jgi.doe.gov/Auran1/Auran1.home.html
Chrysochromulina tobin	CCMP291	Haptophyte	59	16 777	Hovde et al., 2015	http://www.ncbi.nlm.nih.gov/bioproject/263501
Emiliania huxleyii	*CCMP1516*		167,68	38 549	Read et al., 2013	http://genome.jgi.doe.gov/haptophyta/haptophyta.info.html
Tisochrysis lutea	*CCAP 926/14*		58,7	22 969	Carrier et al., *in prep*	

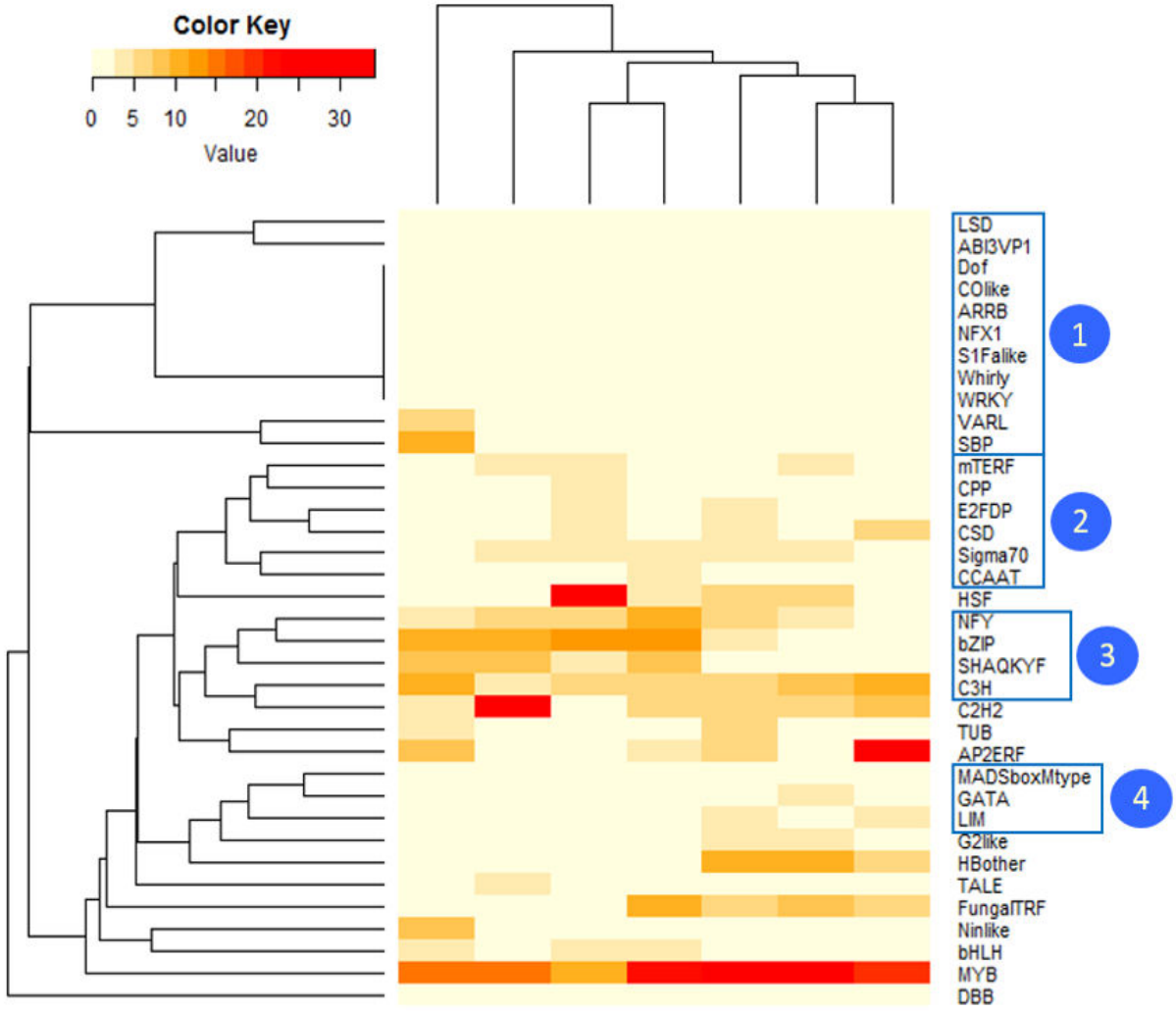

Figure 38 : heatmap générée à partir des proportions de chaque famille de FTs identifiées chez les sept microalgues étudiées.

Le premier (1) regroupe 11 familles de FTs qui étaient jusqu'alors décrites comme spécifiques de la lignée verte (Sharma *et al.*, 2013 ; Lang *et al.*, 2010). Or, certaines d'entre elles ont été identifiées chez des microalgues appartenant aux straménopiles et aux haptophytes. Ceci s'explique par le fait que cette relative spécificité avait été établie à partir d'études comparatives ne comprenant que des plantes terrestres, des microalgues vertes et deux microalgues rouges, *Cyanidioschyzon merolae* et *Galdieria sulfuraria*. Les familles absentes chez ces deux microalgues rouges ont été définies comme spécifiques de la lignée verte. Or, ces deux espèces sont des microalgues extrêmophiles adaptées aux pressions de sélections relatives à leur milieu de vie (telles les sources chaudes du parc Yellowstone) (Schönknecht *et al.*, 2013). En conséquence, l'absence d'une famille de FTs chez ces deux espèces ne peut pas être considérée comme une caractéristique de la lignée rouge. De plus, notre échantillonnage est plus large puisque nous incluons des microalgues haptophytes et straménopiles. Cette exhaustivité accrue permet d'affiner les spécificités de lignées précédemment établies. (2) Le deuxième cluster contient des familles de FTs communes aux sept microalgues et ayant des proportions équivalentes. Des familles communes à tous les eucaryotes, telles les E2F/DP et les NF-Y, y sont notamment représentées (Figure 38). (3) Le troisième cluster comporte des familles communes aux sept microalgues, mais dans des proportions différentes représentant certaines spécificités. Les auréochromes, caractérisés par la présence d'un domaine de liaison à l'ADN bZIP (basique leucine zipper) et d'un domaine senseur LOV (Light, Oxygene, Voltage) sont notamment comptés parmi les bZIP. Ceux-ci sont spécifiques des straménopiles photosynthétiques (Figure 4 de Thiriet-Rupert et al 2016). Enfin, le quatrième cluster (4) représente trois familles (LIM, MADS-box et C2C2-GATA) absentes chez les deux straménopiles mais présentes chez les autres microalgues (Figure 38). Cette absence chez les straménopiles a été confirmée par une recherche spécifique chez les trois diatomées *Thalassiosira pseudonana*, *Pseudo-nitzschia multiseries* et *Fragilariopsis cylindrus*.

Outre ces quatre clusters, l'absence de FTs de la famille bHLH mise en évidence chez *T. lutea* a été confirmée chez les deux autres haptophytes alors que les quatre autres algues en comportent. Trois expansions ont été identifiées : une expansion des HSFs chez *P. tricornutum*, laquelle est caractéristique des diatomées ; une expansion des AP2/ERF chez l'haptophyte *E. huxleyi*, qui avait déjà été remarquée dans une étude précédente (Rayko *et al.*, 2010) ; et une expansion des C2H2 chez la microalgue rouge *P. purpureum*. Des FTs appartenant à une famille connue jusqu'alors uniquement chez les levures et les champignons ont également été identifiés chez les haptophytes

et les straménopiles. De plus, comme chez *T. lutea*, des FT appartenant à des familles retrouvées chez les cyanobactéries ont été identifiées dans le génome de ces sept microalgues. Ces séquences pourraient être le résultat de transferts de gènes endosymbiotiques lors de la domestication de l'algue internalisée, ou d'un transfert de gène horizontal ayant eu lieu au cours de l'histoire évolutive de ces organismes. Enfin, les spécificités de lignées étant illustrées par une présence/absence de familles de FTs, un dendrogramme a été construit à partir de ces critères afin d'évaluer leur concordance avec la phylogénie des microalgues. La position de chaque espèce sur le dendrogramme produit (Figure 3 de Thiriet-Rupert et al 2016) correspond bien à la phylogénie connue des microalgues, prouvant la fiabilité de l'utilisation des FTs dans de telles études comparatives.

III. Transcription factors in microalgae: genome-wide prediction and comparative analysis, Thiriet-Rupert et al 2016

Thiriet-Rupert et al. BMC Genomics (2016) 17:282
DOI 10.1186/s12864-016-2610-9

BMC Genomics

RESEARCH ARTICLE Open Access

Transcription factors in microalgae: genome-wide prediction and comparative analysis

Stanislas Thiriet-Rupert[1*], Grégory Carrier[1], Benoît Chénais[2], Camille Trottier[1], Gaël Bougaran[1], Jean-Paul Cadoret[1], Benoît Schoefs[2] and Bruno Saint-Jean[1]

Abstract

Background: Studying transcription factors, which are some of the key players in gene expression, is of outstanding interest for the investigation of the evolutionary history of organisms through lineage-specific features. In this study we performed the first genome-wide TF identification and comparison between haptophytes and other algal lineages.

Results: For TF identification and classification, we created a comprehensive pipeline using a combination of BLAST, HMMER and InterProScan software. The accuracy evaluation of the pipeline shows its applicability for every alga, plant and cyanobacterium, with very good PPV and sensitivity. This pipeline allowed us to identify and classified the transcription factor complement of the three haptophytes *Tisochrysis lutea*, *Emiliania huxleyi* and *Pavlova* sp.; the two stramenopiles *Phaeodactylum tricornutum* and *Nannochloropsis gaditana*; the chlorophyte *Chlamydomonas reinhardtii* and the rhodophyte *Porphyridium purpureum*. By using *T. lutea* and *Porphyridium purpureum*, this work extends the variety of species included in such comparative studies, allowing the detection and detailed study of lineage-specific features, such as the presence of TF families specific to the green lineage in *Porphyridium purpureum*, haptophytes and stramenopiles. Our comprehensive pipeline also allowed us to identify fungal and cyanobacterial TF families in the algal nuclear genomes.

Conclusions: This study provides examples illustrating the complex evolutionary history of algae, some of which support the involvement of a green alga in haptophyte and stramenopile evolution.

Keywords: Algae, Endosymbiotic gene transfer, Haptophytes, Prediction pipeline, Stramenopiles, *Tisochrysis lutea*, Transcription factors

Background

In every living organism, developmental, morphological and physiological mechanisms, such as those allowing acclimation to environmental changes, are the result of genome expression modulation. One level of this modulation is related to gene expression, in which transcription factors are among the key players [1]. These regulators can be divided into two groups: transcription factors (TFs) and transcriptional regulators (TRs). These groups interact with each other and affect gene transcription. TFs are characterized by a DNA binding domain (DBD), an oligomerization domain (allowing interaction with other TFs, as well as with other transcriptional regulators) and a transcription regulation domain (allowing control of gene expression). These proteins (also called *trans*-factors) control the expression of multiple target genes by binding to specific DNA motifs in their promotor regions. TRs interact with TFs or with chromatin allowing genes to be transcribed either (1) facilitating the recruitment of the basal transcription machinery, or (2) modifying chromatin structure, making genes more accessible [2].

TFs are classified according to their DBD [3]. Most TFs have only one DBD, which can be present in one or

* Correspondence: Stanislas.Thiriet.Rupert@ifremer.fr
[1]IFREMER, Physiology and Biotechnology of Algae Laboratory, rue de l'Ile d'Yeu, 44311 Nantes, France
Full list of author information is available at the end of the article

multiple copies in the same sequence. However, some TFs can have several DBD types in their sequence [4].

Since the first study on the identification of TFs in four archaeal genomes [5], the increase in the number of sequenced genomes facilitates putative TF identification in unrelated taxa through *in silico* studies [6–10]. Such taxonomically diverse data allows comparative analyses between different species or lineages [6, 7, 9–13] and understanding of the evolutionary aspects through TFs [11, 14, 15]. This kind of study can reveal taxonomic characteristics (i.e., the specificity and expansion of TF families) of the TF complement of different organisms. *In silico* analysis of FTs performed on *Arabidopsis thaliana* (*A. thaliana*) showed that 45 % of TFs are plant specific. Moreover, a plant-specific expansion of the MYB superfamily was demonstrated (190 copies in the *A. thaliana* genome compared with 6 and 10 in *Drosophila melanogaster* and *Saccharomyces cerevisiae*, respectively) [6]. Another example of such lineage-specific expansion of a TF family is the retinoic acid receptors in the nematode *Caenorhabditis elegans*. Using the AnimalTFDB database, 239 putative TFs belonging to this family were identified, whereas in other animals, such as *Tetraodon nigroviridis,* this TF family is only represented by 19 members [10].

Among microalgae, TF complement comparative studies have been undertaken for stramenopiles [9] and to investigate the evolutionary history of both red and green algae among photosynthetic organisms [11, 15]. Microalgae arose from the endosymbiosis of a photosynthetic eukaryote, related to today's cyanobacteria, by a primitive eukaryotic heterotroph. Glaucophyta, Rhodophyta and Chlorophyta all originated from this primary endosymbiosis [16, 17]. A series of secondary and tertiary endosymbioses would have then led to the diversity of microalgae observed today [18, 19]. Haptophytes would have appeared, as would stramenopiles, from the secondary endosymbiosis of both a green and a red alga by a heterotrophic eukaryote [19, 20]. Haptophytes are one of the key players in the evolutionary history of photosynthetic organisms [21] and are widely distributed among the photosynthetic unicellular eukaryotes in today's oceans. However, *in silico* comparative studies in haptophytes are limited because few data are available.

Here, we conducted the first genome-wide identification and comparison of the TF complement in haptophytes using an optimized and automated pipeline. This analysis pipeline combines research for similarities with known TFs and protein domains using a large database containing plant, fungal, mammal and cyanobacterial TFs. Using our pipeline, we performed the *in silico* identification of the TF complement in three haptophytes (*Tisochrysis lutea*, *Emiliania huxleyi* and *Pavlova sp*) and two stramenopiles (the eustigmatophycea, *Nannochloropsis gaditana* and the diatom *Phaeodactylum tricornutum*), which are close organism groups [19, 22], as well as in the green alga *Chlamydomonas reinhardtii* and the red alga *Porphyridium purpureum*. We focused on the identification of the main families of TFs found in these microalgal species and compared their respective abundance in each. Moreover, the present study identified, for the first time, the presence of cyanobacterial TFs in each of the microalgal genomes studied.

Results and discussion

Evaluation of transcription factor identification accuracy

Pipeline analysis is essential for whole genome TF identification. Since no universal pipeline exists, each study uses its own. However, every pipeline is based on the same tools: a single identification with BLAST searches against a plant database [9, 15], and/or a single protein domain search with HMMER software focused on plant DBDs [11–13]. Several pipelines combine both methods so as to be more accurate and exhaustive [2, 8]. Moreover, the HMMER software is used either with the Pfam database or the combination of Pfam and another database. Our pipeline also combines the same identification strategies, but with some specificities: our analysis pipeline includes more protein domain databases (the eleven databases of the InterProScan consortium) and the research is not restricted to plants, but enlarged to fungi, algae and cyanobacteria.

In order to estimate the accuracy of our pipeline (Fig. 1), we applied it to the predicted proteome of *A. thaliana* and three cyanobacteria (see Methods section). The sensitivity and the PPV were measured in the same way as [23] and [24].

The analysis of the pipeline accuracy against eleven plant TF families showed that nine were identified with a good sensitivity and PPV values equal to one (Tables 1 and 2). Only, MADS and bHLH TF families were identified with a low sensitivity and a PPV value of 0.99, respectively. Using a more recent gold standard than [23] and [24], our sensitivity and PPV values are equivalent or better than previous pipelines [24, 25].

Concerning the cyanobacterial TF families, the sensitivity value was one for all families (no false negative identified). The PPV values were equal to one for cyanobacterial TFs, except for the GntR and Crp families (0.83 and 0.88, respectively). These lower PPV values are mostly due to the lower number of TFs in these organisms (i.e., only one and two false positives for families GntR and Crp). These results indicate the high accuracy (low false positives identified) and performance (low false negatives) of our analysis pipeline for the *in silico* identification of TFs not only in plants and cyanobacteria but also for other organisms such as algae.

Fig. 1 Identification pipeline. The pipeline is divided into three steps. Step One uses two strategies: i) a similarity search against an algae-based self-built database of known TFs with BLAST software; ii) functional domain annotation with InterProScan and HMMER software. The protein list obtained is the subject of the Step Two: the filtration of false positives according to specific parameters (see Methods). The last step consists in the classification of the putative TF list obtained in Step Two using a homemade perl script followed by manual curation for specific cases (see Methods)

Table 1 Evaluation of the pipeline accuracy for each TF family for plant TFs. A sensitivity value less than one means inclusion of false negatives, and a PPV value less than one means inclusion of false positives

A. thaliana

	This study		Riaño-Pachón et al., 2007 [24]	
TF family	sensitivity	PPV	sensitivity	PPV
AP2/ERF	169/169 = 1	169/169 = 1	0.99	1
ARF	37/37 = 1	37/37 = 1	0.91	0.95
bZIP	127/127 = 1	127/127 = 1	0.92	0.97
C2C2-Dof	47/47 = 1	47/47 = 1	0.97	0.97
C2C2-GATA	41/41 = 1	41/41 = 1	1	1
GARP	85/85 = 1	85/85 = 1	NA	NA
GRAS	37/37 = 1	37/37 = 1	0.97	0.97
MADS	145/146 = 0.99	145/145 = 1	0.92	0.95
NAC	138/138 = 1	138/138 = 1	1	0.99
WRKY	90/90 = 1	90/90 = 1	0.99	0.99
bHLH	225/225 = 1	224/225 = 0.99	0.80	0.92

Table 2 Evaluation of the pipeline accuracy for each TF family for cyanobacterial TFs. A sensitivity value less than one means inclusion of false negatives, and a PPV value less than one means inclusion of false positives

Cyanobacteria		
TF family	sensitivity	PPV
arsR	12/12 = 1	12/12 = 1
Bac_DNA_binding	6/6 = 1	6/6 = 1
BolA	3/3 = 1	3/3 = 1
Crp	15/15 = 1	15/17 = 0.88
FUR	9/9 = 1	9/9 = 1
GerE	34/34 = 1	34/34 = 1
GntR	5/5 = 1	5/6 = 0.83
LysR	15/15 = 1	15/15 = 1
SfsA	3/3 = 1	3/3 = 1

Transcription factor content in algae

In this study, predicted TFs from seven algae representing four different lineages were identified and classified using our analysis pipeline (Table 3). In total, 155,128 and 478 TFs were identified in the haptophytes *Tisochrysis lutea* (*T. lutea*), *Pavlova* sp. and *Emiliania huxleyi* (*E. huxleyi*), respectively. Concerning the two stramenopiles, 196 and 93 TFs were identified in *Phaeodactylum tricornutum* (*P. tricornutum*) and *Nannochloropsis gaditana* (*N. gaditana*), respectively. Finally, 199 and 212 TFs were identified in the rhodophyte *Porphyridium purpureum* (*P. purpureum*) and the chlorophyte *Chlamydomonas reinhardtii* (*C. reinhardtii*), respectively. All TFs identified belong to common families that are largely distributed between species studied. Here, the predicted TFs of the haptophytes *T. lutea*, *Pavlova* sp. and *E. huxleyi* were divided into 27, 24 and 25 families, respectively. Twenty-two families were reported for each of the stramenopiles (*P. tricornutum* and *N. gaditana*), while 25 and 37 families were identified for *P. purpureum* and *C. reinhardtii*. According to predicted proteomes, the proportion of TFs was estimated between 0.8 and 2.4 % (Fig. 2). Such percentages in microalgae are consistent with previous studies [9, 13]. By way of comparison across the eukaryotic world, the unicellular organism *Saccharomyces cerevisiae* dedicates 3.5 % of its proteome to TFs [26]; whereas the multicellular eukaryotes such as *Drosophilia melanogaster*, *A. thaliana* and *Homo sapiens*, contain 4.6, 5.9 and 8 to 9 % TFs, respectively [6, 26, 27]. In accordance with the fact that TFs play a role in morphology diversification of organisms [28–30] these proportions show a correlation between the complexity of organisms and the proportion of TFs found in the proteome of these organisms [2, 14, 31–33]. This is illustrated by the coincidence of TF families' expansion with divergence of great eukaryotic lineages [11]. Indeed, it is well known that the evolutionary history of eukaryotes, especially plants, is punctuated by multiple biological processes, such as duplication [34–36] or domain shuffling, allowing modifications resulting in the emergence of new TF families [6, 11, 37]. These whole or partial genome duplications and domain shuffling have not been shown in algae. However, it can be reasonably assumed that such phenomena, leading to the emergence of new TF families, have also occurred in algae. This is suggested by the presence of TF families found only in green algae compared to the other algal lineages.

These lineage-specific gains and losses of TF families are a kind of mirror of their evolutionary history. To illustrate this idea, a binary table representing the presence/absence of TF families in seven algae representing four different lineages was performed. On this basis, a similarity matrix was computed to infer a dendrogram using R version 3.1.0 (Fig. 3). The resultant dendrogram (deposited in TreeBase: http://purl.org/phylo/treebase/phylows/study/TB2:S19079) confirms the relationship between algae derived from the four different lineages. Haptophytes, stramenopiles, red algae and green algae are clearly separated. We also found that *T. lutea* is more related to *E. huxleyi* than *Pavlova sp.,* as has been described in the literature [38, 39]. The rhodophyte *P. purpureum* is located between haptophytes and stramenopiles. This position is mostly due to the absence of MADS-box and C2C2-GATA families in stramenopiles, which makes them a more distant group from the four previous algae. Finally, the chlorophyte *C. reinhardtii* is the most distant from the others because of the presence of the TF families specific to the green lineage. This illustrates that the composition of this TF content is partly lineage specific. To discriminate the TF families, a haetmap was built using the data of Table 3. TF families were clustered according to their given proportions in the seven algal genomes (Fig. 4). Four interesting clusters were found: (i) TF families described as specific to green lineage. (ii) TF families with equivalent proportions among the 7 algal genomes. (iii) TF families present in the 7 algae but with different proportions. (iv) Finally, TF families only absent in stramenopiles.

In the following section, the TF content of the seven algae and their specificities of lineage, based on Table 3 and Fig. 4, are examined in more detail.

Comparison of TF families among microalgae lineages

Common TF families with equivalent proportions

The proportions of each TF family in the seven algae were compared. We found that four families were present in similar proportions throughout the algal lineage (Table 3). Among these, the Cold Shock Domain (CSD) family is distributed around 1 to 5 % in analyzed algae. Our analysis pipeline identified for the first time

Table 3 Transcription factor families identified and their proportions in seven microalgae

TF family		*Tisochrysis lutea*	*Pavlova* sp	*Emiliania huxleyi*	*Phaeodactylum tricornutum*	*Nannochloropsis gaditana*	*Porphyridium purpureum*	*Chlamydomonas reinhardtii*
B3	ABI3/VP1	1 (0.65)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.47)
AP2/ERF	AP2	1 (0.65)	1 (0.78)	58 (12.13)	0 (0)	2 (2.15)	0 (0)	6 (2.83)
	ERF	1 (0.65)	6 (4.69)	99 (20.71)	2 (1.02)	2 (2.15)	0 (0)	9 (4.25)
bHLH		0 (0)	0 (0)	0 (0)	8 (4.08)	3 (3.23)	3 (1.51)	8 (3.77)
bZIP		3 (1.94)	3 (2.34)	6 (1.26)	25 (12.76)	11 (11.83)	21 (10.55)	20 (9.43)
C2C2	CO-like	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.47)
	Dof	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.47)
	GATA	5 (3.23)	1 (0.78)	4 (0.84)	0 (0)	0 (0)	2 (1.01)	12 (0.66)
	LSD	1 (0.65)	1 (0.78)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.47)
C2H2		8 (5.16)	8 (6.25)	37 (7.74)	4 (2.04)	5 (5.38)	60 (30.15)	5 (2.36)
C3H		13 (8.39)	7 (5.47)	47 (9.83)	11 (5.61)	5 (5.38)	8 (4.02)	22 (10.38)
CCAAT		3 (1.94)	0 (0)	2 (0.42)	3 (1.53)	3 (3.23)	3 (1.51)	1 (0.47)
CPP		1 (0.65)	0 (0)	4 (0.84)	5 (2.55)	1 (1.08)	2 (1.01)	3 (1.42)
CSD		3 (1,94)	4 (3.13)	25 (5.23)	5 (2.55)	1 (1.08)	3 (1.51)	2 (0.94)
DBB		0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.50)	0 (0)
E2F/DP		2 (1.29)	3 (2.34)	3 (0.63)	5 (2.55)	1 (1.08)	3 (1.51)	3 (1.42)
Fungal TRF		14 (9.03)	8 (6.25)	27 (5.65)	1 (0.51)	10 (10.75)	0 (0)	0 (0)
GARP	G2-like	4 (2.58)	4 (3.13)	5 (1.05)	2 (1.02)	0 (0)	2 (1.01)	4 (1.89)
	ARR-B	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.47)
Homeobox	HB-other	16 (10.32)	14 (10.94)	28 (5.86)	0 (0)	0 (0)	2 (1.01)	1 (0.47)
	TALE	1 (0.65)	1 (0.78)	0 (0)	4 (2.04)	0 (0)	9 (4.52)	3 (1.42)
HSF		9 (5.81)	8 (6.25)	8 (1.67)	67 (34.18)	4 (4.30)	1 (0.50)	2 (0.94)
LIM		2 (1.29)	3 (2.34)	11 (2.30)	0 (0)	0 (0)	2 (1.01)	1 (0.47)
MADS-box	M-type	3 (1.94)	1 (0.78)	1 (0.21)	0 (0)	0 (0)	2 (1.01)	2 (0.94)
mTERF		5 (3.23)	0 (0)	6 (1.26)	5 (2.55)	2 (2.15)	5 (2.51)	4 (1.89)
MYB	MYB (3R)	1 (0.65)	0 (0)	3 (0.63)	2 (1.02)	5 (5.38)	1 (0.50)	1 (0.47)
	MYB (2R)	25 (16.13)	20 (15.63)	39 (8.16)	11 (5.61)	8 (8.60)	23 (11.56)	10 (4.72)
	MYB-rel	21 (13.55)	15 (11.90)	51 (10.69)	7 (3.57)	7 (7.53)	7 (3.52)	18 (8.65)
	MYB-SHAQKYF	1 (0.65)	2 (1.56)	1 (0.21)	7 (3.57)	8 (8.60)	16 (8.04)	4 (1.89)
NF-X1		0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.47)
NF-Y	NF-YA	0 (0)	1 (0.78)	1 (0.21)	1 (0.51)	1 (1.08)	1 (0.50)	0 (0)
	NF-YB	1 (0.65)	1 (0.78)	4 (0.84)	2 (1.02)	2 (2.15)	3 (1.51)	3 (1.42)
	NF-YC	3 (1.94)	4 (3.13)	1 (0.21)	8 (4.08)	6 (6.45)	6 (3.02)	2 (0.94)
Nin-like		0 (0)	1 (0.78)	0 (0)	0 (0)	1 (1.08)	4 (2.01)	15 (7.08)
S1Fa-like		0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.47)
SBP		0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	23 (10.85)
Sigma-70		4 (2.58)	4 (3.13)	2 (0.42)	8 (4.08)	4 (4.30)	8 (4.02)	1 (0.47)
TUB		3 (1.94)	7 (5.47)	5 (1.05)	3 (1.53)	1 (1.08)	0 (0)	6 (2.83)
VARL		0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	12 (5.66)
Whirly		0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.47)
WRKY		0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.47)
	Total	155	128	478	196	93	199	212

ERF Ethylene Response Factor, *bHLH* basic helix-loop-helix, *bZIP* basic leucine zipper, *CSD* Cold Shock Domain, *DBB* Double B-box, *TRF* Transcriptional Regulatory Factor, *HSF* Heat Shock Factor, *mTERF* mitochondrial transcription termination factor, *SBP* SQUAMOSA promotor binding protein, *VARL* Volvocine Algal RegA Like. Numbers in parentheses correspond to percentage of each family for each species. For the total number of TFs, number in parentheses corresponds to percentage of the predicted proteome dedicated to TFs

three CSD TFs in the rhodophyte *P. purpureum,* representing 1.5 % of the predicted proteome. Moreover, this family was previously described as absent from red microalgae [15]. The absence of identification of CSD TFs from the red lineage may be explained by the fact that research on red microalgae was performed only in the genome of the extremophiles *Galderia sulfuraria* (*G. sulfuraria*) and *Cyanidioschyzon merolae* (*C. merolae).* These organisms are adapted to the particular selection pressure due to their living environment (in hot springs such as in Yellowstone National Park) [40]. Consequently, the absence of this TF family from *G. sulfuraria* and *C. merolae* cannot be taken as a common characteristic of the red lineage.

The E2F/DP family, present in all eukaryotes and known for its involvement in the cell cycle [41], is also equally distributed among algae (around 1 to 3 %).

The MYB family is large, functionally diverse and represented in all eukaryote, such as algae (around 30 %). MYB factors are characterized by a highly conserved DNA-binding domain: the MYB domain. MYB TFs can be divided into different classes depending on the number of adjacent repeats. Three repeats of MYB protein are referred to as R1, R2 or R3, and repeats identified on other related MYB proteins are named in accordance with their similarity with R1, R2 and R3. Although most of these TFs are not functionally characterized in plants, some have been identified as involved in key mechanisms, such as cellular morphogenesis, secondary metabolism, response to biotic and abiotic stresses and signal transduction [42–45]. Finally, the last family equally distributed among algae is the Sigma-70 family. Members of the Sigma-70 family of sigma factors serve as components of the RNA polymerase that direct it to specific promoter elements. In photosynthetic eukaryotes, these Sigma-70 TFs are nuclear encoded and play a role in plastid transcription [46].

Common TF families with different proportions

Four cases of TF families exhibit a difference of proportion between species and are grouped in the cluster number 3 in the Fig. 4. Among these, the C3H type zinc finger family, whose DBD forms a zinc finger, is twice as common in haptophytes and green algae (around 10 %, except for *Pavlova* sp. (5.5 %)) as in stramenopiles and red algae (around 5 %) (Table 3). This protein family is widespread in the tree of life [47–49] and involved in the response to biotic and abiotic stresses [50, 51]. The second family that shows different proportions is the basic leucin-zipper (bZIP) TF family, which accounts for about 2 % in the three haptophytes analyzed in this study, while its proportion is about 10 % in the other algae (*P. tricornutum*: 12.8 %, *N. gaditana*: 11.8 %, *P. purpureum*: 10.6 % and *C. reinhardtii*: 9.4 %).

The third case is that of a particular class of MYB-related TFs: the SHAQKYF-like TFs. This family was described in plants, green algae, as well as in stramenopiles and Amoebozoa [9, 52, 53]. MYB-SHAQKYF is a minority among MYB-rel in *E. huxleyi* and *T. lutea* (2 and 4.7 %, respectively). For *Pavlova* sp. and *C. reinhardtii,* non-negligible amounts of MYB-SHAQKYF were identified among MYB-rel (13.3 and 22.2 %, respectively). In contrast, MYB-SHAQKYF represent almost half of the MYB-rel TFs in the two stramenopiles *P. tricornutum* and *N. gaditana,* as well as in the rhodophyte *P. purpureum* (50, 53.3 and 69.6 %, respectively) (Fig. 5). Such a distribution, together with the presence of such TFs in

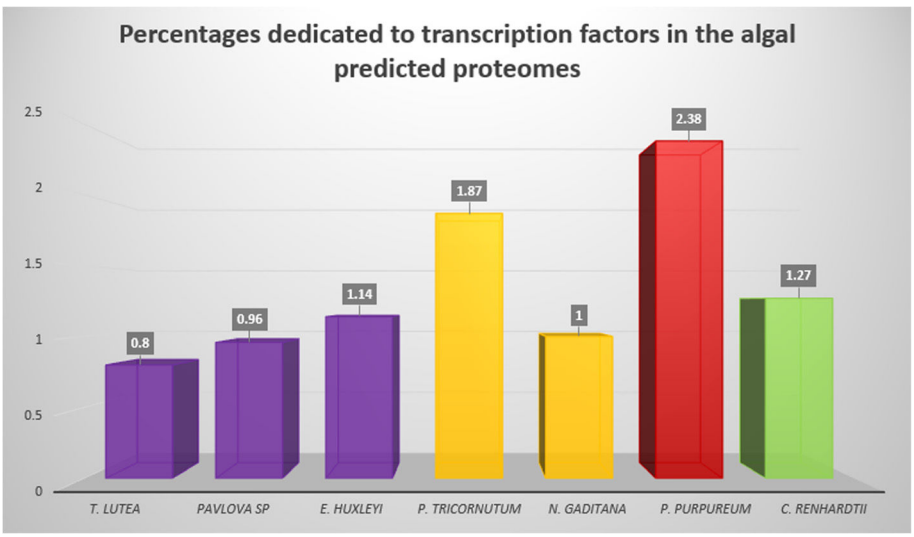

Fig. 2 Percentages of the predicted proteomes dedicated to transcription factors in the 7 algae

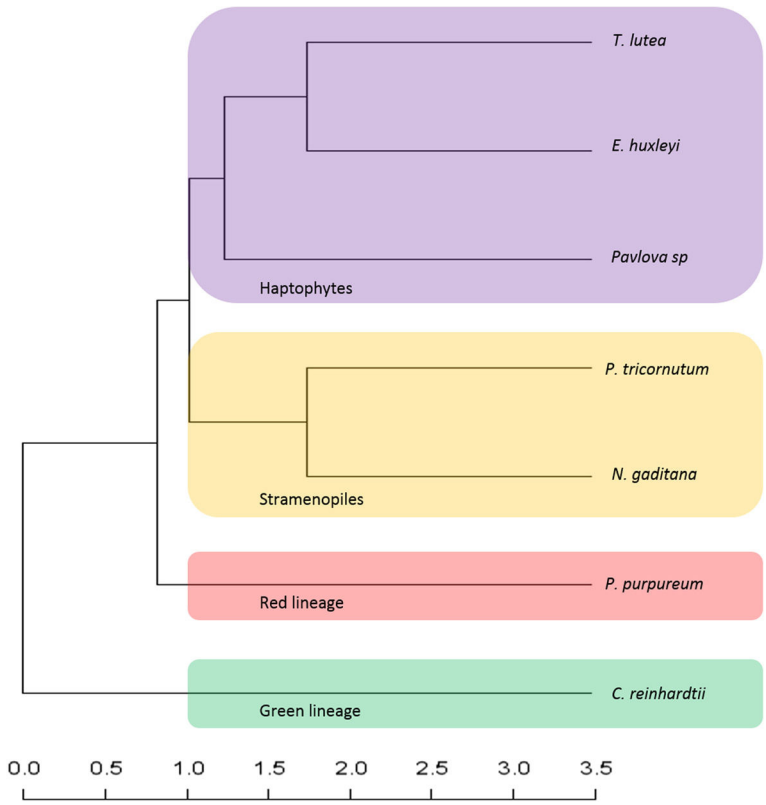

Fig. 3 Dendrogram representing the repartition of the four lineages according to the presence/absence of TF families. The green lineage is colored in green, stramenopiles in orange, red lineage in red and haptophytes in purple. The scale indicates distance measurement

Amoebozoa, suggests that MYB-SHAQKYF proteins have an ancient origin.

Finally, the Nuclear Factor-Y (NF-Y) family, also present in all eukaryotes is divided into three subunits: NF-YA, NF-YB and NF-YC. In plants, three subunits were identified [13]; these TRs are involved in mechanisms as diverse as chloroplast biogenesis, stress response, nodule formation, flowering time control, fatty acid biosynthesis, or response to absisc acid and blue light [54–59]. Subunits NF-YB and NF-YC form a dimer in the cytosol, which is then translocated into the nucleus. The NF-YB/NF-YC dimer interacts in the nucleus with the NF-YA subunit. The functional trimer binds to a *cis*-element called CCAAT-box in the promoter of its target genes [60, 61]. However, no NF-YA subunit was identified in *T. lutea* and *C. reinhardtii*. Such an absence in chlorophyte was previously reported using a similar approach for *C. reinhardtii*, *Volvox carteri* and *Ostreococcus tauri* [13]. The absence of the NF-YA subunit would therefore imply that it is impossible to form the functional trimer. However, it was demonstrated that other TFs are able to interact with the NF-YB and NF-YC subunits. For example, the NF-YB/NF-YC complex can interact with a TF belonging to the C2C2-CO-like family thanks to its CCT domain [62]. Moreover, the interaction between the NF-YB/NF-YC complex and bZIP TFs of *A. thaliana* is sufficient to activate the transcription of target genes, either in the presence or absence of abscisic acid (ABA) [63]. Alternatively, the NF-YB/NF-YC dimer could be active without NF-YA in these taxa.

TF family expansion

During evolutionary history, duplication events occur. Following these duplications, the number of genes of a given family increases. These gene family expansions may be lineage or species specific [64]. Contrary to the other algae in which the MYB family is the most represented, in *P. tricornutum*, *E. huxleyi* and *P. purpureum,* another TF family is more represented because of the expansion phenomenon. In the stramenopile *P. tricornutum*, the Heat Shock Factor family (HSF) was the most represented among the TF families (34.2 % of the TF content) (Table 3). Such a proportion of HSF was previously shown in the diatoms *P. tricornutum*, *T. pseudonana* and *Fistulifera solaris* [9, 65]. This expansion seems to be specific to diatoms since neither *N. gaditana* nor other photosynthetic stramenopiles exhibit such expansion of HSFs [9].

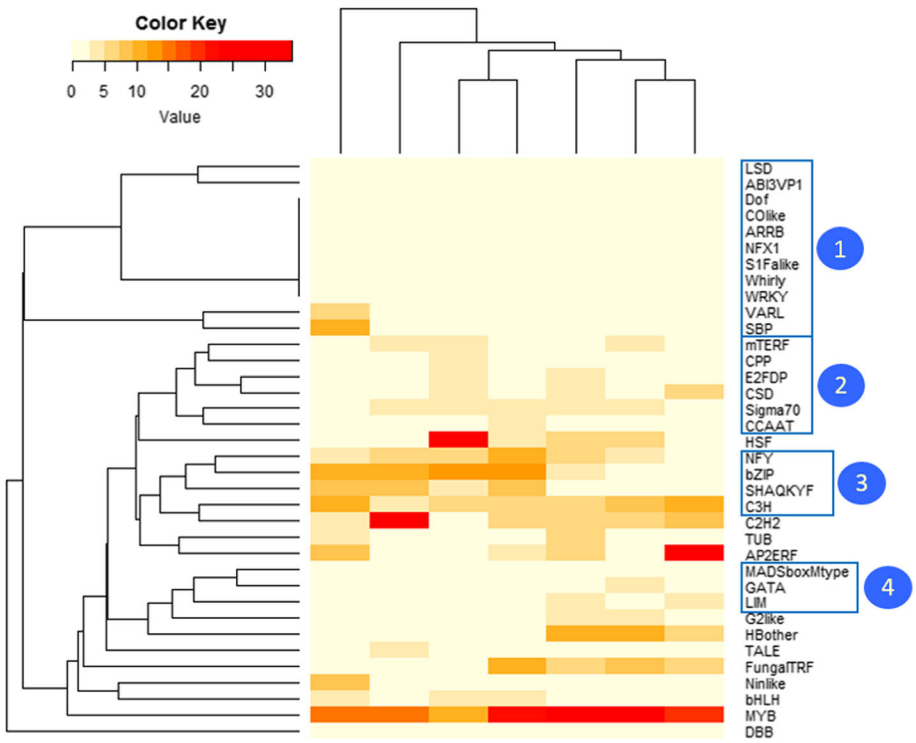

Fig. 4 Heatmap showing the clustering of TF families according to their proportion in the algal genomes. Cluster 1 comprises TF families described as specific to the green lineage. Cluster 2 is composed of families with equivalent proportions across algal genomes. Cluster 3 is composed of families present in the 7 algae but in different proportions. Cluster 4 is composed of 3 families that are absent in stramenopiles

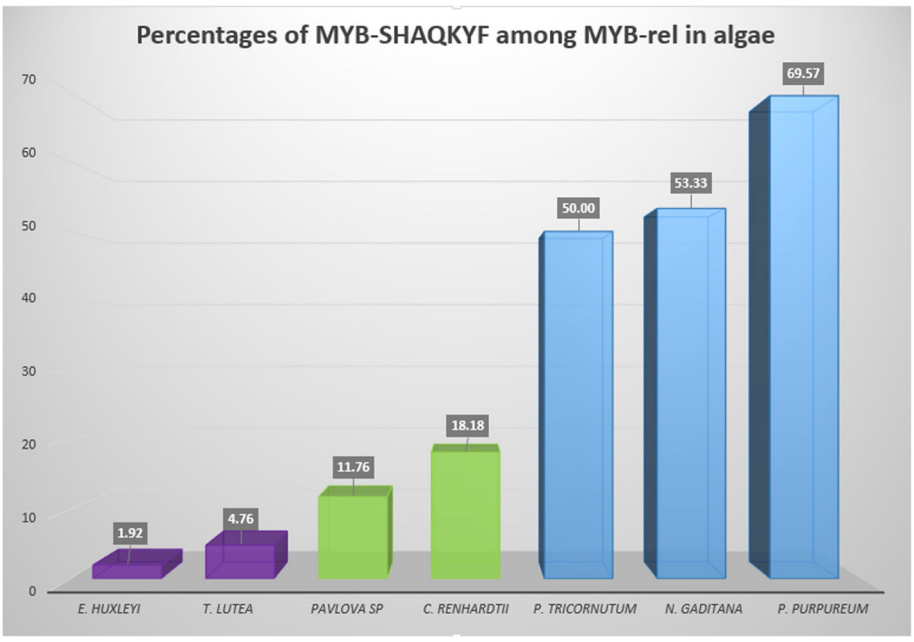

Fig. 5 Percentages of MYB-SHAQKYF among MYB-related TFs in algae

In the haptophyte *E. huxleyi*, the most represented family, accounting for 33 %, is AP2/ERF, involved in growth and development as well as various responses to environmental stimuli. This family was described as specific to the green lineage [15] and its expansion in *E. huxleyi* was also previously described [9]. However, such a proportion of the AP2/ERF is not common to all haptophytes since *T. lutea* and *Pavlova sp*. have AP2/ERF proportions of 1.8 and 5.5 %, respectively, which are close to values recovered for stramenopiles and green algae, respectively. The non-detection of the AP2/ERF family in the Rhodophyta *P. purpureum* is noteworthy, confirming the absence of AP2/ERF in algae belonging to the red lineage [12, 15].

Finally, the C2H2 type zinc finger family was identified as the most represented family in the rhodophyte *P. purpureum*. We found that the C2H2 proportion represents 30.2 % compared to less than 8 % in the other algae. Interestingly, in the two extremophiles, *G. sulfuraria* and *C. merolae*, the C2H2 family was reported to account for less than 5 % [12].

These examples of lineage or species-specific TF expansion illustrate the phenomena that govern the story of TF evolution: gene duplication [66] and diversification through the emergence of lineage-specific families via functional domain shuffling [4, 6, 14, 67]. In the algal world, one of the best examples of lineage-specific TF families is the "green TFs family", which are specific to the green lineage.

Lineage-specific TF families

Are TF families specific to the green lineage highly specific? Previous comparative studies of the TF content of diverse photosynthetic organisms reveal that some TF families are specific to the green lineage because of their absence from red microalgae [11, 15]. Among all green lineage-specific TF families identified in this study, only nine families were present in the green algae *C. reinhardtti*: NF-X1, S1Fa-like, SBP, VARL, Whirly, WRKY, GARP-ARR-B, C2C2-CO-like and C2C2-Dof (Table 3). However, some TF families previously described as specific to the green lineage were also identified in haptophytes, stramenopiles or in the rhodophyte *P. purpureum*. First of all, one TF belonging to the ABI3/VP1 family was identified in *T. lutea* and the C2C2-LSD family have one member in both *T. lutea* and *Pavlova sp*. In the heatmap (Fig. 4), these two TF families are clustered with the nine families only identified in *C. reinhardtii*. Moreover, the CSD family was identified in all predicted proteomes and the AP2/ERF and TUB families are absent in *P. purpureum,* but present in the six other algae. Another interesting finding is the unique identification of a member of the Double B-box (DBB) family in *P. purpureum*. This family had only previously been identified in land plants [68] and was thought to be involved in light signal transduction mechanisms, such as early photomorphogenic development of *A. thaliana* [69–72].

This presence of "green TFs" in algae that do not belong to the green lineage could be explained either (i) by a loss of these families during evolutionary history of rhodophytes, or (ii) by the acquisition of these families by horizontal gene transfer from a green algal endosymbiont to the nuclear genome. This last hypothesis is consistent with the endosymbiosis of a green and a red alga in the evolutionary history of haptophytes and stramenopiles [19].

Specific features of stramenopiles The stramenopiles *P. tricornutum* and *N. gaditana* are distinguished by the absence of the C2C2-GATA family and the MADS-box family, which are involved in plant homeotic functions [73–75] (Table 3). These results confirm those of Rayko et al. [9] for stramenopile micro- and macro-algae. Moreover, our results also highlight the absence of TFs from the LIM family in stramenopiles, while LIM TFs are present in all other studied algae. LIM, C2C2-GATA and MADS-box families are clustered together in Fig. 4. To examine whether these features are shared by other stramenopiles not investigated in this work, a specific research of LIM, MADS-box and C2C2-GATA TFs was carried out in the two diatoms *Pseudo-nitzschia multiseries* and *Fragilariopsis cylindrus*. No member of these families was identified (data not shown). By contrast, the MADS-box, C2C2-GATA and LIM families were identified in *P. purpurem* and *C. reinhardtii* (this study), as well as in other chlorophytes and rhodophytes (the green algae *Bathycoccus prasinos*, *Micromonas pusilla*, *Micromonas sp*, *Ostreococcus lucimarinus*, *Ostreococcus sp*, *Ostreococcus tauri* and *Volvox carteri*; the red algae *C. merolae* and *G. sulfuraria*) [12, 13]. This repartition suggests that the MADS-box, C2C2-GATA and LIM families were present in the hypothetical ancestor of the algae and secondarily lost in stramenopiles.

Another feature of stramenopiles concerns some particular combinations of functional domains. Two domain associations shared by both stramenopiles *N. gaditana* and *P. tricornutum* were identified. The first is composed of a bHLH domain and a PAS domain (named after the three first sequences in which it was identified (Per, Arnt, Sim)) and the second by a bZIP and LOV (Light, Oxygen, Voltage) domain combination. The bHLH-PAS TFs are well known in vertebrate TFs in which two PAS domains are present, contrary to the stramenopile sequences that have only one PAS [9, 76]. In vertebrates, the PAS domains are involved in the dimerization of PAS domains containing TFs, such as

the Hypoxia Inducible Factor [77, 78]. The presence of bHLH and PAS domains in the same sequence in both vertebrates and stramenopiles may be an example of convergent evolution, which suggests that this fusion occurred in a parallel fashion in different lineages.

The second stramenopile specific combination is that of the bZIP and LOV domains. These sequences, called aureochromes, are an atypical case that couple both blue light receptor and transcription factor functions [79]. We identified three and four aureochromes in *N. gaditana* and in *P. tricornutum*, respectively. Such sequences have only been identified in photosynthetic stramenopiles [9, 79–82]. In marine environments, the sea water absorbs wavelengths other than blue, which are the only wavelengths to travel long distances within the water column [83]. Blue light is thus expected to play an important role in algae, as suggested by the involvement of aureochromes in key mechanisms such as the cell cycle [84]. Moreover, mechanisms like photomorphogenesis and phototropism observed in algae [85] are influenced in land plants by phototropins [86]. These are blue light receptors harboring two LOV domains and have a role in signal transduction [87]. Thus, aureochromes are lineage-specific TFs evolved by photosynthetic stramenopiles that confer an adaptive capacity for success in an aquatic environment.

Specific features of haptophytes The bHLH TFs were identified in the predicted proteome of *P. tricornutum*, *N. gaditana*, *C. reinhardtii* and *P. purpureum,* but not in the three haptophytes (Table 3). Nevertheless, bHLH is one of the most widespread TF families in eukaryotes and the second most represented in plants [13, 88]. This repartition suggests that the bHLH TF family was secondarily lost in *T. lutea*, *E. huxleyi* and *Pavlova* sp. These results confirm previous conclusions derived from the comparison of the TF content composition of six stramenopiles with *E. huxleyi* [9], and extends the number of haptophyte organisms sharing this common absence of bHLH families.

Interestingly, we identified two and four Heat Shock transcription factors (HSFs) in *E. huxleyi* and *T. lutea*, respectively, that share the association of a HSF DBD with a PAS domain. Moreover, two other HSF proteins, harboring two PAS domains, were identified only in *T. lutea*.

The HSF domain is known for playing a role in stress perception in all categories of living organisms [89]. Its sensor function is applied to stimuli such as light, oxygen or redox potential. Such stimuli are also known to induce HSF expression. In plants in particular, HSFs are involved in response to oxidative stress and redox stat changes [90, 91]. This functional convergence led us to hypothesize that the sensor function of the PAS domain may play a role in the detection of stimuli involved in HSF activation. The PAS domain also enables protein-protein interactions, especially with other PAS-containing proteins [89, 92]. This function may stabilize the homotrimer formed by activated HSFs. Likewise, four TFs have the undescribed association of a PAS domain and a homeobox domain in *T. lutea*.

Potential gene transfer cases

Identification of cyanobacterial TFs in the nuclear genome of algae

Remarkably, our TFs prediction pipeline allowed the identification of cyanobacterial TFs in the predicted proteome of all the microalgae studied (Table 4). We investigated whether the presence of these genes could be due to bacterial contamination, and if not, whether these genes are localized in the nuclear, chloroplastic or mitochondrial genome. Because information concerning bacterial contamination are only available for *T. lutea* (G. Carrier, pers. Com.), *Pavlova* sp. (transcriptomic data) and *C. reinhardtii* (JGI portal), it only was possible to answer the contamination question for these three algae. It allowed us to conclude that *T. lutea*, *Pavlova* sp. and *C. reinhardtii* cyanobacterial TFs identification are not due to bacterial contamination. Concerning the localization of the cyanobacterial TFs in the algae, we cannot draw any conclusions for *Pavlova* sp., for which no mitochondrial or chloroplastic genome are available. For *P. purpureum* the TFs are not localized in the chloroplastic genome; however, since the mitochondrial genome is not available, we cannot make a conclusion about a mitochondrial localization. We found that these TFs are nuclear genes for *T. lutea*, *E. huxleyi*, *P. tricornutum*, *N. gaditana* and *C. reinhardtii*.

Only one TF belonging to the arsenic resistance operon regulator (arsR) family was identified, in *N. gaditana*. This family is involved in stress response to metal ions in cyanobacteria [93]. Considering the Bac_DNA_-Binding family, one member was identified in all the algae except in *P. purpureum*. This protein family is involved in transcription regulation, transposition and DNA chaperones [94, 95]. Several members of the BolA family were identified in all algae. BolA is a widespread family identified in all groups of the tree of life [2] and is involved in cell cycle regulation and abiotic stress response in cyanobacteria [96]. The GerE family which is part of a two component response regulator was only identified in haptophytes *T. lutea*, *E. huxleyi* (except for *Pavlova sp.*), and in the two stramenopiles *N. gaditana* and *P. tricornutum*. This family is characterized by the presence of a LuxR DBD and involved in processes such as signal transduction [97], quorum sensing [98] and sporulation [99]. One member of LysR protein was identified in *N. gaditana*. In cyanobacteria, this family is

Table 4 Number of cyanobacterial transcription factors (TFs) identified in the seven algae for each TF family

TF family	*T. lutea*	*Pavlova* sp	*E. huxleyi*	*P. tricornutum*	*N. gaditana*	*P. purpureum*	*C. reinhardtii*
arsR	0	0	0	0	1	0	0
Bac_DNA_binding	1	1	1	1	1	0	1
BolA	2	2	7	4	4	3	5
GerE	1	0	2	2	1	0	0
LysR	0	0	0	0	1	0	0
SfsA	1	3	2	0	1	2	2

BolA TFs were previously identified in the chlorophyte *C. reinhardtii*, the rhodophyte *Cyanidoschyzon merolae*, the diatom *Thalassiosira pseudonana* and the cryptophyte *Guillardia theta* [24]

involved in CO_2 fixation [100] and nitrate assimilation [101]. Finally, the SfsA family was identified in all algae except *P. tricornutum*. SfsA TF is known to be involved in sugar fermentation [102].

So far, no genome-wide TF identification study has shown the presence of such sequences in microalgae, except for the BolA family in the chlorophyte *C. reinhardtii*, the diatom *Thalassiosira pseudonana*, the rhodophyte *C. merolae* and the cryptophyte *Guillardia theta* [2]. Since these TF families are found either in cyanobacteria or bacteria, their presence in the algal genomes could be explained either by an endosymbiotic gene transfer (EGT), which is a gene transfer taking place from the chloroplastic genome to the nuclear genome during evolutionary history [103, 104], or a horizontal gene transfer (HGT) from a prokaryotic organism to the algal genome [105].

Fungal TRF: fungus in algae

The TF families described above are of bacterial type, but TFs from the fungal TRF family (also called Zn-clus) were also identified. These TFs are abundant and well described in fungi [106]. Their DBD is characterized by a conserved CysX2CysX6CysX5–16CysX2CysX6–8Cys motif. The six conserved cysteines coordinate two Zn(II) ions allowing correct folding of the domain [107]. This DBD was first identified in the *Saccharomyces cerevisiae* Gal4 TF [108]. Members of this TF family are implicated in the regulation of genes involved in diverse mechanisms, such as amino acid biosynthesis [109], multidrug resistance [110], ethanol catabolism [111] or lipid catabolism [112, 113].

Fungal TRF were identified in *T. lutea*, *Pavlova sp.*, *E. huxleyi*, *N. gaditana* and *P. tricornutum*. However, no fungal TRF were identified in either *C. reinhardtii* or in *P. purpureum*. In previous studies TFs from this family were identified in the rhodophyte *G. sulfuraria* [12, 15].

This presence of fungal type TFs in algal genomes is another illustration of the complex evolutionary history of algae [114]. Multiple endosymbiosis resulting in the algal diversity [18] is punctuated by numerous gene transfer events. These gene transfer events comprise both EGT [115, 116], as the original case of HGT from bacteria to the plastid genome [117], or from bacteria or archaebacteria to the nuclear genome [40, 105, 118, 119]. In these HGT, the donor organism is prokaryotic, but interesting cases of HGT from a fungus to an alga were recently shown [120]. All these gene transfers give rise to metabolic and regulatory diversity, leading to adaptation of algae to a wide variety of environments and conditions.

Conclusion

Using a pipeline with very good sensitivity and PPV for both plant and cyanobacterial TFs, we undertook the first genome-wide identification of TFs in haptophytes, coupled with a comparison of TF content between haptophytes and other algal lineages. The identification highlighted the presence of cyanobacterial TFs in algal nuclear genomes, which is likely to originate from either an EGT or an HGT. Moreover, members of the Fungal TRF family were identified in *T. lutea*, *Pavlov* asp, *E. huxleyi*, *P. tricornutum* and *N. gaditana*. The presence of fungal type TFs in algal genomes also illustrates the complex evolutionary history of these organisms. This comparison study confirms and extends lineage-specific features highlighted between haptophytes and stramenopiles by previous work [9] and extends the panel of genomes used for this comparison (Fig. 6). In order to investigate the evolutionary history of organisms and genome-wide studies, some gaps need to be filled and the red algae are one of them. In this kind of study, the only two red algae used are the two extremophiles *G. sulfuraria* and *C. merolae*. The extreme environmental pressures they face make these two algae peculiar cases that should not be considered representative of the red lineage. Here, we used mesophilic species *P. purpureum*. Availability of genomic data from haptophytes is also lacking. In this study, we provide the first genomic data of *T. lutea*. The characteristics revealed include some clues consistent with the hypothesis of an endosymbiosis of green and red algae in the evolutionary history of haptophytes and stramenopiles [19]. Therefore, this work provides a basis to better understand gene

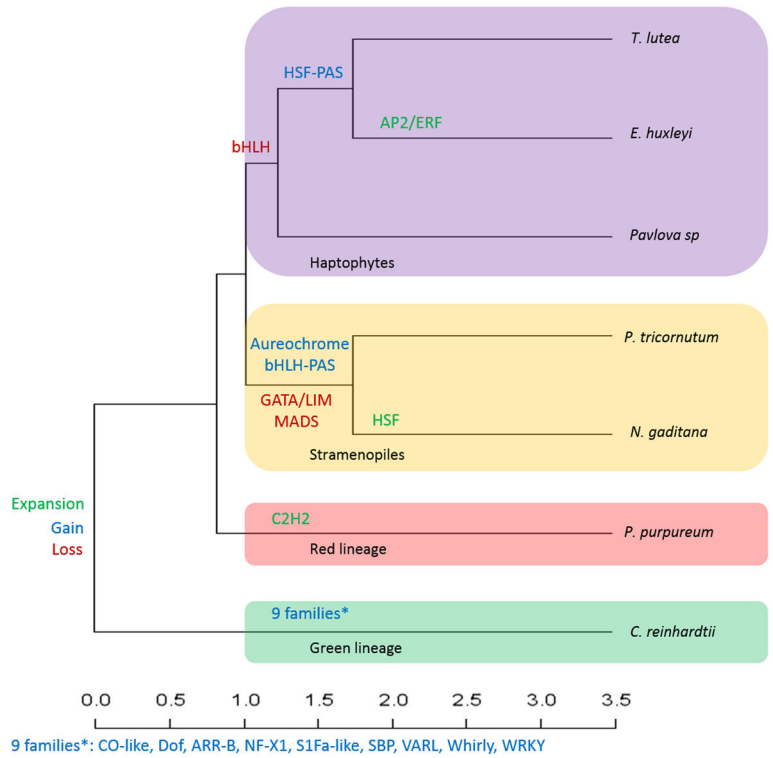

Fig. 6 Expansion, gain and loss of TF families during the evolutionary history of microalgae

regulation in *T. lutea,* which is a species of ecological interest as part of haptophytes, a diverse and often ecologically dominant group in the planktonic photic realm [121].

Methods

Source datasets

The predicted proteomes used in this study were downloaded from different sources (Additional file 1: Table S1). The *C. reinhardtii* CC-503 cw92 mt+, *P. tricornutum* CCAP1055/1 and *E. huxleyi* CCMP1516 predicted proteomes were downloaded from the JGI genome portal at http://genome.jgi.doe.gov/. The *N. gaditana* CCMP526, *Pavlova sp.* CCMP459 and *P. purpureum* DBLAB2 predicted proteomes were downloaded from http://nannochloropsis.genomeprojectsolutions-databases.com/, http://data.imicrobe.us/project/view/104 and http://cyanophora.rutgers.edu/porphyridium/, respectively. The genome of the *T. lutea* CCAP927/14 strain was recently sequenced and annotated in our laboratory (data not shown). Raw read data are available at SRA (RUN: SRR3156597).

Identification and classification of transcription factors

The TF identification and classification pipeline was calibrated with the model plant *A. Thaliana* (TAIR 10). Overall, the pipeline uses two strategies: (1) a similarity research with BLAST software against a self-built database of known TFs from algae, *A. thaliana*, *Saccharomyces cerevisiae* and cyanobacteria; (2) identification of TF DBDs with InterProScan and HMMER software. The compilation of software results allowed us to obtain a putative list of TFs (Fig. 1).

Construction of a TF database for BLAST software

The TF database is composed of TFs from different organisms (the model plant *A. thaliana*; the green algae *Bathycoccus prasinos*, *Chlorella sp*, *Coccomyxa sp*, *Micromonas pusilla*, *Micromonas sp*, *Ostreococcus lucimarinus*, *Ostreococcus sp*, *Ostreococcus tauri* and *Volvox carteri*; the red algae *Cyanidioschyzon merolae* and *Galdieria sulfuraria*; the diatom *Thalassiosira pseudonana* and the yeast *Saccharomyces cerevisiae*). These sequences were retrieved from online databases (Additional file 2: Table S2). Since algae originate from the engulfment of a cyanobacteria-like organism by a primitive eukaryotic heterotroph, we added all cyanobacterial TFs of the cTFbase [8] to the self-built database.

Identification of protein functional domains

Each protein domain contained in the protein domain databases is stored as a Hidden Markov Model (HMM) and linked to a putative function. This statistical method computes a matrix based on the multiple alignments of

a protein domain [122]. For functional domain annotation of all the predicted proteomes, we employed InterProScan 5 version 5.4-47.0 [123], which uses a consortium of eleven protein domain databases (PROSITE, HAMP, Pfam, PRINTS, ProDom, SMART, TIGRFAMs, PIRSF, SUPERFAMILY, CATH-Gene3D and PANTHER). However, twelve DBDs (G2-like, BELL, HD-ZIP, HRT, NF-YB, NF-YC, SAP, STAT, Trihelix, VOZ, WOX and VARL) are not supported by the eleven databases of the consortium and were added through multiple alignments available in the TF databases PlantTFDB [13] and PlnTFDB [12] with HMMER3, v3.1b1 [124].

Pipeline description

First step Sequences of each predicted proteome were analyzed in parallel by HMMER (*hmmscan*, default parameters), InterProScan (default parameters) for protein functional domains and by BLAST (*e-value* threshold 10^{-10}) for a similarity search against known TFs (Fig. 1).

Second step The results of each software analysis were filtered using different homemade PERL scripts. For InterProScan, false positives were filtered out to keep only annotated domains that had an *e-value* above or equal to 10^{-3}. Among these, only TFs DBDs were conserved. For HMMER, filtration was done on the score value. Sequences with a significant *hmmscan* match (according to the database thresholds) were added as TF candidates. For BLAST searches, the filtering step was applied with an identity percentage threshold of 35 % and an alignment length threshold of 100 residues. Then, the best-BLAST hit was taken for each query. Finally, the results of all software processes were combined in one file.

Third step Once identified, putative TFs were classified into specific families according to their DBD(s). We used a compilation of the "family assignment rules" described by the web databases PlantTFDB [13], PlnTFDB [12] and cTFbase [8], as well as previous studies [9, 11]. A PERL script was used to automatically classify the putative TFs in families following the assignment rules.

Final step Manual curation was necessary, in particular for three complex cases: (1) MYB, where the calibration stage revealed that filtration of the *e-value* score generated false negatives. To overcome this, MYB identification was performed using the same protocol, with the exception of the validation step of the *e-value* scores on the InterProScan result. Moreover, each candidate was manually inspected (BLAST) to confirm each MYB domain and classify putative TFs in each family (MYB-3R, MYB-2R and MYB-related). (2) G2-like, due to the absence of a G2-like domain in the InterProScan database and its close similarity to the MYB-SHAQKYF domain, cross-annotation between these two domains was manually checked using HMMER. (3) TF families characterized by the repetition of a single domain; for proteins identified as belonging to the DBB and AP2/ERF families, the presence of two or more B-Box or AP2/ERF domains, respectively, was verified.

Evaluation of pipeline accuracy

To estimate the accuracy and reliability of our identification method, we applied our pipeline to the predicted proteome of *A. thaliana* (TAIR 10) and compared the identification of eleven well-annotated families to published datasets [13], used as a gold standard. For the identification of cyanobacterial TFs, we applied our pipeline to *Synechocystis* sp. PCC 6803 (GeneBank Assembly: GCA_000009725.1), *Synechococcus* sp. CC9605 (downloaded from cyanobase) and *Nostoc punctiforme* PCC73102 (GeneBank Assembly: GCA_000020025.1) predicted proteomes and compared our prediction results with published data [8]. The accuracy was evaluated by the measurement of sensitivity:

$$\frac{True\ positives}{True\ positives + False\ negatives}$$

and Positive Predictive Value (PPV):

$$\frac{True\ positives}{True\ positives + False\ positives}$$

A sensitivity value of less than one means inclusion of false negatives and a PPV of less than one means inclusion of false positives.

Availability of data and material

The datasets supporting the conclusions of this article are included within the article (and its additional files).

Additional files

Additional file 1: Table S1. Source datasets. Table listing the reference of the genomic data used in this study. (XLSX 11 kb)

Additional file 2: Table S2. Sources for the self-build TF database. Table listing sources for the building of the self-build transcription factor database. (XLSX 11 kb)

Abbreviations

DBD: DNA binding domain; EGT: endosymbiotic gene transfer; HGT: horizontal gene transfer; HMM: Hidden Markov Model; TFs: transcription factor; TRs: transcriptional regulators.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions
STR elaborated the pipeline, carried out the identification and the comparative study and drafted the manuscript. GC participated in the coordination of the study, the assemblage of the genome and helped to draft the manuscript. CT carried out the automation of the pipeline. BC and BS participated in the coordination and helped to draft the manuscript. JPC and GB participated in the design and coordination of the study. BSJ participated in the design of the study and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgment
The authors are grateful to the anonymous reviewers for their critical comments, which have greatly improved the manuscript. Thanks to Ms Deborah McCombie for the English reviewing of the manuscript. This work was supported by the French region of Pays de la Loire and the French Research Institute for Exploitation of the Sea (IFREMER).

Author details
[1]IFREMER, Physiology and Biotechnology of Algae Laboratory, rue de l'Ile d'Yeu, 44311 Nantes, France. [2]MicroMar, Mer Molécules Santé, IUML - FR 3473 CNRS, University of Le Mans, Le Mans, France.

Received: 31 October 2015 Accepted: 5 April 2016
Published online: 11 April 2016

References

1. Heydarizadeh P, Marchand J, Chenais B, Sabzalian MR, Zahedi M, Moreau B, Schoefs B: Functional investigations in diatoms need more than a transcriptomic approach. Diatom Res. 2014;29:75–89.
2. Richardt S, Lang D, Reski R, Frank W, Rensing SA. PlanTAPDB, a phylogeny-based resource of plant transcription-associated proteins. Plant Physiol. 2007; 143:1452–66.
3. Luscombe NM, Austin SE, Berman HM, Thornton JM. An overview of the structures of protein-DNA complexes. Genome Biol. 2000;1:REVIEWS001.
4. Charoensawan V, Wilson D, Teichmann SA. Lineage-specific expansion of DNA-binding transcription factor families. Trends Genet. 2010;26:388–93.
5. Aravind L, Koonin EV. DNA-binding proteins and evolution of transcription regulation in the archaea. Nucleic Acids Res. 1999;27:4658–70.
6. Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu G: Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science. 2000;290: 2105–10.
7. Martínez-Bueno M, Molina-Henares AJ, Pareja E, Ramos JL, Tobes R. BacTregulators: a database of transcriptional regulators in bacteria and archaea. Bioinforma Oxf Engl. 2004;20:2787–91.
8. Wu J, Zhao F, Wang S, Deng G, Wang J, Bai J, et al. cTFbase: a database for comparative genomics of transcription factors in cyanobacteria. BMC Genomics. 2007;8:104.
9. Rayko E, Maumus F, Maheswari U, Jabbari K, Bowler C. Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. New Phytol. 2010;188:52–66.
10. Zhang H-M, Chen H, Liu W, Liu H, Gong J, Wang H, et al. AnimalTFDB: a comprehensive animal transcription factor database. Nucleic Acids Res. 2012;40(Database issue):D144–9.
11. Lang D, Weiche B, Timmerhaus G, Richardt S, Riano-Pachon DM, Correa LGG, et al. Genome-Wide Phylogenetic Comparative Analysis of Plant Transcriptional Regulation: A Timeline of Loss, Gain, Expansion, and Correlation with Complexity. Genome Biol Evol. 2010;2:488–503.
12. Pérez-Rodríguez P, Riaño-Pachón DM, Corrêa LGG, Rensing SA, Kersten B, Mueller-Roeber B. PlnTFDB: updated content and new features of the plant transcription factor database. Nucleic Acids Res. 2010;38(Database issue):D822–7.
13. Jin J, Zhang H, Kong L, Gao G, Luo J. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. Nucleic Acids Res. 2014;42:D1182–7.
14. Charoensawan V, Wilson D, Teichmann SA. Genomic repertoires of DNA-binding transcription factors across the tree of life. Nucleic Acids Res. 2010; 38:7364–77.
15. Sharma N, Bhalla PL, Singh MB. Transcriptome-wide profiling and expression analysis of transcription factor families in a liverwort, Marchantia polymorpha. BMC Genomics. 2013;14:915.
16. Delwiche CF. Tracing the Thread of Plastid Diversity through the Tapestry of Life. Am Nat. 1999;154:S164–77.
17. Keeling PJ. Diversity and evolutionary history of plastids and their hosts. Am J Bot. 2004;91:1481–93.
18. Archibald JM. The puzzle of plastid evolution. Curr Biol CB. 2009;19:R81–8.
19. Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D. Genomic Footprints of a Cryptic Plastid Endosymbiosis in Diatoms. Science. 2009;324:1724–6.
20. Wang D, Ning K, Li J, Hu J, Han D, Wang H, et al. Nannochloropsis Genomes Reveal Evolution of Microalgal Oleaginous Traits. PLoS Genet. 2014;10:e1004094.
21. Burki F, Okamoto N, Pombert J-F, Keeling PJ. The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. Proc Biol Sci. 2012;279:2246–54.
22. Andersen RA. Biology and systematics of heterokont and haptophyte algae. Am J Bot. 2004;91:1508–22.
23. Iida K, Seki M, Sakurai T, Satou M, Akiyama K, Toyoda T, et al. RARTF: database and tools for complete sets of Arabidopsis transcription factors. DNA Res Int J Rapid Publ Rep Genes Genomes. 2005;12:247–56.
24. Riaño-Pachón DM, Ruzicic S, Dreyer I, Mueller-Roeber B. PlnTFDB: an integrative plant transcription factor database. BMC Bioinformatics. 2007;8:42.
25. Guo A-Y, Chen X, Gao G, Zhang H, Zhu Q-H, Liu X-C, et al. PlantTFDB: a comprehensive plant transcription factor database. Nucleic Acids Res. 2008; 36(Database issue):D966–9.
26. Messina DN, Glasscock J, Gish W, Lovett M. An ORFeome-based Analysis of Human Transcription Factor Genes and the Construction of a Microarray to Interrogate Their Expression. Genome Res. 2004;14:2041–7.
27. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, et al. The Genome Sequence of Drosophila melanogaster. Science. 2000; 287:2185–95.
28. Doebley J, Lukens L. Transcriptional Regulators and the Evolution of Plant Form. Plant Cell Online. 1998;10:1075–82.
29. Lespinet O, Wolf YI, Koonin EV, Aravind L. The role of lineage-specific gene family expansion in the evolution of eukaryotes. Genome Res. 2002;12:1048–59.
30. Nitta KR, Jolma A, Yin Y, Morgunova E, Kivioja T, Akhtar J, et al. Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. eLife. 2015;4:e04837.
31. Carroll SB. Chance and necessity: the evolution of morphological complexity and diversity. Nature. 2001;409:1102–9.
32. Van Nimwegen E. Scaling laws in the functional content of genomes. Trends Genet. 2003;19:479–84.
33. Vogel C, Chothia C. Protein Family Expansions and Biological Complexity. PLoS Comput Biol. 2006;2:e48.
34. Wendel JF. Genome evolution in polyploids. Plant Mol Biol. 2000;42:225–49.
35. Paterson AH, Chapman BA, Kissinger JC, Bowers JE, Feltus FA, Estill JC. Many gene and domain families have convergent fates following independent whole-genome duplication events in Arabidopsis, Oryza, Saccharomyces and Tetraodon. Trends Genet TIG. 2006;22:597–602.
36. Edger PP, Pires JC. Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosome Res Int J Mol Supramol Evol Asp Chromosome Biol. 2009;17:699–717.
37. Carretero-Paulet L, Galstyan A, Roig-Villanova I, Martínez-García JF, Bilbao-Castro JR, Robertson DL. Genome-Wide Classification and Evolutionary Analysis of the bHLH Family of Transcription Factors in Arabidopsis, Poplar, Rice, Moss, and Algae. Plant Physiol. 2010;153:1398–412.
38. Shalchian-Tabrizi K, Reier-Røberg K, Ree DK, Klaveness D, Bråte J. Marine-freshwater colonizations of haptophytes inferred from phylogeny of environmental 18S rDNA sequences. J Eukaryot Microbiol. 2011;58:315–8.
39. Bendif EM, Probert I, Schroeder DC, de Vargas C. On the description of Tisochrysis lutea gen. nov. sp. nov. and Isochrysis nuda sp. nov. in the Isochrysidales, and the transfer of Dicrateria to the Prymnesiales (Haptophyta). J Appl Phycol. 2013;25:1763–76.
40. Schönknecht G, Chen W-H, Ternes CM, Barbier GG, Shrestha RP, Stanke M, et al. Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. Science. 2013;339:1207–10.
41. Inzé D, De Veylder L. Cell Cycle Regulation in Plant Development 1. Annu Rev Genet. 2006;40:77–105.
42. Ito M, Araki S, Matsunaga S, Itoh T, Nishihama R, Machida Y, et al. G2/M-phase-specific transcription during the plant cell cycle is mediated by c-Myb-like transcription factors. Plant Cell. 2001;13:1891–905.
43. Yoshioka S, Taniguchi F, Miura K, Inoue T, Yamano T, Fukuzawa H. The novel Myb transcription factor LCR1 regulates the CO2-responsive gene Cah1,

encoding a periplasmic carbonic anhydrase in Chlamydomonas reinhardtii. Plant Cell. 2004;16:1466–77.
44. Zhao L, Gao L, Wang H, Chen X, Wang Y, Yang H, et al. The R2R3-MYB, bHLH, WD40, and related transcription factors in flavonoid biosynthesis. Funct Integr Genomics. 2013;13:75–98.
45. Pattanaik S, Patra B, Singh SK, Yuan L. An overview of the gene regulatory network controlling trichome development in the model plant, Arabidopsis. Front Plant Sci. 2014;5:259.
46. Allison LA. The role of sigma factors in plastid transcription. Biochimie. 2000; 82:537–48.
47. De J, Lai WS, Thorn JM, Goldsworthy SM, Liu X, Blackwell TK, Blackshear PJ: Identification of four CCCH zinc finger proteins in Xenopus, including a novel vertebrate protein with four zinc fingers and severely restricted expression. Gene. 1999;228:133–45.
48. Chai G, Hu R, Zhang D, Qi G, Zuo R, Cao Y, et al. Comprehensive analysis of CCCH zinc finger family in poplar (Populus trichocarpa). BMC Genomics. 2012;13:253.
49. Yeh P-A, Yang W-H, Chiang P-Y, Wang S-C, Chang M-S, Chang C-J. Drosophila eyes absent is a novel mRNA target of the tristetraprolin (TTP) protein DTIS11. Int J Biol Sci. 2012;8:606–19.
50. Peng X, Zhao Y, Cao J, Zhang W, Jiang H, Li X, et al. CCCH-type zinc finger family in maize: genome-wide identification, classification and expression profiling under abscisic acid and drought treatments. PLoS ONE. 2012;7:e40120.
51. Deng H, Liu H, Li X, Xiao J, Wang S. A CCCH-type zinc finger nucleic acid-binding protein quantitatively confers resistance against rice bacterial blight disease. Plant Physiol. 2012;158:876–89.
52. Schaffer R, Ramsay N, Samach A, Corden S, Putterill J, Carré IA, Coupland G: The late elongated hypocotyl Mutation of Arabidopsis Disrupts Circadian Rhythms and the Photoperiodic Control of Flowering. Cell. 1998;93:1219–29.
53. Ehrenkaufer GM, Hackney JA, Singh U. A developmentally regulated Myb domain protein regulates expression of a subset of stage-specific genes in Entamoeba histolytica. Cell Microbiol. 2009;11:898–910.
54. Miyoshi K, Ito Y, Serizawa A, Kurata N. OsHAP3 genes regulate chloroplast biogenesis in rice. Plant J. 2003;36:532–40.
55. Combier J-P, Frugier F, de Billy F, Boualem A, El-Yahyaoui F, Moreau S, Vernié T, Ott T, Gamas P, Crespi M, Niebel A: MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in Medicago truncatula. Genes Dev. 2006;20:3084–8.
56. Warpeha KM, Upadhyay S, Yeh J, Adamiak J, Hawkins SI, Lapik YR, Anderson MB, Kaufman LS: The GCR1, GPA1, PRN1, NF-Y Signal Chain Mediates Both Blue Light and Abscisic Acid Responses in Arabidopsis. Plant Physiol. 2007; 143:1590–600.
57. Cai X, Ballif J, Endo S, Davis E, Liang M, Chen D, DeWald D, Kreps J, Zhu T, Wu Y: A Putative CCAAT-Binding Transcription Factor Is a Regulator of Flowering Timing in Arabidopsis. Plant Physiol. 2007;145:98–105.
58. Nelson DE, Repetti PP, Adams TR, Creelman RA, Wu J, Warner DC, Anstrom DC, Bensen RJ, Castiglioni PP, Donnarummo MG, Hinchey BS, Kumimoto RW, Maszle DR, Canales RD, Krolikowski KA, Dotson SB, Gutterson N, Ratcliffe OJ, Heard JE: Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. Proc Natl Acad Sci. 2007;104:16450–5.
59. Mu J, Tan H, Zheng Q, Fu F, Liang Y, Zhang J, et al. LEAFY COTYLEDON1 Is a Key Regulator of Fatty Acid Biosynthesis in Arabidopsis. Plant Physiol. 2008; 148:1042–54.
60. Frontini M, Imbriano C, Manni I, Mantovani R. Cell cycle regulation of NF-YC nuclear localization. Cell Cycle Georget Tex. 2004;3:217–22.
61. Kahle J, Baake M, Doenecke D, Albig W. Subunits of the Heterotrimeric Transcription Factor NF-Y Are Imported into the Nucleus by Distinct Pathways Involving Importin β and Importin 13. Mol Cell Biol. 2005;25: 5339–54.
62. Wenkel S, Turck F, Singer K, Gissot L, Gourrierec JL, Samach A, Coupland G: CONSTANS and the CCAAT Box Binding Complex Share a Functionally Important Domain and Interact to Regulate Flowering of Arabidopsis. Plant Cell Online. 2006;18:2971–84.
63. Yamamoto A, Kagaya Y, Toyoshima R, Kagaya M, Takeda S, Hattori T. Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. Plant J. 2009; 58:843–56.
64. Jacquemin J, Ammiraju JSS, Haberer G, Billheimer DD, Yu Y, Liu LC,Rivera LF, Mayer K, Chen M, Wing RA: Fifteen million years of evolution in the Oryza genus shows extensive gene family expansion. Mol Plant. 2014;7:642–56.
65. Tanaka T, Maeda Y, Veluchamy A, Tanaka M, Abida H, Maréchal E, E, Bowler C, Muto M, Sunaga Y, Tanaka M, Yoshino T, Taniguchi T, Fukuda Y, Nemoto M, Matsumoto M, Wong PS, Aburatani S, Fujibuchi W: Oil accumulation by the oleaginous diatom Fistulifera solaris as revealed by the genome and transcriptome. Plant Cell. 2015;27:162–76.
66. Shiu S-H, Shih M-C, Li W-H. Transcription Factor Families Have Much Higher Expansion Rates in Plants than in Animals. Plant Physiol. 2005;139:18–26.
67. Kersting AR, Bornberg-Bauer E, Moore AD, Grath S. Dynamics and adaptive benefits of protein domain emergence and arrangements during plant genome evolution. Genome Biol Evol. 2012;4:316–29.
68. Khanna R, Kronmiller B, Maszle DR, Coupland G, Holm M, Mizuno T, Wu S-H: The Arabidopsis B-box zinc finger family. Plant Cell. 2009;21:3416–20.
69. Kumagai T, Ito S, Nakamichi N, Niwa Y, Murakami M, Yamashino T,Mizuno T: The common function of a novel subfamily of B-Box zinc finger proteins with reference to circadian-associated events in Arabidopsis thaliana. Biosci Biotechnol Biochem. 2008;72:1539–49.
70. Crocco CD, Holm M, Yanovsky MJ, Botto JF. Function of B-BOX under shade. Plant Signal Behav. 2011;6:101–4.
71. Huang J, Zhao X, Weng X, Wang L, Xie W. The rice B-box zinc finger gene family: genomic identification, characterization, expression profiling and diurnal analysis. PLoS ONE. 2012;7:e48242.
72. Bowler C, Botto J, Deng X-W. Photomorphogenesis, B-Box Transcription Factors, and the Legacy of Magnus Holm. Plant Cell. 2013;25:1192–5.
73. Gregis V, Sessa A, Colombo L, Kater MM. AGAMOUS-LIKE24 and SHORT VEGETATIVE PHASE determine floral meristem identity in Arabidopsis. Plant J. 2008;56:891–902.
74. Immink RGH, Posé D, Ferrario S, Ott F, Kaufmann K, Valentim FL, Folter S de, Wal F van der, Dijk ADJ van, Schmid M, Angenent GC: Characterization of SOC1's Central Role in Flowering by the Identification of Its Upstream and Downstream Regulators. Plant Physiol. 2012;160:433–49.
75. Maejima K, Iwai R, Himeno M, Komatsu K, Kitazawa Y, Fujita N, Ishikawa K, Fukuoka M, Minato N, Yamaji Y, Oshima K, Namba S: Recognition of floral homeotic MADS domain transcription factors by a phytoplasmal effector, phyllogen, induces phyllody. Plant J. 2014;78:541–54.
76. Kewley RJ, Whitelaw ML, Chapman-Smith A. The mammalian basic helix–loop–helix/PAS family of transcriptional regulators. Int J Biochem Cell Biol. 2004;36:189–204.
77. Lindebro MC, Poellinger L, Whitelaw ML. Protein-protein interaction via PAS domains: role of the PAS domain in positive and negative regulation of the bHLH/PAS dioxin receptor-Arnt transcription factor complex. EMBO J. 1995; 14:3528–39.
78. Erbel PJA, Card PB, Karakuzu O, Bruick RK, Gardner KH. Structural basis for PAS domain heterodimerization in the basic helix–loop–helix-PAS transcription factor hypoxia-inducible factor. Proc Natl Acad Sci U S A. 2003;100:15504–9.
79. Takahashi F, Yamagata D, Ishikawa M, Fukamatsu Y, Ogura Y, Kasahara M, Kiyosue T, Kikuyama M, Wada M, Kataoka H: AUREOCHROME, a photoreceptor required for photomorphogenesis in stramenopiles. Proc Natl Acad Sci U S A. 2007;104:19625–30.
80. Ishikawa M, Takahashi F, Nozaki H, Nagasato C, Motomura T, Kataoka H. Distribution and phylogeny of the blue light receptors aureochromes in eukaryotes. Planta. 2009;230:543–52.
81. Vieler A, Wu G, Tsai C-H, Bullard B, Cornish AJ, Harvey C, Reca I-B, Thornburg C, Achawanantakun R, Buehl CJ, Campbell MS, Cavalier D, Childs KL, Clark TJ, Deshpande R, Erickson E, Armenia Ferguson A, Handee W, Kong Q, Li X, Liu B, Lundback S, Peng C, Roston RL, Sanjaya, Simpson JP, TerBush A, Warakanont J, Zäuner S, Farre EM, et al. Genome, Functional Gene Annotation, and Nuclear Transformation of the Heterokont Oleaginous Alga Nannochloropsis oceanica CCMP1779. PLoS Genet. 2012;8:e1003064.
82. Schellenberger Costa B, Sachse M, Jungandreas A, Bartulos CR, Gruber A, Jakob T, et al. Aureochrome 1a is Involved in the Photoacclimation of the Diatom Phaeodactylum tricornutum. PLoS ONE. 2013;8:e74451.
83. Austin RW, Petzold TJ. Spectral Dependence of the Diffuse Attenuation Coefficient of Light in Ocean Waters. Opt Eng. 1986;25:253471–9.
84. Huysman MJJ, Fortunato AE, Matthijs M, Costa BS, Vanderhaeghen R, Van den Daele H, Sachse M, Inzé D, Bowler C, Kroth PG, Wilhelm C, Falciatore A, Vyverman W, De Veylder L. AUREOCHROME1a-mediated induction of the diatom-specific cyclin dsCYC2 controls the onset of cell division in diatoms (Phaeodactylum tricornutum). Plant Cell. 2013;25:215–28.
85. Hegemann P. Algal Sensory Photoreceptors. Annu Rev Plant Biol. 2008;59:167–89.
86. Briggs WR, Christie JM. Phototropins 1 and 2: versatile plant blue-light receptors. Trends Plant Sci. 2002;7:204–10.

87. Christie JM, Blackwood L, Petersen J, Sullivan S. Plant Flavoprotein Photoreceptors. Plant Cell Physiol. 2015;56:401–13.
88. Feller A, Machemer K, Braun EL, Grotewold E. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. Plant J Cell Mol Biol. 2011;66:94–116.
89. Taylor BL, Zhulin IB. PAS Domains: Internal Sensors of Oxygen, Redox Potential, and Light. Microbiol Mol Biol Rev. 1999;63:479–506.
90. Miller G, Mittler R. Could heat shock transcription factors function as hydrogen peroxide sensors in plants? Ann Bot. 2006;98:279–88.
91. Liu Y, Zhang C, Chen J, Guo L, Li X, Li W, Yu Z, Deng J, Zhang P, Zhang K, Zhang L: Arabidopsis heat shock factor HsfA1a directly senses heat stress, pH changes, and hydrogen peroxide via the engagement of redox state. Plant Physiol Biochem PPB Société Fr Physiol Végétale. 2013;64:92–8.
92. Partch CL, Gardner KH. Coactivator recruitment: a new role for PAS domains in transcriptional regulation by the bHLH-PAS family. J Cell Physiol. 2010;223:553–7.
93. Liu T, Golden JW, Giedroc DP. A zinc(II)/lead(II)/cadmium(II)-inducible operon from the Cyanobacterium anabaena is regulated by AztR, an alpha3N ArsR/SmtB metalloregulator. Biochemistry (Mosc). 2005;44:8673–83.
94. Lavoie BD, Shaw GS, Millner A, Chaconas G. Anatomy of a Flexer–DNA Complex inside a Higher-Order Transposition Intermediate. Cell. 1996;85:761–71.
95. Aki T, Adhya. Repressor induced site-specific binding of HU for transcriptional regulation. EMBO J. 1997;16:3666–74.
96. Santos JM, Freire P, Vicente M, Arraiano CM. The stationary-phase morphogene bolA from Escherichia coli is induced by stress during early stages of growth. Mol Microbiol. 1999;32:789–98.
97. Maris AE, Sawaya MR, Kaczor-Grzeskowiak M, Jarvis MR, Bearson SMD, Kopka ML, et al. Dimerization allows DNA target site recognition by the NarL response regulator. Nat Struct Mol Biol. 2002;9:771–8.
98. Chai Y, Winans SC. Site-directed mutagenesis of a LuxR-type quorum-sensing transcription factor: alteration of autoinducer specificity. Mol Microbiol. 2004;51:765–76.
99. Cangiano G, Mazzone A, Baccigalupi L, Isticato R, Eichenberger P, De Felice M, et al. Direct and indirect control of late sporulation genes by GerR of Bacillus subtilis. J Bacteriol. 2010;192:3406–13.
100. Takahashi Y, Yamaguchi O, Omata T. Roles of CmpR, a LysR family transcriptional regulator, in acclimation of the cyanobacterium Synechococcus sp. strain PCC 7942 to low-CO(2) and high-light conditions. Mol Microbiol. 2004;52:837–45.
101. Frías JE, Flores E, Herrero A. Activation of the Anabaena nir operon promoter requires both NtcA (CAP family) and NtcB (LysR family) transcription factors. Mol Microbiol. 2000;38:613–25.
102. Kawamukai M, Utsumi R, Takeda K, Higashi A, Matsuda H, Choi YL, et al. Nucleotide sequence and characterization of the sfs1 gene: sfs1 is involved in CRP*-dependent mal gene expression in Escherichia coli. J Bacteriol. 1991;173:2644–8.
103. Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, et al. Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc. Natl. Acad. Sci. U.S.A. 2002;99:12246–51.
104. Leliaert F, Smith DR, Moreau H, Herron MD, Verbruggen H, Delwiche CF, et al. Phylogeny and Molecular Evolution of the Green Algae. Critical Reviews in Plant Sciences. 2012;31:1–46.
105. Nosenko T, Bhattacharya D. Horizontal gene transfer in chromalveolates. BMC Evol Biol. 2007;7:173.
106. MacPherson S, Larochelle M, Turcotte B. A Fungal Family of Transcriptional Regulators: the Zinc Cluster Proteins. Microbiol Mol Biol Rev. 2006;70:583–604.
107. Todd RB, Andrianopoulos A. Evolution of a fungal regulatory gene family: the Zn(II)2Cys6 binuclear cluster DNA binding motif. Fungal Genet Biol FG B. 1997;21:388–405.
108. Pan T, Coleman JE. GAL4 transcription factor is not a "zinc finger" but forms a Zn(II)2Cys6 binuclear cluster. Proc Natl Acad Sci U S A. 1990;87:2077–81.
109. Martens JA, Laprade L, Winston F. Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. Nature. 2004;429:571–4.
110. Moye-Rowley WS. Transcriptional control of multidrug resistance in the yeast Saccharomyces. Prog Nucleic Acid Res Mol Biol. 2003;73:251–79.
111. Felenbok B, Flipphi M, Nikolaev I. Ethanol catabolism in Aspergillus nidulans: a model system for studying gene regulation. Prog Nucleic Acid Res Mol Biol. 2001;69:149–204.
112. Hynes MJ, Murray SL, Duncan A, Khew GS, Davis MA. Regulatory genes controlling fatty acid catabolism and peroxisomal functions in the filamentous fungus Aspergillus nidulans. Eukaryot Cell. 2006;5:794–805.
113. Garrido SM, Kitamoto N, Watanabe A, Shintani T, Gomi K. Functional analysis of FarA transcription factor in the regulation of the genes encoding lipolytic enzymes and hydrophobic surface binding protein for the degradation of biodegradable plastics in Aspergillus oryzae. J Biosci Bioeng. 2012;113:549–55.
114. McFadden GI. Origin and Evolution of Plastids and Photosynthesis in Eukaryotes. Cold Spring Harb Perspect Biol. 2014;6:a016105.
115. Richards TA, Soanes DM, Foster PG, Leonard G, Thornton CR, Talbot NJ. Phylogenomic Analysis Demonstrates a Pattern of Rare and Ancient Horizontal Gene Transfer between Plants and Fungi. Plant Cell Online. 2009;21:1897–911.
116. Chan CX, Reyes-Prieto A, Bhattacharya D. Red and green algal origin of diatom membrane transporters: insights into environmental adaptation and cell evolution. PLoS ONE. 2011;6:e29138.
117. Mackiewicz P, Bodył A, Moszczyński K. The case of horizontal gene transfer from bacteria to the peculiar dinoflagellate plastid genome. Mob Genet Elem. 2013;3:e25845.
118. Qiu H, Yoon HS, Bhattacharya D. Algal endosymbionts as vectors of horizontal gene transfer in photosynthetic eukaryotes. Plant Physiol. 2013;4:366.
119. Qiu H, Price DC, Weber APM, Reeb V, Chan Yang E, Lee JM, et al. Adaptation through horizontal gene transfer in the cryptoendolithic red alga Galdieria phlegrea. Curr Biol. 2013;23:R865–6.
120. Beck A, Divakar PK, Zhang N, Molina MC, Struwe L. Evidence of ancient horizontal gene transfer between fungi and the terrestrial alga Trebouxia. Org Divers Evol. 2014;15:235–48.
121. Liu H, Probert I, Uitz J, Claustre H, Aris-Brosou S, Frada M, et al. Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. Proc Natl Acad Sci U S A. 2009;106:12803–8.
122. Krogh A, Brown M, Mian IS, Sjölander K, Haussler D. Hidden Markov models in computational biology. Applications to protein modeling. J Mol Biol. 1994;235:1501–31.
123. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinforma Oxf Engl. 2014;30:1236–40.
124. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011;39 suppl 2:W29–37.

IV. Résultats complémentaires : Apport du génome de l'haptophyte *Chrysochromulina tobin* à cette étude comparative.

Depuis l'écriture de cette publication, un nouveau génome d'haptophyte a récemment été publié : celui de *Chrysochromulina tobin* (Hovde *et al.*, 2015). Cet haptophyte, non coccolithophore, appartient à l'ordre des Prymnesiales et à la famille des Chrysochromulinaceae. Le pipeline d'identification et de classification des FTs a donc été appliqué au protéome prédit de *C. tobin*, et les résultats obtenus ont été ajoutés aux précédents (Tableau 2).

Comme dans le cas des autres microalgues étudiées, des familles précédemment décrites comme spécifiques de la lignée verte ont été identifiées (AP2/ERF, CSD, TUB et C2C2-YABBY). De plus, des séquences appartenant aux Fungal TRF ont également été identifiées chez *C. tobin*, ainsi que des séquences appartenant à des familles de FTs retrouvées chez les cyanobactéries.

A l'exception d'une séquence appartenant à la famille des C2C2-YABBY, *C. tobin* ne comporte pas de particularité par rapport aux autres espèces. L'ajout de ces données au dendrogramme (Figure 3 de Thiriet-Rupert et al 2016 fondée sur la présence/absence des différentes familles de FTs chez chaque espèce) montre un embranchement de *C. tobin* entre *Pavlova* sp. d'une part, et *T. lutea* et *E. huxleyi* d'autre part (Figure 39). Cette structure étant conforme à la phylogénie des haptophytes. De plus, l'ajout des données obtenues pour *C. tobin* à la construction de la « heatmap » (Figure 4 de Thiriet-Rupert et al 2016) montre une discrimination plus efficace des familles de FTs en fonction de leurs proportions dans le génome des différentes microalgues étudiées (Figure 40). Le cluster N°1 regroupe les familles décrites comme spécifiques de la lignée verte. Le N°2 regroupe les familles communes aux huit microalgues ou majoritairement réparties et dont les proportions varient peu. Le cluster N°3 correspond aux familles communes aux huit microalgues et dont les proportions varient d'une espèce à l'autre. Le N°4 regroupe les trois familles absentes chez les straménopiles. Et le cluster N°5 regroupe les trois cas d'expansion ainsi que la famille des MYB qui, hormis ces trois expansions, est la plus représentée chez chacune des espèces.

Tableau 2 : FTs identifiés dans le génome des huit microalgues de l'étude. Pour chaque espèce, la colonne de gauche correspond au nombre de séquences appartenant à chaque famille. Celle de droite correspond à leur proportion par rapport au nombre total de FTs identifiés dans le génome de l'organisme en question.

FT famille		*Tisochrtsis lutea*		*Pavlova sp*		*Emiliania huxleyi*		*Chrysochromulina tobin*		*Phaeodactylum tricornutum*		*Nannochloropsis gaditana*		*Porphyridium purpureum*		*Chlamydomonas renhardtii*	
B3	ABI3/VP1	1	0.65	0	0	0	0	0	0	0	0	0	0	0	0	1	0.47
AP2/ERF	AP2	1	0.65	1	0.78	58	12.13	11	6.83	0	0	2	2.15	0	0	6	2.83
	ERF	1	0.65	6	4.69	99	20.71	7	4.35	2	1.02	2	2.15	0	0	9	4.25
bHLH		0	0	0	0	0	0	1	0.62	8	4.08	3	3.23	3	1.51	8	3.77
bZIP		3	1.94	3	2.34	6	1.26	2	1.24	25	12.76	11	11.83	21	10.55	20	9.43
C2C2	CO-like	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.47
	Dof	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.47
	GATA	5	3.23	1	0.78	4	0.84	1	0.62	0	0	0	0	2	1.01	12	0.66
	LSD	1	0.65	1	0.78	0	0	0	0	0	0	0	0	0	0	1	0.47
	YABBY	0	0	0	0	0	0	1	0.62	0	0	0	0	0	0	1	0.47
C2H2		8	5.16	8	6.25	37	7.74	18	11.18	4	2.04	5	5.38	60	30.15	5	2.36
C3H		13	8.39	7	5.47	49	9.83	13	8.07	11	5.61	5	5.38	8	4.02	22	10.38
CCAAT		3	1.94	0	0	2	0.42	2	1.24	3	1.53	3	3.23	3	1.51	1	0.47
CPP		1	0.65	0	0	4	0.84	3	1.86	5	2.55	1	1.08	2	1.01	3	1.42
CSD		3	1.94	4	3.13	25	5.23	5	3.11	5	2.55	1	1.08	3	1.51	2	0.94
DBB		0	0	0	0	0	0	0	0	0	0	0	0	1	0.5	0	0
E2F/DP		2	1.29	3	2.34	3	0.63	4	2.48	5	2.55	1	1.08	3	1.51	3	1.42
Fungal TRF		14	9.03	8	6.25	27	5.65	3	1.86	1	0.51	10	10.75	0	0	0	0
GARP	G2-like	4	2.58	4	3.13	5	1.05	7	4.35	2	1.02	0	0	2	1.01	4	1.89
	ARR-B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.47
Homeobox	HB-other	16	10.32	14	10.94	28	5.86	13	8.07	0	0	0	0	2	1.01	1	0.47
	TALE	1	0.65	1	0.78	0	0	1	0.62	4	2.04	0	0	9	4.52	3	1.42
HSF		9	5.81	8	6.25	8	1.67	1	0.62	67	34.18	4	4.3	1	0.5	2	0.94
LIM		2	1.29	3	2.34	11	2.3	2	1.24	0	0	0	0	2	1.01	1	0.47
MADS-box	M-type	3	1.94	1	0.78	1	0.21	1	0.62	0	0	0	0	2	1.01	2	0.94
mTERF		5	3.23	0	0	6	1.26	3	1.86	5	2.55	2	2.15	5	2.51	4	1.89
MYB	MYB (3R)	1	0.65	0	0	3	0.63	0	0	2	1.02	5	5.38	1	0.5	1	0.47
	MYB (2R)	25	16.13	20	15.63	39	8.16	24	14.91	11	5.61	8	8.6	23	11.56	10	4.72
	MYB-rel	21	13.55	15	11.9	51	10.69	12	7.45	7	3.57	7	7.53	7	3.52	18	8.65
	SHAQKYF	1	0.65	2	1.56	1	0.21	3	1.86	7	3.57	8	8.6	16	8.04	4	1.89
NF-X1		0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.47
NF-Y	NF-YA	0	0	1	0.78	1	0.21	3	1.86	1	0.51	2	1.08	1	0.5	0	0
	NF-YB	1	0.65	1	0.78	4	0.84	7	4.35	2	1.02	6	2.15	3	1.51	3	1.42
	NF-YC	3	1.94	4	3.13	1	0.21	4	2.48	8	4.08	1	6.45	6	3.02	2	0.94
Nin-like		0	0	1	0.78	0	0	0	0	0	0	1	1.08	4	2.01	15	7.08
S1Fa-like		0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.47
SBP		0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	10.85
Sigma-70		4	2.58	4	3.13	2	0.42	5	3.11	8	4.08	4	4.3	8	4.02	1	0.47
TUB		3	1.94	7	5.47	5	1.05	4	2.48	3	1.53	1	1.08	0	0	6	2.83
VARL		0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	5.66
Whirly		0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.47
WRKY		0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.47

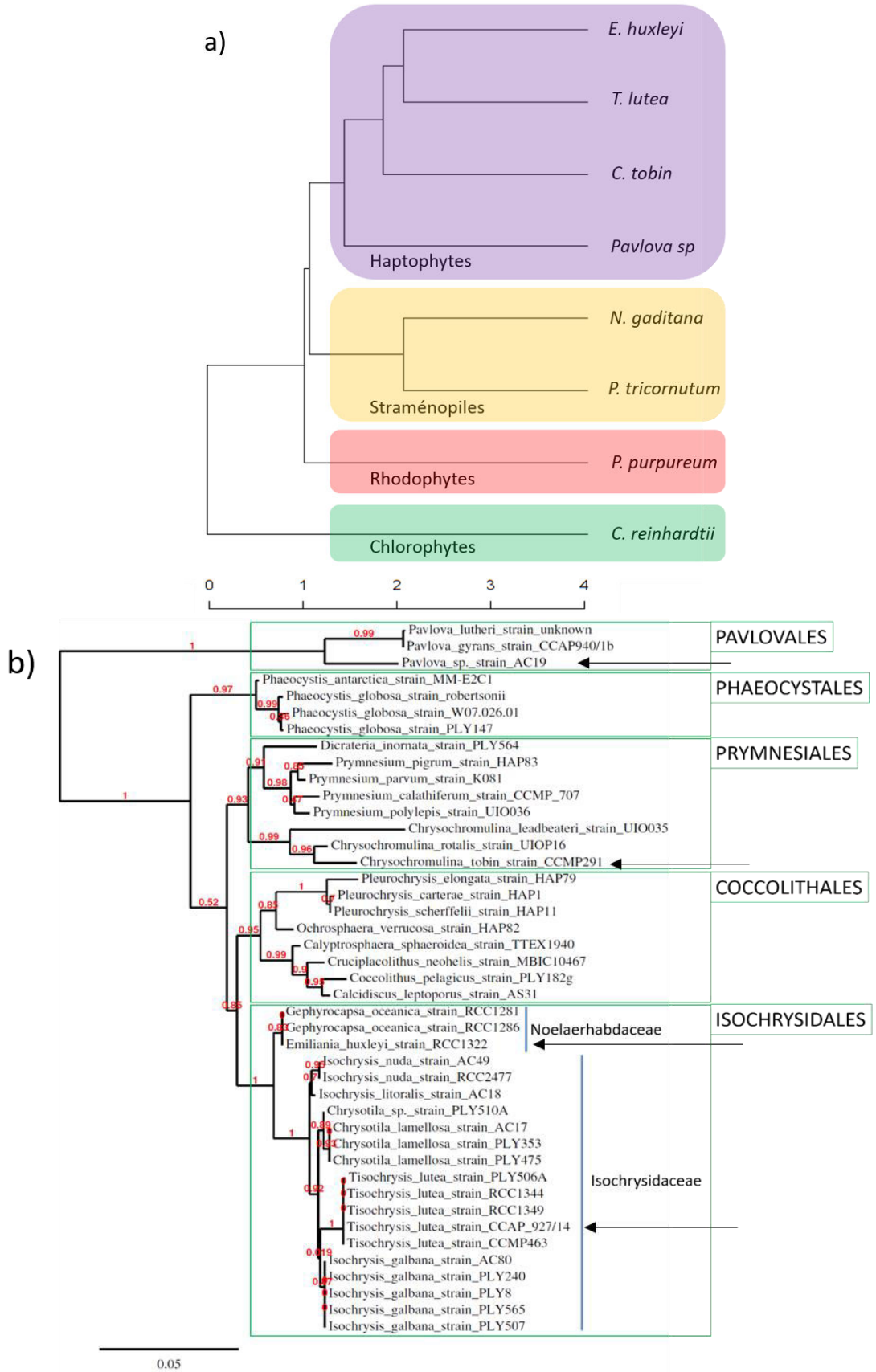

Figure 39 : ajout des données de *C. tobin* à la construction du dendrogramme de Thiriet-Rupert *et al.*, 2016. a) Dendrogramme construit à partir des données de présence/absence des familles de FTs chez chaque espèce de microalgues. b) Phylogénie des haptophytes construite à partir de leur séquence 16S. Les quatre haptophytes présents sur le dendrogramme sont pointés par une flèche.

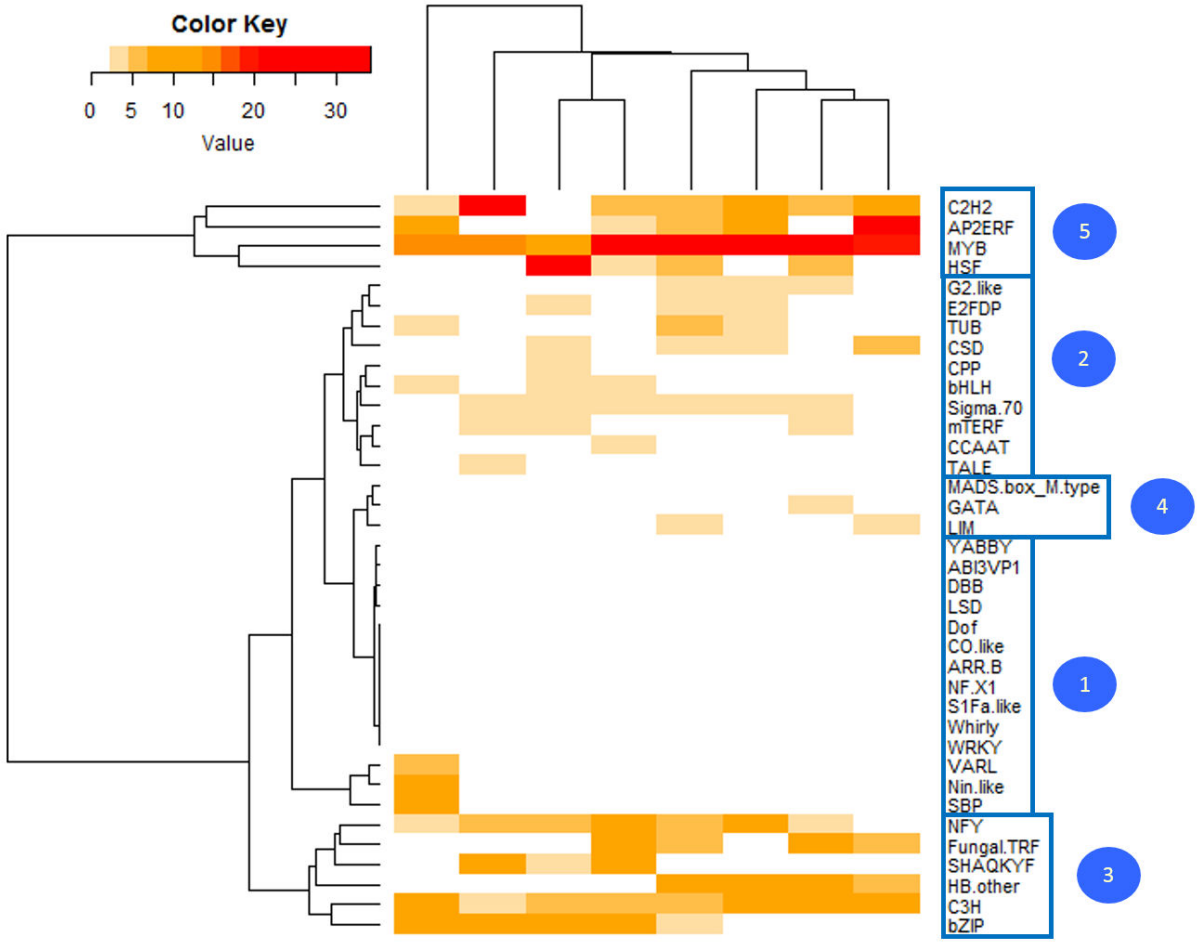

Figure 40 : ajout des données de *C. tobin* à la construction de la heatmap de Thiriet-Rupert *et al.*, 2016. a) Représentation « heatmap » réalisée à partir des proportions de chaque famille de FTs dans le génome des sept algues de la publication (Thiriet-Rupert et al 2016). b) Représentation « heatmap » après ajout du génome de la microalgue *C. tobin* à l'analyse.

L'ensemble de ces résultats conforte donc les conclusions de la publication et montre que plus le nombre d'espèces utilisées est élevé, plus les résultats obtenus sont précis. En revanche, l'hypothèse de l'absence des FTs de la famille des bHLH chez les haptophytes n'est plus valable. En effet, la présence d'un FT bHLH dans le génome de *C. tobin* contredit la perte de cette famille chez les haptophytes. Afin d'évaluer si cette espèce constitue un cas isolé, une recherche de séquences appartenant à la famille des bHLH chez des espèces d'haptophytes a été menée. Pour cela, les données transcriptomiques générées par le projet Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) (http://data.imicrobe.us/project/view/104) (Keeling *et al.*, 2014) ont été utilisées. Les séquences de bHLH ont été recherchées dans l'annotation fonctionnelle du transcriptome de 14 espèces d'haptophytes représentant cinq des sept ordres de cette lignée (Tableau 3). De manière intéressante, aucune séquence de bHLH n'a été identifiée chez les six Isochrysidales, les deux Coccolithales ni chez *Pavlova* sp. alors que les deux Phaeocystales et quatre Prymnesiales comportent au moins un bHLH. En revanche, aucune séquence de bHLH n'a été identifiée dans le transcriptome de *Chrysochromulina ericina*, qui appartient pourtant également aux Prymnesiales. La répartition des bHLH au sein des haptophytes semble donc restreinte aux seules Phaeocystales et Prymnesiales (Figure 41). Une répartition aussi atypique peut s'expliquer par deux hypothèses. (1) Une perte de cette famille de FTs chez l'ancêtre commun aux Coccolithales et aux Isochrysidales, ainsi qu'une perte au cours de l'histoire évolutive des Pavlovales. (2) Ou bien, à l'inverse, le gain de cette famille chez l'ancêtre commun aux Phaeocystales et aux Prymnesiales. Des arguments supportent chacune de ces deux hypothèses. Tout d'abord, expliquer cette répartition par l'intervention d'un seul événement de gain chez l'ancêtre commun aux Phaeocystales et aux Prymnesiales est plus probable que par deux évènements de perte de cette famille de FTs. Cependant, le gain de la famille de bHLH chez l'ancêtre commun aux Phaeocystales et aux Prymnesiales impliquerait l'absence de cette famille chez l'ancêtre commun à l'ensemble des haptophytes. Or, les bHLH sont parmi les FTs les plus répandus puisqu'ils sont retrouvés chez l'ensemble des eucaryotes (Massari & Murre, 2000 ; Sailsbery & Dean, 2012). Il est donc difficile de définir laquelle des deux propositions suivantes est la plus probable : que l'ancêtre des haptophytes ait été dépourvu de bHLH, ou que deux événements de perte de cette famille soient intervenus au cours de leur histoire évolutive. Quoi qu'il en soit, ce ne sont pour le moment que des hypothèses qui demandent, en premier lieu, une

confirmation de la répartition des bHLH parmi les différents ordres des haptophytes d'un point de vue génomique.

Tableau 3 : espèces d'haptophytes utilisées pour la recherche spécifique de FTs de la famille de bHLH.

Ordre	Espèce	Souche
Pavlovale	Pavlova sp.	CCMP459
Phaeocystale	Phaeocystis antarctica	Caron Lab Isolate
		CCMP1374
	Phaeocystis cordata	RCC1383
Prymnesiale	Prynesium parvum	Texoma1
	Chrysochromulina brevifilum	UTEX LB 985
	Chrysochromulina ericina	CCMP281
	Chrysochromulina rotalis	UIO044
	Chrysochromulina polylepis	CCMP1757
Coccolithale	Pleurochrysis carterae	CCMP645
	Coccolithus pelagicus ssp braarudi	PLY182g
Isochrysidale	Emiliania huxleyi	CCMP370
		PLYM2019
	Gephyrocapsa oceanica	RCC1303
	Isochrysis galbana	CCMP1323
	Isochrysis sp.	CCMP1244
		CCMP1324

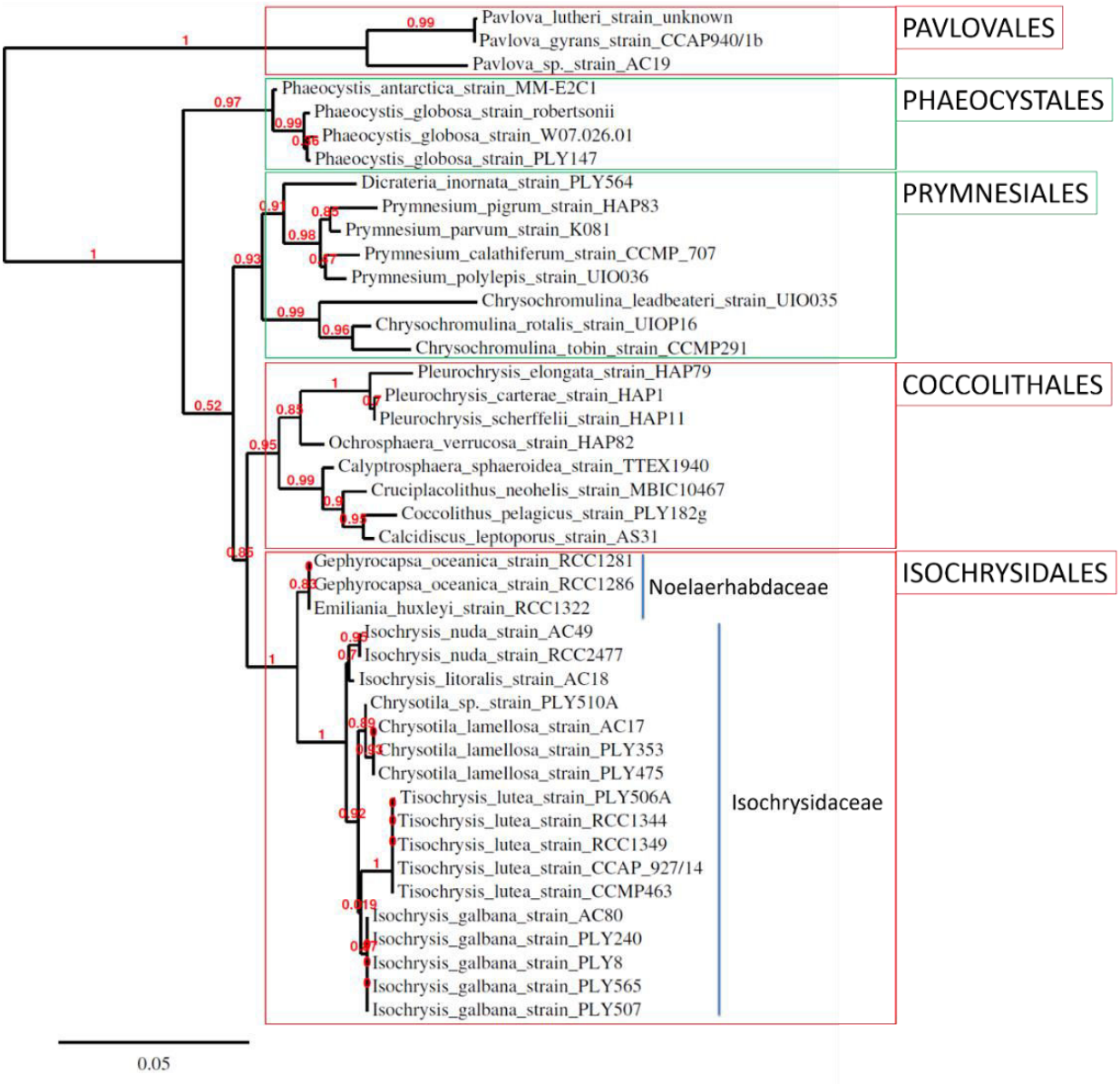

Figure 41 : Répartition des FTs de la famille des bHLH chez les cinq ordres des haptophytes représentés dans leur recherche spécifique. L'arbre a été construit à partir des séquences 16S du plus large échantillon trouvé dans les bases de données. Les ordres encadrés en rouges correspondent à ceux au sein desquels aucun bHLH n'a été identifié. En vert, sont encadrés les ordres dont des membres présentent ces FTs.

V. Bilan et perspectives

L'ensemble de ces résultats montre bien que la taille de l'échantillonnage est primordiale dans ce type d'études et que plus le nombre de génomes séquencés sera important et taxonomiquement diversifié, plus les résultats obtenus seront précis et fiables. La comparaison des FTs identifiés chez des organismes représentant quatre lignées de microalgues constitue, en effet, le principal point fort de cette étude. Cette couverture, plus élargie comparativement aux études précédentes permet d'affiner et renforcer les spécificités de lignées, ainsi que d'en identifier de nouvelles. Toutefois, utiliser davantage d'organismes renforcerait encore plus ces dernières. Le fait que *P. purpureum* soit la seule microalgue rouge mésophile dont le génome est séquencé rend difficile l'intégration de cette lignée, pourtant à l'origine du chloroplaste de nombreuses lignées (cf Introduction), aux études comparatives visant à étudier l'histoire évolutive des microalgues. Un autre exemple intéressant est celui de la répartition des bHLH chez les haptophytes. Bien que cette surprenante caractéristique soit étayée par l'analyse de données concernant 14 espèces représentant cinq des sept ordres des haptophytes (des données concernant les Syracosphaerales et les Zygodiscales manquent dans notre analyse) un plus grand échantillonnage permettrait de préciser cette répartition atypique. De plus, hormis pour *T. lutea*, *E. huxleyi* et *C. tobin* pour lesquelles le génome est séquencé, cette répartition est fondée sur l'analyse de données transcriptomiques. L'absence de bHLH de ce jeu de données ne prouve donc que l'absence d'expression de tels FTs dans les conditions appliquées à la culture utilisée pour le séquençage. Des données génomiques seraient nécessaires à la confirmation de l'absence de bHLH chez les Isochrysidales, les Coccolithales et les Pavlovales.

Enfin, concernant les familles définies comme spécifiques de la lignée verte, certaines sont retrouvées chez la plupart des microalgues de cette étude (familles CSD, AP2/ERF ou TUB). Cette répartition s'explique par l'exhaustivité des lignées incluses dans cette étude par rapport à celles qui avaient décrit ces familles comme spécifiques de la lignée verte (Sharma *et al.*, 2013 ; Lang *et al.*, 2010). D'autres de ces familles ne sont retrouvées que très ponctuellement et en petit nombre en dehors de la lignée verte (une séquence de la famille ABI3/VP1 chez *T. lutea*, une séquence de C2C2-LSD chez *T. lutea* et *Pavlova* sp., et une séquence de C2C2-YABBY chez *C. tobin*). Ce cas est, en revanche, plus délicat à élucider. Ces familles ne sont retrouvées que chez une ou deux espèces et non communes aux quatre haptophytes. De plus, chacune de ces familles

n'est représentée que par une seule séquence. Autant d'indices qui suggèrent un transfert horizontal récent. Un tel événement a également été suggéré par Hunsperger *et al.*, (2015) à partir de données génomiques concernant la présence des gènes *por* (protochlorophyllide oxidoreductases), impliqués dans la synthèse de chlorophylle a, chez les haptophytes. Lors de la rédaction de la publication, la présence de ces séquences de FTs spécifiques de la lignée verte a été associée à l'hypothèse d'une endosymbiose cryptique. Celle-ci propose que l'endosymbiose d'une microalgue verte aurait précédé celle d'une microalgue rouge chez les haptophytes et straménopiles (Moustafa *et al.*, 2009 ; Dorrell & Smith, 2011). Toutefois cette hypothèse reçoit peu de crédit dans les études récentes concernant l'histoire évolutive des algues. Une autre hypothèse, plus soutenue, est celle d'un transfert horizontal lié à la prédation, certains représentants des lignées CASH (Cryptophytes, Alvéolées, Straménopiles, Haptophytes) étant communément prédatrices (Doolittle, 1998). Récemment, Keeling (2013) a repris cette hypothèse selon laquelle, chez les lignées CASH, préalablement à l'événement d'endosymbiose, des transferts de gènes auraient eu lieu depuis le génome de microalgues vertes phagocytées par prédation (Figure 42). Ces mécanismes expliquant la présence de gènes originaires de la lignée verte chez les lignées CASH dont le chloroplaste est dérivé d'une microalgue rouge.

Une autre hypothèse concernant l'origine de ces transferts de gènes horizontaux fait intervenir un vecteur viral. En effet, les virus sont connus pour être impliqués dans le transfert de gènes entre organismes (Gilbert *et al.*, 2014 ; Gao *et al.*, 2014 ; Chen *et al.*, 2016) et, parmi les nombreux virus océaniques, une grande variété cible les algues (Short, 2012). La présence de gènes du photosystème I, impliqués dans la photosynthèse, ont été identifiés dans le génome de virus marins (Sharon *et al.*, 2009), et le transfert de sept gènes de la voie de biosynthèse des sphingolipides de l'haptophyte *E. huxleyi* dans le génome de son virus à ADN a été rapporté (Monier *et al.*, 2009). A l'inverse, des gènes viraux ont été identifiés dans le génome de plantes (Bertsch *et al.*, 2009) ainsi que chez d'autres eucaryotes, parmi lesquels des algues (Filée, 2014). Des transferts de gènes sont donc possibles depuis le génome des algues vers celui de virus et inversement, suggérant un rôle très probable de ces vecteurs dans les transferts de gènes chez les algues.

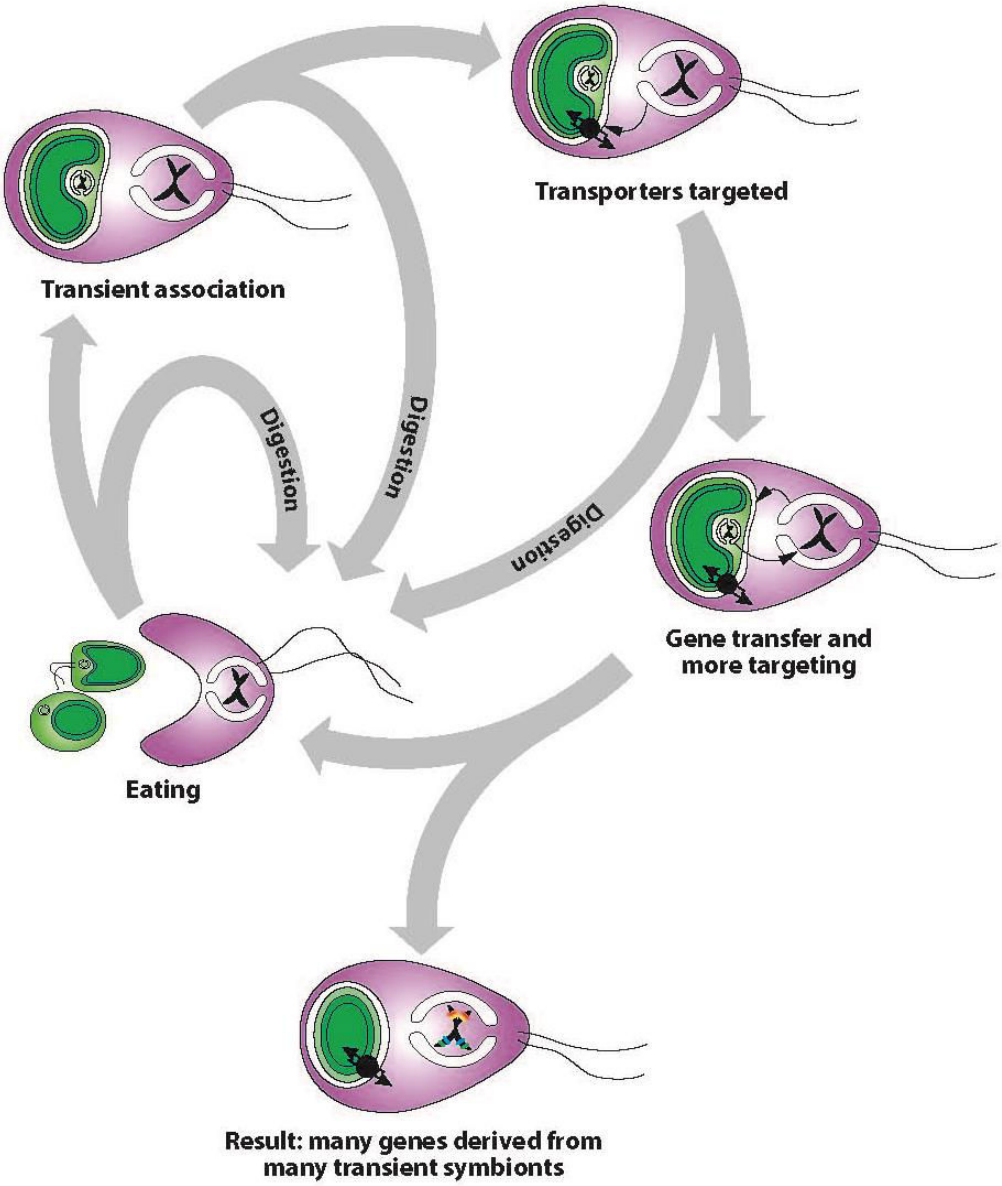

Figure 42 : Théorie proposée par Keeling (2013) selon laquelle les gènes provenant de la lignée verte et identifiés dans le génome des algues des lignées CASH auraient été acquis par prédation. L'organisme hétérotrophe (en violet) consomme diverses algues. S'en suit une rétention transitoire de la proie précédent sa digestion et permettant des transferts de gènes. Ce mécanisme pouvant se répéter plusieurs fois avant que l'hétérotrophe ne retienne définitivement l'une de ses proies.

De plus, des gènes dérivés de virus et de bactéries parasites sont contenus dans des plasmides d'algues, et ont également été retrouvés dans le génome nucléaire et chloroplastique de ces organismes hôtes (Lee *et al.*, 2016). Ces vecteurs sont bien connus pour diriger les transferts de gènes chez les bactéries (Ochman *et al.*, 2000) et des mécanismes, parfois impliqués dans la conjugaison bactérienne, permettent également aux bactéries parasites de transférer du matériel biologique (protéique ou nucléique) dans la cellule hôte (Lacroix & Citovsky, 2016). La présence de nombreux gènes de Chlamidiae, des bactéries parasites intracellulaires obligatoires, dans le génome d'algues (Huang & Gogarten, 2007 ; Becker *et al.*, 2008) suggère donc l'implication de ces mécanismes dans les transferts de gènes horizontaux chez les algues.

Sans être exhaustif, ces hypothèses sont parmi les principales élaborées concernant les mécanismes sous-jacents aux transferts de gènes horizontaux. Toutefois, aucune preuve décisive n'a pu être apportée en faveur de l'une d'elles.

Chapitre 2 : Identification de FTs impliqués dans l'orchestration des remaniements métaboliques constituant la réponse spécifique de la souche mutante de *T. lutea* à un stress azoté

La mise au point d'un pipeline bio-informatique a donc permis l'identification des FTs présents dans le génome de *T. lutea* (voir Chapitre 1). Dans ce chapitre, l'expression de l'ensemble de ces FTs chez les deux souches de *T. lutea* (la souche sauvage WTc1 et la souche mutante 2Xc1) et dans différentes conditions physiologiques a été étudiée. L'objectif consiste à identifier des FTs potentiellement impliqués dans l'établissement du phénotype de la souche 2Xc1 de *T. lutea*. Dans ce but, la construction et l'analyse de réseaux de co-expression et de régulation des gènes a été réalisée. Afin de pallier à l'annotation fonctionnelle parcellaire inhérente aux organismes non-modèles, une priorisation de gènes a été mise au point afin de lier le profil d'expression des gènes à la dynamique des paramètres physiologiques de l'algue. La fonction d'un FT étant identifiée à travers celle de ses gènes cibles, assigner une fonction putative aux gènes de *T. lutea* est une première étape indispensable à l'identification de FTs candidats. Une analyse de RT-PCR quantitative (q-RT-PCR) a ensuite permis de mieux caractériser le rôle de deux de ces régulateurs candidats dans la réponse spécifique de la souche 2Xc1 à un stress azoté. Cette étude fait l'objet d'une publication en cours de rédaction.

I. Introduction

La production de lipides par les microalgues étant provoquée par un stress azoté, les deux souches de *T. lutea* (WTc1 et 2Xc1) ont été cultivées durant 85 jours en photobioréacteur en chémostat limité par l'azote. Le chémostat est un type de culture en mode continu. Les cellules sont cultivées dans un réacteur alimenté en continu par du milieu nutritif dont un élément est présent en quantité limitante. Le volume de culture au sein du réacteur est maintenu constant par un flux sortant équivalent au flux entrant. Le contrôle de ce taux de dilution permet de fixer la concentration cellulaire de la culture. L'ensemble des paramètres de culture étant contrôlé, l'équilibre physiologique atteint par les cellules cultivées peut être maintenu indéfiniment. Une culture en chémostat offre donc l'avantage de maintenir de façon stable un état physiologique précis et contrôlé. Dans le cas présent, cet équilibre est caractérisé par une limitation azotée. Suite à la stabilisation de la biomasse algale au sein du photobioréacteur (état d'équilibre), un ajout d'azote ponctuel a été réalisé. A partir de modèles prédictifs (Mairet *et al.*, 2011), une hypothèse concernant la dynamique de biomasse algale et du contenu des cellules en lipides neutres a été

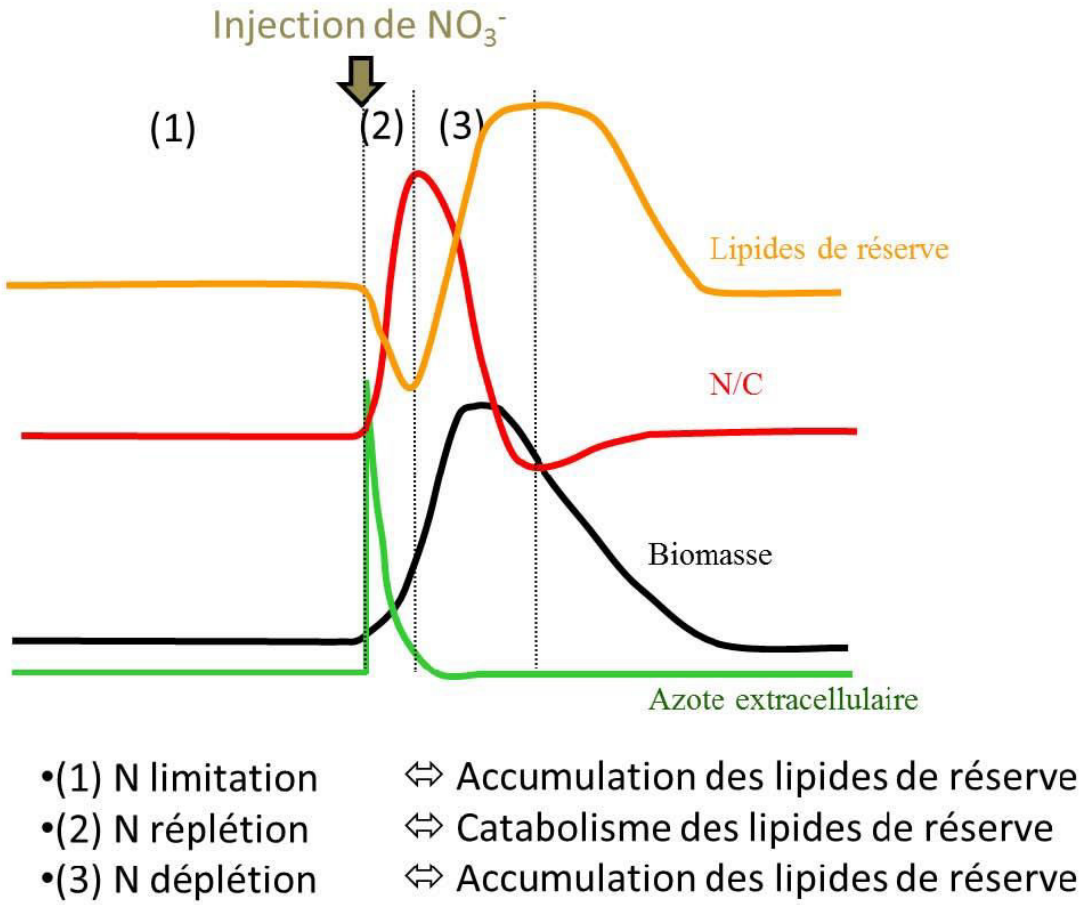

Figure 43 : représentation schématique de la dynamique des différents paramètres physiologiques de l'algue (la quantité de lipides de réserve intracellulaires, le rapport N/C et la quantité de biomasse), suite à une injection de nitrate lors d'une culture de *T. lutea* dans un chémostat limité par l'azote. Ces dynamiques correspondent aux prédictions de modèles écophysiologiques. (Garnier, 2016)

établie. Selon cette hypothèse, la culture étant limitée par l'azote, l'ajout de cet élément (réplétion azotée) provoque une augmentation de la biomasse (Figure 43). Parallèlement à cette augmentation de biomasse, une diminution des réserves lipidiques des cellules a lieu (phase de dégradation lipidique). La dégradation des lipides de réserve permet de produire de l'énergie sous forme d'ATP, du pouvoir réducteur sous forme de NADPH et NADH, et des précurseurs métaboliques pour la synthèse de biomasse fonctionnelle (protéines, acides nucléiques, lipides membranaires). Suite à cette phase de croissance, et une fois l'azote injecté entièrement consommé par les algues (déplétion azotée), une phase d'accumulation lipidique a lieu (Figure 43). Le manque d'azote disponible dans le milieu ne permet pas à l'algue de satisfaire ses besoins en azote. En conséquence, les algues stockent le carbone sous forme de lipides de réserve afin d'éviter la consommation d'énergie que provoquerait la synthèse de biomasse fonctionnelle. Suite au pic de biomasse, celle-ci diminue jusqu'à atteindre son niveau initial (phase d'équilibre en limitation azotée) du fait du taux de dilution de la culture continue. Au cours de l'expérience, les cultures ont été soumises à trois ajouts ponctuels d'azote successifs (Figure 44). Les données physiologiques récoltées durant l'expérience ont confirmé cette dynamique (Figure 44) ainsi que la suraccumulation lipidique de la souche mutante 2Xc1 (Figure 45). De plus, de manière inattendue, une augmentation de la capacité de stockage du carbone de cette même souche après chaque ajout d'azote a été mise en évidence. Cette augmentation du carbone intracellulaire après chaque ajout d'azote correspond à une augmentation des réserves de la cellule en carbohydrates (Figure 45). De nombreux prélèvements ont été réalisés tout au long des 85 jours de culture, dont six pour chaque souche (douze au total) ont fait l'objet d'analyses transcriptomiques par RNA-seq (Figure 46). Pour chacun de ces prélèvements des analyses physiologiques ont également été réalisées (dosage des lipides neutres et carbohydrates intracellulaires notamment).

Le phénotype de la souche mutante de *T. lutea* est caractérisé par la dynamique des quantités de lipides neutres et de carbohydrates intracellulaires au fil des 85 jours de culture. Les deux souches présentent des différences génétiques se traduisant par des variations d'expression de gènes, à l'origine du phénotype mutant. Par conséquent, identifier les gènes dont le profil d'expression est corrélé à la dynamique des caractéristiques physiologiques de la souche mutante permettrait de mieux comprendre l'établissement de ce phénotype.

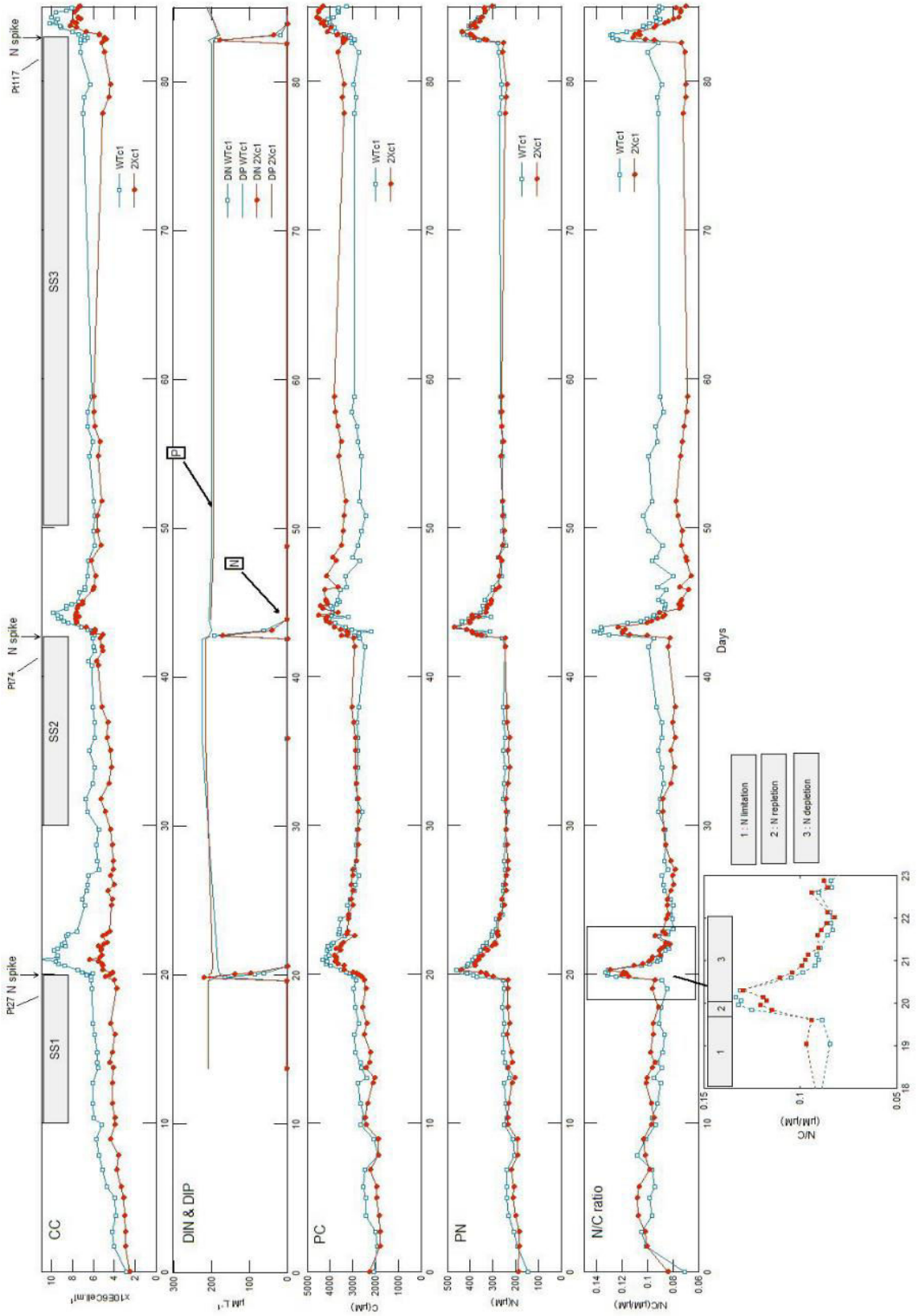

Figure 44 : suivi des paramètres physiologiques chez chaque souche au cours des 85 jours de culture (Garnier *et al.*, acceptée) (a). CC : concentration cellulaire, DIN & DIP : azote et phosphore inorganique dissous, PC : carbone particulaire, PN : azote particulaire. SS1, SS2 et SS3 correspondent aux trois états d'équilibre. En b), un zoom sur l'évolution du rapport N/C suite à la première injection d'azote.

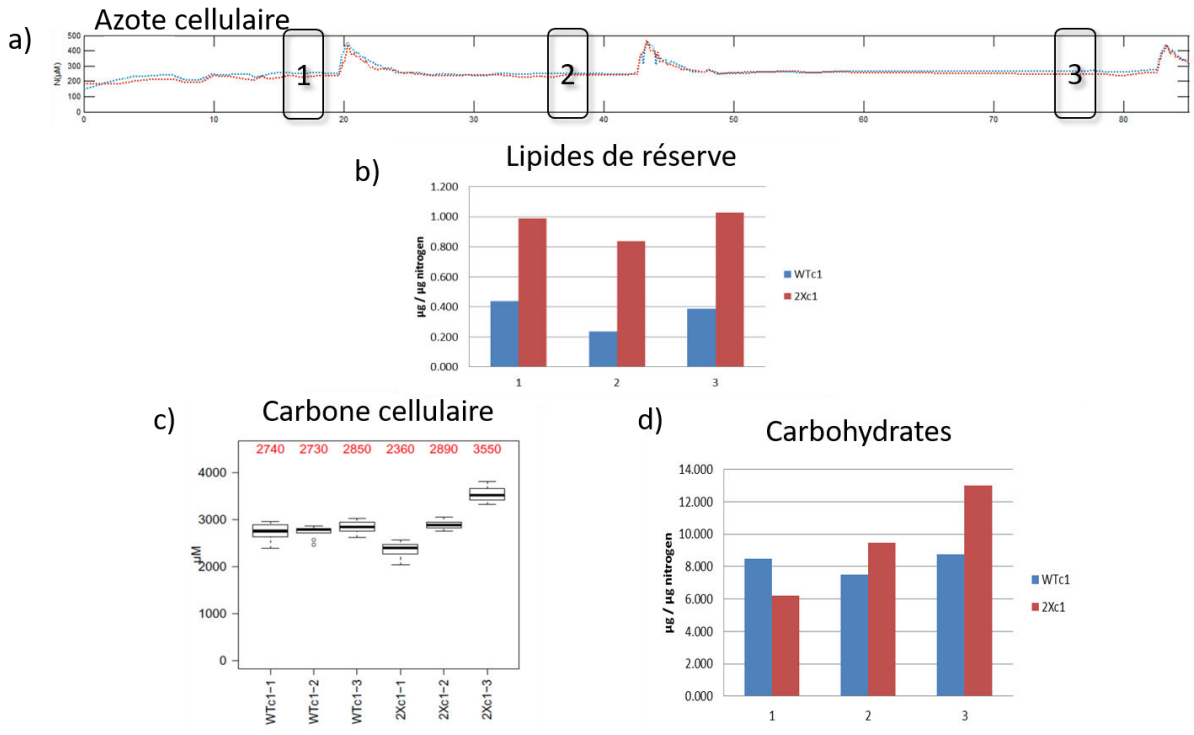

Figure 45 : caractérisation physiologique du phénotype mutant. Dosage des lipides de stockage et des carbohydrates dans les trois états d'équilibre (a) au cours des 85 jours de culture. La suraccumulation en lipides de stockage de la souche mutante a été confirmée (b). Une augmentation de la capacité de stockage du carbone chez la souche 2Xc1 a été observée après chaque injection d'azote (c). Cette augmentation du carbone stocké correspond à une augmentation de la quantité de carbohydrates cellulaires (d). (Garnier, 2016)

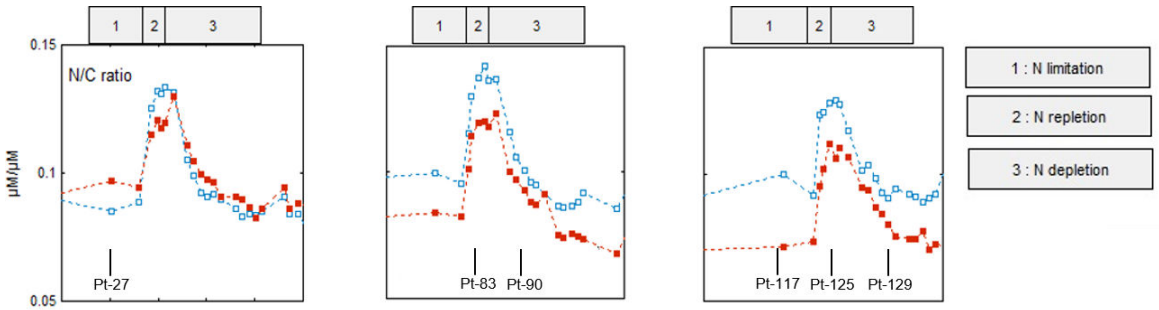

Figure 46 : échantillons utilisés pour l'analyse RNA-seq au long de la cinétique des trois injections d'azote. Deux échantillons (Pt-27 et Pt-117) correspondent à une limitation azotée, deux (Pt-83 et Pt-125) correspondent à une réplétion azotée, et deux (Pt-90 et et Pt-129) correspondent à une déplétion azotée.

De plus, les FTs jouent un rôle clé dans la régulation de l'expression de gènes. Ainsi, identifier ceux qui seraient impliqués dans la régulation de l'expression des gènes de façon différentielle exprimés chez la souche mutante permettrait d'identifier les réseaux de régulation sous-jacents à l'établissement de son phénotype. Métaphoriquement, Les FTs sont les chefs d'orchestre de l'établissement du phénotype dont la partition est jouée par un orchestre constitué par leurs gènes cibles. Ce chapitre se propose donc d'identifier les FTs impliqués dans l'établissement du phénotype mutant via l'annotation fonctionnelle de leurs gènes cibles putatifs. Dans cette optique, une stratégie couplant la construction et l'analyse de réseaux de co-expression et de régulation des gènes a été élaborée.

II. Application d'une stratégie visant à identifier les régulateurs potentiels de l'établissement d'un phénotype mutant chez un organisme non-modèle : *Tisochrysis lutea*.

1. Corrélation entre profil d'expression des gènes et dynamique des caractères phénotypiques pour compléter l'annotation fonctionnelle

a) Stratégie

La stratégie élaborée débute par la construction d'un réseau de co-expression des gènes grâce au package R WGCNA (Langfelder & Horvath, 2008) à partir des données transcriptomiques. Pour cela, les 12 échantillons RNA-seq (six pour chaque souche) ont été utilisés (Figure 46). La méthode implémentée dans ce package permet une identification robuste de modules de gènes co-exprimés. Ceux-ci regroupent des gènes dont la dynamique d'expression est identique au fil des échantillons étudiés. De plus, le grand avantage de WGCNA est qu'il réalise également la corrélation du profil d'expression des gènes avec la dynamique de caractères physiologiques quantifiables. Dans le cas présent, les gènes dont le profil d'expression est corrélé à la dynamique des quantités de lipides de stockage ou de carbohydrates cellulaires ont été identifiés. L'annotation fonctionnelle de ces gènes permet d'identifier les fonctions potentiellement liées à ces caractéristiques physiologiques. Cependant, *T. lutea* n'étant pas un organisme modèle, l'annotation fonctionnelle de ses gènes reste parcellaire. Dans un tel cas, l'utilisation de WGCNA permet de pallier à ce problème d'annotation fonctionnelle en liant les gènes annotés ou non à des

traits physiologiques. De telles approches utilisant WGCNA ont été appliquées avec succès à différentes espèces, aboutissant à des résultats biologiquement significatifs (Kogelman *et al.*, 2014 ; Hollender *et al.*, 2014 ; Wang & Huang, 2014 ; El-Sharkawy *et al.*, 2015). Afin d'utiliser au mieux cet avantage, une priorisation de gènes visant à identifier les gènes liés aux caractéristiques physiologiques du phénotype mutant a été mise au point à partir des données générées par WGCNA. Cette priorisation permettant d'identifier les gènes liés aux caractéristiques physiologiques du phénotype mutant indépendamment de leur annotation, elle constitue un atout certain chez les organismes non-modèles. Dans cette étude, la priorisation des gènes a été fondée sur deux critères : (i) le coefficient de corrélation entre le profil d'expression des gènes et la dynamique de la quantité de lipides de stockage ou de carbohydrates cellulaires. Et (ii) la connectivité intra-modulaire permettant de quantifier, au sein d'un module de gènes co-exprimés, la connexion d'un gène aux autres gènes du module. Les gènes les plus connectés aux autres gènes du module sont appelés gènes « hub ». La position centrale de ces gènes « hub » au sein d'un module joue un rôle déterminant dans la fonction cellulaire du module en question (Jeong *et al.*, 2001 ; Carter *et al.*, 2004b ; Cooper *et al.*, 2006). Du fait de ce rôle clé, les gènes « hub » des modules d'intérêt (liés aux caractéristiques physiologiques du phénotype mutant) ont également été sélectionnés.

b) Résultats

Cette stratégie a été appliquée aux 15 333 gènes exprimés dans au moins un des 12 échantillons transcriptomiques (six conditions physiologiques par souche) de *T. lutea*. Vingt modules de gènes co-exprimés ont été identifiés (Figure 47 a). Trois d'entre eux sont significativement corrélés à la dynamique d'une des caractéristiques physiologique du phénotype mutant : un module significativement corrélé à la quantité de lipides neutres intracellulaires et deux modules significativement corrélés à la quantité de carbohydrates intracellulaires (Figure 47 b). Une analyse des fonctions « Gene Ontology (GO) » a été menée afin de mettre en évidence les fonctions enrichies parmi les gènes appartenant à chacun de ces trois modules. Bien qu'étant corrélés à un caractère phénotypique d'intérêt, ces modules présentent des fonctions variées (Annexe A).

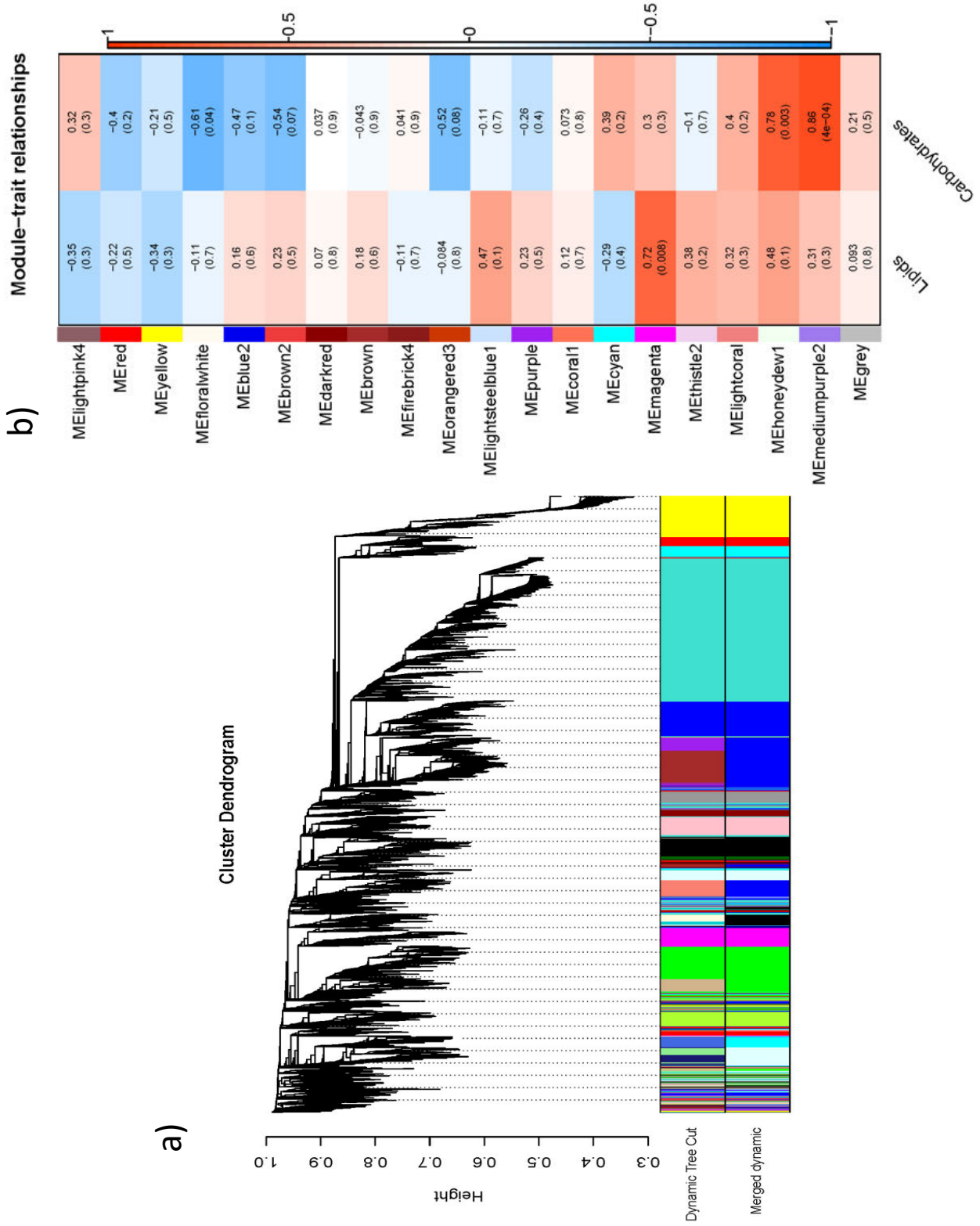

Figure 47 : Construction et analyse du réseau de co-expression des gènes. a) représentation de l'arbre de clusterisation hiérarchique, chaque branche correspond à un gène. Les couleurs assignées à chaque module par les methodes "Dynamic Tree Cut" et "Merged Dynamic" sont représentées en dessous de l'arbre. b) Corrélations entre le profil d'expression caractérisant chaque module et la dynamique des paramètres physiologiques. Chaque ligne correspond à un module et chaque colonne à un paramètre physiologique. Chaque case contient le coefficient de corrélation ainsi que la p-value associée.

Une telle diversité des fonctions liées au phénotype mutant peut s'expliquer par le fait que celui-ci est provoqué par un stress azoté. Ce type de conditions physiologiques est connu pour induire des changements variés, tant en terme d'expression de gènes que du point de vue des remaniements métaboliques, lesquels ne sont pas forcément liés directement à la synthèse de lipides de réserve ou de carbohydrates (Valledor *et al.*, 2014 ; Schmollinger *et al.*, 2014 ; Gargouri *et al.*, 2015). De plus, la mutagénèse aléatoire à l'origine de la souche mutante a provoqué de nombreuses mutations affectant des régions variées du génome. En conséquence, les modifications métaboliques relatives aux fonctions impactées sont, elles aussi, variées. Enfin, de nombreux gènes ayant une fonction putative inconnue, leur présence biaise l'enrichissement en fonctions GO.

Cette démarche a permis d'identifier des gènes liés aux caractéristiques physiologiques du phénotype mutant. La production de lipides de réserve et de carbohydrates étant conséquente à un stress azoté, les gènes priorisés sont donc également liés à la réponse de *T. lutea* à ce stress. Un enrichissement en fonctions GO au sein de ces gènes a mis en évidence des fonctions connues pour être impactées lors d'un stress azoté chez les microalgues (Annexe B), confirmant que cette stratégie est appropriée à l'identification de gènes liés à des caractères physiologiques d'intérêt. Parmi l'ensemble des gènes priorisés, quatre FTs ont été identifiés (LIM_1, MYB-3R et MYB-2R_14 liés aux lipides de réserve, et MYB-rel_12 lié aux carbohydrates). Ces FTs représentent des candidats intéressants pour la compréhension des mécanismes de régulation de *T. lutea* en réponse à un stress azoté.

2. Identifier les FTs candidats et leur rôle dans l'établissement du phénotype mutant par l'analyse de réseaux de régulation des gènes

a) Stratégie

La première étape de la stratégie a permis d'associer les gènes de *T. lutea* à une fonction putative ou à un caractère physiologique donné. Grâce à cette démarche, quatre FTs candidats ont pu être identifiés. La deuxième étape a pour but d'identifier les FTs impliqués dans l'établissement du phénotype mutant, puis de les caractériser via la fonction putative de leurs gènes cibles. Dans cette

optique, les gènes exprimés de façon différentielle entre la souche WTc1 et la souche 2Xc1 ont été identifiés. Ces gènes exprimés de façon différentielle étant supposés jouer un rôle clé dans l'établissement du phénotype mutant, un réseau de régulation des gènes a été construit pour chaque souche à partir de leur profil d'expression. L'analyse de ces réseaux permettra d'identifier les FTs potentiellement impliqués dans la régulation de ces gènes et, par conséquent, dans l'établissement du phénotype mutant. Un enrichissement en fonctions GO des gènes cibles des FTs de chaque réseau a été réalisé. La comparaison des deux réseaux a permis d'identifier des fonctions spécifiques à certains FTs du réseau de la souche 2Xc1. Afin de compléter l'analyse, les gènes priorisés pour leur lien avec les caractéristiques physiologiques du phénotype mutant ont été identifiés au sein des deux réseaux. Un enrichissement de ces gènes priorisés au sein des cibles des FTs de chaque réseau a ensuite été réalisé.

b) Résultats

Les gènes liés au phénotype mutant ayant été identifiés dans la première étape de la stratégie, les spécificités de régulation des gènes de la souche 2Xc1 ont été recherchées. Dans ce but, les gènes exprimés de façon différentielle entre les deux souches ont été identifiés pour chacun des six points de cinétique (échantillon Pt-27 WTc1 *vs* échantillon Pt-27 2Xc1, etc.) en utilisant le logiciel Gfold (Feng *et al.*, 2012). Parmi les 527 gènes exprimés de façon différentielle pour au moins un des six points de cinétique, se trouvent sept FTs (*MYB-2R_14*, *GATA_2*, *MYB-2R_20*, *MYB-rel_11*, *NF-YB_2*, *Fungal-TRF_8* et *HB-other_9_PAS*) dont deux exprimés uniquement chez la souche 2Xc1 (*MYB-2R_14* et *GATA_2*). Ces familles de FTs ont été identifiées comme étant impliquées dans le métabolisme de l'azote et la réponse des microalgues à un stress azoté (Marzluf, 1997 ; Avila *et al.*, 1998, 2002 ; Imamura *et al.*, 2009 ; Gargouri *et al.*, 2015). De plus, ces sept FTs étant exprimés de façon différentielle chez la souche 2Xc1, leur implication dans la réponse aux variations de la disponibilité en azote semble spécifique chez cette souche. Cette expression différentielle suggère donc une implication de ces FTs dans l'établissement du phénotype mutant. Parmi eux, le FT MYB-2R_14 occupe une position de gène « hub » au sein du module WGCNA, corrélée à la dynamique de la quantité de lipides de réserve. De plus, son propre profil d'expression est corrélé avec la dynamique de ce même paramètre physiologique, et il n'est

exprimé que chez la souche mutante. Autant de critères qui en font un candidat particulièrement intéressant.

Afin d'identifier le rôle de ces sept FTs dans les spécificités de régulations de l'expression des gènes de la souche 2Xc1, les 527 gènes exprimés de façon différentielle ont été utilisés pour construire un réseau de régulation des gènes pour chacune des deux souches (Figure 48). Chacun des deux réseaux est constitué de quatre communautés. Ces communautés sont constituées de gènes plus liés les uns aux autres qu'ils ne le sont aux autres gènes du réseau. La topologie du réseau de la souche 2Xc1 est très proche de celle du réseau de la souche WTc1. Dans les deux cas, la communauté violette apparait dirigée par le FT *Fungal-TRF_8*. Il en va de même concernant la communauté bleue dirigée par le FT *HB-other_9_PAS*. La communauté rouge est dirigée par le seul FT *NF-YB_2* chez la souche WTc1. En revanche, chez la souche 2Xc1, cette communauté est également dirigée par le FT *MYB-2R_14* (celui-ci n'étant exprimé que chez la souche mutante). De même, la communauté verte est dirigée par les deux FTs *MYB-rel_11* et *MYB-2R_20* chez la souche WTc1, alors que chez la souche 2Xc1, elle l'est également par le FT *GATA_2* (uniquement exprimé chez la souche mutante). Afin de mieux caractériser le rôle de ces FTs dans l'établissement du phénotype mutant, un enrichissement en fonctions GO a été réalisé dans les communautés de chaque réseau (Annexe C et D). Leur comparaison a ensuite permis de mettre en évidence les fonctions enrichies uniquement au sein d'une communauté du réseau de la souche 2Xc1. En complément de cette annotation, la priorisation de gènes a été appliquée aux deux réseaux de régulation (Tableau 4). Un enrichissement de ces gènes priorisés au sein des cibles potentielles de chaque FTs a ensuite été réalisé.

Ainsi, la communauté bleue, dirigée par le FT *HB-other_9_PAS*, semble liée à la dégradation, la maturation et le trafic intracellulaire des protéines. Cette observation est en adéquation avec la littérature. En effet, en condition de privation azotée, le protéome des microalgues est fortement affecté, notamment via une diminution de la synthèse protéique et une augmentation de la protéolyse (Dong *et al.*, 2013 ; López García de Lomana *et al.*, 2015 ; Alipanah *et al.*, 2015). Ces mécanismes permettent de fournir de l'énergie pour la synthèse de lipides (Msanne *et al.*, 2012).

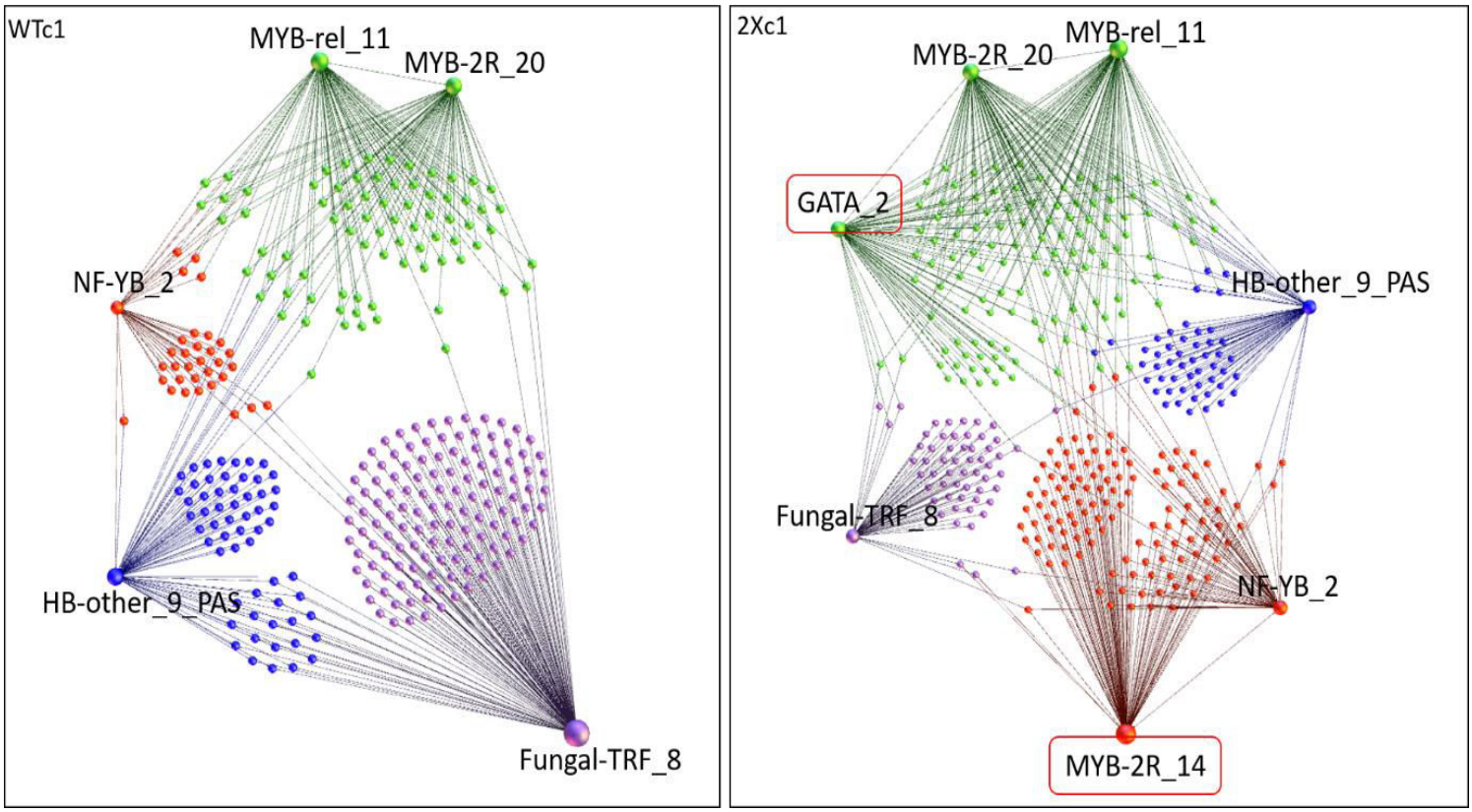

Figure 48 : Représentation des réseaux de régulation des gènes des deux souches de *T. lutea* grâce au logiciel Gephi. Chaque nœud du réseau représente un gène. La co-expression d'un gène avec un FT est visualisée par un lien entre ces deux gènes. La taille des gènes est proportionnelle au nombre de connexion qu'ils comportent. Chaque communauté est représentée par une couleur. Les deux FTs encadrés en rouge ne sont exprimés que chez la souche mutante.

Tableau 4 : gènes priorisés au sein des gènes exprimés de façon différentielle. Plus la valeur de KME est proche de 1, plus le gène est central au sein de son module. Les gènes ayant une valeur de KME supérieur ou égale à 0.95 ont été considérés comme des gènes hub. Les gènes ont été priorisés si leur profil d'expression était corrélé à la dynamique de la quantité de lipides ou de carbohydrates cellulaire, ou s'ils étaient des gènes hub de l'un des trois modules d'intérêt. Nd : non déterminée.

Gène	Corrélation avec la dynamique carbohydrates	Corrélation avec la dynamique lipides	KME	Module	Fonction putative
Tiso_gene_3094	0,77		0,96	CR1 module	nd
Tiso_gene_5367	0,70		0,96	CR1 module	nd
Tiso_gene_5511	0,74		0,95	CR1 module	nd
Tiso_gene_5871	0,86		0,95	CR2 module	nd
Tiso_gene_14258	0,84		0,95	CR2 module	nd
Tiso_gene_11673	0,89		0,95	CR2 module	nd
Tiso_gene_9025		0,76	0,97	LR module	Acetyl-CoA synthetase-like
Tiso_gene_20182		0,70	0,96	LR module	nd
Tiso_gene_18537		0,73	0,96	LR module	nd
Tiso_gene_8095		0,67	0,95	LR module	MYB-2R_14
Tiso_gene_19516		0,61	0,95	LR module	nd
Tiso_gene_19445		0,68	0,95	LR module	nd
Tiso_gene_6707		0,70	0,95	LR module	nd
Tiso_gene_18899		0,71	0,95	LR module	Tyrosine-protein kinase ephrin type A/B receptor-like
Tiso_gene_11571	0,92			CR2 module	Regulator of chromosome condensation RCC1
Tiso_gene_13889	0,86			CR2 module	nd
Tiso_gene_6544	0,84			CR2 module	nd
Tiso_gene_6760	0,84			LR module	nd
Tiso_gene_13518	0,83			CR2 module	ATP-binding cassette transporter
Tiso_gene_7210	0,82			CR2 module	ATP-binding cassette transporter
Tiso_gene_9366	0,80			CR2 module	Saccharopine dehydrogenase
Tiso_gene_17273	0,80			CR2 module	nd
Tiso_gene_20489	0,80				nd
Tiso_gene_4808	0,78			CR2 module	Calcium-independent phospholipase A2-gamma
Tiso_gene_16415	0,77			CR2 module	nd
Tiso_gene_2919	0,77			CR2 module	Neurotransmitter-gated ion-channel ligand-binding domain containing protein
Tiso_gene_20165	0,76			CR2 module	nd
Tiso_gene_3564	0,76			CR2 module	nd
Tiso_gene_11856	0,75			CR2 module	nd
Tiso_gene_4035		0,91		CR2 module	nd
Tiso_gene_19975		0,79		LR module	nd
Tiso_gene_9964		0,77		CR1 module	nd
Tiso_gene_13835		0,76		LR module	nd
Tiso_gene_14115		0,76		LR module	nd
Tiso_gene_15057		0,76		LR module	ADP-ribosylation factor
Tiso_gene_20336		0,75		LR module	nd

De plus, parmi les gènes cibles du FT *HB-other_9_PAS*, un gène codant pour la Diacylglycerol-O-acyltransferase, enzyme impliquée dans la synthèse de triglycérides en conditions de privation d'azote (Msanne *et al.*, 2012), ainsi que deux gènes impliqués dans la synthèse de carbohydrates ont été identifiés (un gène codant une énolase et un gène codant une mannose-6-phosphate isomérase).

Quant à la communauté rouge dirigée par les FTs *MYB-2R_14* et *NF-YB_2*, elle semble liée à la prévention et la réparation des dommages causés par le stress oxydatif induit par la photosynthèse. En effet, quatre gènes impliqués dans la structure des antennes collectrices de lumière sont présents dans cette communauté. La photosynthèse est largement affectée lors d'un stress azoté chez les microalgues (Msanne *et al.*, 2012 ; Schmollinger *et al.*, 2014 ; Juergens *et al.*, 2015 ; Alipanah *et al.*, 2015 ; Gargouri *et al.*, 2015). De plus, deux gènes codant une DNA photolyase et une cryptochrome DASH appartiennent à cette communauté. Ces familles de protéines sont impliquées dans la réparation de l'ADN et, plus spécifiquement, les dimères de pyrimidines (Selby & Sancar, 2006). Or, l'apparition de ces dimères de pyrimidines est induite par une irradiation aux UV ainsi que par un stress oxydatif (Hochberg *et al.*, 2006), lequel est induit par la privation d'azote chez les algues (Liu *et al.*, 2012). De plus, le stress oxydatif a été proposé comme étant un des mécanismes régulateurs de l'accumulation de lipides en conditions de stress azoté chez les microalgues dont les mécanismes sous-jacents sont encore inconnus (Zhang *et al.*, 2013 ; Yilancioglu *et al.*, 2014). Il est suggéré que les composés oxydatifs générés par la photosynthèse sont consommés par la synthèse de TAG, prévenant ainsi un stress oxydatif (Solovchenko, 2012). Enfin, les gènes priorisés pour leur lien avec la quantité cellulaire de lipides de réserve sont enrichis au sein des cibles des deux FTs de cette communauté (*MYB-2R_14* et *NF-YB_2*) dans le réseau de la souche 2Xc1. En revanche, ce n'est pas le cas dans le réseau de la souche WTc1. Autant d'indices convergents vers un lien des FTs *MYB-2R_14* et *NF-YB_2* avec l'accumulation de lipides de réserve chez la souche 2Xc1.

Dans le cas de la communauté verte, dirigée par les FTs *MYB-2R_20*, *MYB-rel_11* et *GATA_2*, les fonctions enrichies sont plus variées. Toutefois, trois fonctions associées à l'utilisation de l'azote en condition de privation azotée sont enrichies : l'absorption d'azote avec un transporteur d'ammonium, et la remobilisation d'azote cellulaire avec des gènes impliqués dans la dégradation des protéines et un transporteur d'urée. Le cycle de l'urée est utilisé pour le recyclage de l'azote

cellulaire chez les diatomées (Allen *et al.*, 2011) ainsi que chez l'haptophyte *E. huxleyi* (Rokitta *et al.*, 2011). De plus, l'ensemble des enzymes de ce cycle a été identifié dans le génome de *T. lutea*, ce qui renforce l'hypothèse de son utilisation chez cette espèce. Quant à la protéolyse, elle permet de recycler l'azote des acides-aminés et nourrir les voies de synthèse des composés de stockage du carbone (da Silva *et al.*, 2009 ; Dong *et al.*, 2013 ; Alipanah *et al.*, 2015). De plus, une UDP-glucose pyrophosphorylase impliquée dans la synthèse de carbohydrates a également été identifiée dans cette communauté. Chez les haptophytes, les carbohydrates peuvent être stockés sous forme de chrysolaminarine (Sadovskaya *et al.*, 2014 ; Wang *et al.*, 2014). Chez la diatomée *P. tricornutum*, une UDP-glucose pyrophosphorylase a justement été proposé comme enzyme clé dans l'allocation du carbone et la synthèse de chrysolaminarine (Zhu *et al.*, 2016). Enfin, les gènes priorisés pour leur lien avec la quantité de carbohydrates cellulaires étaient enrichis au sein des cibles du FT *MYB-rel_11* dans le réseau de la souche 2Xc1. Ces résultats suggèrent donc une implication des FTs de la communauté verte, et particulièrement du FT *MYB-rel_11*, dans l'absorption et le recyclage de l'azote cellulaire ainsi que dans la synthèse de carbohydrates chez la souche 2Xc1.

Cette association complémentaire de l'analyse des données du réseau de co-expression des gènes à celles des réseaux de régulation a donc permis d'identifier trois communautés de gènes potentiellement associés au phénotype de la souche mutante de *T. lutea*. La communauté verte semble particulièrement intéressante puisque liée au recyclage de l'azote, un mécanisme clé dans la production de lipides et de carbohydrates en condition de privation azotée chez les microalgues. Afin de mieux caractériser le rôle des FTs *MYB-2R_20* et *MYB-rel_11* dirigeant cette communauté, leur expression a été suivie par q-RT-PCR au fil de la cinétique complète de la deuxième injection d'azote.

III. Confirmation de l'implication des FTs *MYB-2R_20* et *MYB-rel_11* dans le recyclage de l'azote et du carbone lors d'une privation azotée chez la souche 2Xc1 de *T. lutea*

Chez les microalgues exposées à une privation azotée, le recyclage et l'absorption de l'azote sont des mécanismes clés qui font également intervenir des transporteurs à haute affinité. Dans de telles

conditions, ces transporteurs à haute affinité nitrate/nitrite (NRT2) permettent d'augmenter l'efficacité de l'absorption de l'azote (Hildebrand & Dahlin, 2000 ; Song & Ward, 2007 ; Kang *et al.*, 2007). Quatre gènes de la famille *Nrt2* (*Nrt*2.1, *Nrt*2.2, *Nrt*2.3 et *Nrt*2.4) ont précédemment été identifiés chez *T. lutea* (Charrier *et al.*, 2015). Cependant, leurs séquences nucléotidiques sont trop proches pour permettre la quantification de leur expression respective par RNA-seq. La communauté verte étant liée à ces mécanismes chez la souche 2Xc1 de *T. lutea*, leur co-expression avec les *Nrt*2 au cours de la deuxième injection d'azote a été évaluée par q-RT-PCR. Deux gènes supplémentaires ont été ajoutés à cette analyse : la Periplasmic L-Amino-Acid Oxidase (PLAAOx) et la Coccolite Scale Associated Protein (CSAP). Ces deux protéines, accumulées de façon différentielle chez la souche 2Xc1 par rapport à la souche WTc1 en condition de carence azotée, sont supposées jouer un rôle dans la réponse différentielle de ces deux souches (Garnier *et al.*, 2014). Les expressions relatives de chacun de ces huit gènes ont été utilisées afin de d'évaluer leur co-expression.

Dans chacune des deux souches, les trois gènes codant pour la PLAAOx, la CSAP et le transporteur *Nrt*2.1 sont fortement co-exprimés (Tableau 5). Cette co-expression commune suggère donc l'implication de mécanismes de régulation communs aux deux souches. Toutefois, ces trois gènes sont sous-exprimés chez la souche 2Xc1 tout au long de la cinétique de l'injection d'azote (Tableau 6). Cette expression différentielle suggère donc la présence de mécanismes de régulation spécifiques à la souche 2Xc1. Ce schéma de régulation, spécifique à cette souche, est retrouvé vis-à-vis des deux FTs MYB-2R_20 et MYB-rel_11. En effet, le FT MYB-2R_20 est co-exprimé avec les gènes de la PLAAOx, de la CSAP et du transporteur de nitrate *Nrt*2.1 chez les deux souches (Tableau 5). Le FT MYB-2R_20 serait donc impliqué dans la co-régulation de l'expression de ces trois gènes chez les deux souches de *T. lutea*. En revanche, le FT MYB-rel_11 n'est co-exprimé avec aucun de ces gènes chez la souche WTc1 mais seulement co-exprimé avec les gènes de la PLAAOx, de la CSAP, de *Nrt*2.1 et du FT MYB-2R_20 chez la souche 2Xc1 (Tableau 5). Cette co-expression spécifique à la souche mutante suggère donc l'implication du FT MYB-rel_11 dans l'expression différentielle des gènes de la PLAAOx, de la CSAP et du *Nrt*2.1 chez cette souche.

Tableau 5 : corrélation des profils d'expression obtenus par q-RT-PCR. Pour chaque couple de gènes, le coefficient de corrélation de Spearman est donné ainsi que la p-value associée.

		2Xc1		WTc1	
		CCS	p-value	CCS	p-value
MYB-rel_11	MYB-2R_20	0,83	0,000024	0,64	0,005070
	CSAP	0,76	0,000877	0,37	0,127717
	PLAOOX	0,73	0,001455	0,48	0,048095
	NRT2.1	0,74	0,001342	0,48	0,048546
MYB-2R_20	MYB-rel_11	0,83	0,000024	0,64	0,005070
	CSAP	0,84	0,000058	0,70	0,001566
	PLAOOX	0,82	0,000135	0,75	0,000507
	NRT2.1	0,79	0,000205	0,71	0,001325
CSAP	MYB-rel_11	0,76	0,000877	0,37	0,127717
	MYB-2R_20	0,84	0,000058	0,70	0,001566
	PLAOOX	0,97	0,000007	0,95	0,000001
	NRT2.1	0,92	0,000001	0,94	0,000001
PLAOOX	MYB-rel_11	0,73	0,001455	0,48	0,048095
	MYB-2R_20	0,82	0,000135	0,75	0,000507
	CSAP	0,97	0,000007	0,95	0,000001
	NRT2.1	0,93	0,000002	0,97	0,000008
NRT2.1	MYB-rel_11	0,74	0,001342	0,48	0,048546
	MYB-2R_20	0,79	0,000205	0,71	0,001325
	CSAP	0,92	0,000001	0,94	0,000001
	PLAOOX	0,93	0,000002	0,97	0,000008

Tableau 6 : expression différentielle des gènes codant la CSAP, la PLAAOx et le Nrt2.1 dans chacun des 17 échantillons représentant la cinétique de la deuxième injection d'azote.

sample	CSAP	PLAAOx	NRT2.1
74	-5,16	-5,16	-2,74
81	-2,61	-2,18	-2,17
82	-3,74	-3,61	-3,30
83	-3,54	-2,67	-2,04
84	-2,66	-2,06	-1,82
85	-1,14	-1,11	1,04
88	-3,98	-3,95	-5,43
89	-1,45	-1,69	-1,48
90	-1,61	-1,49	-1,45
91	-2,78	-1,98	-2,39
92	-2,50	-3,30	-2,47
94	-1,36	-1,74	-1,59
95	-3,05	-2,90	-2,73
96	-4,81	-5,52	-3,91
97	-2,41	-2,23	-2,17
99	-1,42	-1,19	-1,09
101	-1,77	-1,63	-0,27

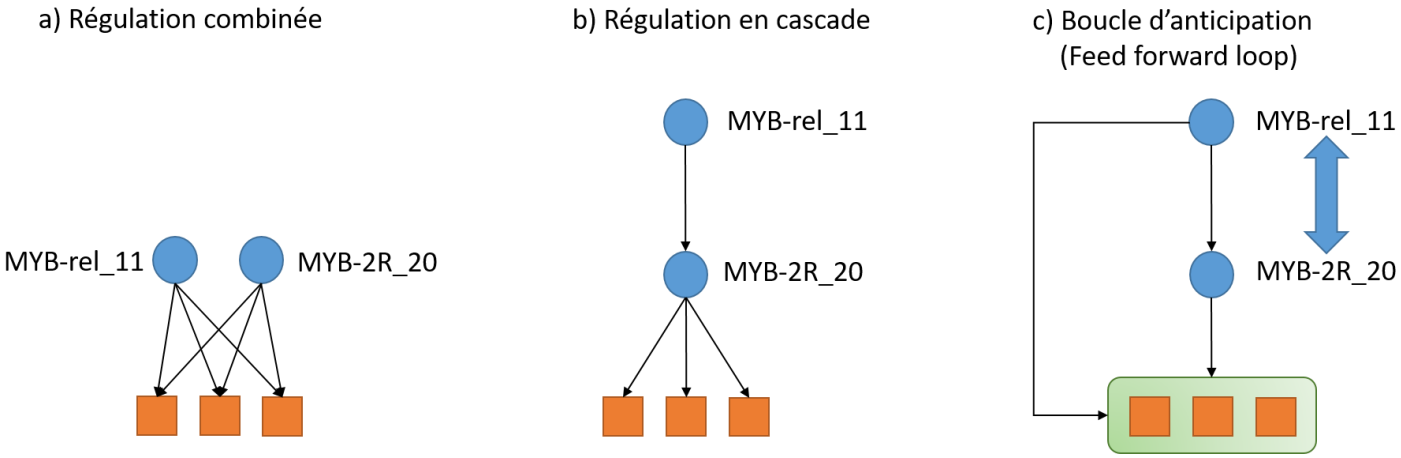

Figure 49 : hypothèses concernant l'action du FT MYB-rel_11 dans la régulation de la réponse spécifique de la souche 2Xc1. Les FTs MYB-rel_11 et MYB-2R_20 sont représentés par des cercles bleus et leurs gènes cibles potentiels (PLAAOx, CSAP et Nrt2.1) par des carrés oranges. a) régulation combinée des trois gènes cibles par les deux FTs. b) régulation en cascade dans laquelle le FT MYB-rel_11 régule l'expression des trois gènes cibles par l'intermédiaire du FT MYB-2R_20. c) boucle d'anticipation (feed forward loop) dans laquelle le FT MYB-rel_11 régule l'expression des trois gènes cibles à la fois directement et par l'intermédiaire du FT MYB-2R_20. Dans cette dernière hypothèse, le rôle des deux FTs peut être inversé sans démentir les co-expressions observées.

Ce mécanisme de régulation peut revêtir trois modes d'action différents : (i) la régulation directe de l'expression des gènes de la PLAAOx, de la CSAP et du *Nrt*2.1 en coordination avec le FT MYB-2R_20 (Figure 49, a). (ii) La régulation en cascade de ces trois gènes par l'intermédiaire du FT MYB-2R_20, celui-ci régulant l'expression des trois gènes cibles, comme chez la souche WT (Figure 49, b). Ou alors (iii) la régulation de l'expression des gènes de la PLAAOx, de la CSAP et du *Nrt*2.1 à la fois directement et via l'intermédiaire du FT MYB-2R_20, formant une boucle d'anticipation (« feed forward loop ») (Figure 49, c). Dans cette dernière hypothèse, le rôle des deux FTs peut être inversé sans pour autant démentir les co-expressions observées : le FT MYB-2R_20 régulant les trois gènes cibles directement ainsi que par l'intermédiaire du FT MYB-rel_11. Dans chacune de ces hypothèses, le FT MYB-rel_11 semble jouer un rôle clé dans la réponse spécifique de la souche 2Xc1 à un stress azoté.

Outre ces régulations spécifiques de souche, le profil d'expression de ces gènes est très particulier et semble dépendre de l'état physiologique de la microalgue. Leur expression est réduite ou réprimée très rapidement à la suite de l'injection d'azote (Figure 50). Leur expression augmente ensuite graduellement à mesure que le rapport N/C se stabilise (Figure 50). Ces valeurs élevées de N/C sont la conséquence de l'absorption par les microalgues de l'azote injecté (réplétion azotée). En conséquence, une augmentation du rapport N/C est synonyme d'une diminution de la quantité d'azote disponible dans le milieu de culture. Une telle expression dépendante de la disponibilité en azote a été montrée précédemment pour le transporteur *Nrt*2.1 chez *T. lutea* (Charrier *et al.*, 2014). Le fait que les gènes de la PLAAOx et de la CSAP possèdent un profil d'expression similaire, suggère une fonction liée à la disponibilité en azote. Une protéine homologue de la PLAAOx est différentiellement accumulée chez la chlorophyte *C. reinhardtii* en condition de privation azotée et est supposée fournir de l'ammonium aux cellules via la déamination des acides-aminés (Wase *et al.*, 2014 ; Aksoy *et al.*, 2014). De plus, un gène homologue de la CSAP est exprimé de façon différentielle chez la diatomée *P. tricornutum* en condition de privation d'azote. Ce gène est supposé jouer un rôle dans l'homéostasie du carbone grâce à sa fonction putative de décarboxylase (Valenzuela *et al.*, 2012). Les protéines PLAAOx et CSAP pourraient donc participer au recyclage de l'azote et du carbone à partir des acides-aminés libres produits par protéolyse, permettant ainsi à la cellule de faire face au manque d'azote extracellulaire. Toutefois, la localisation subcellulaire de ces deux protéines n'est pas identifiée.

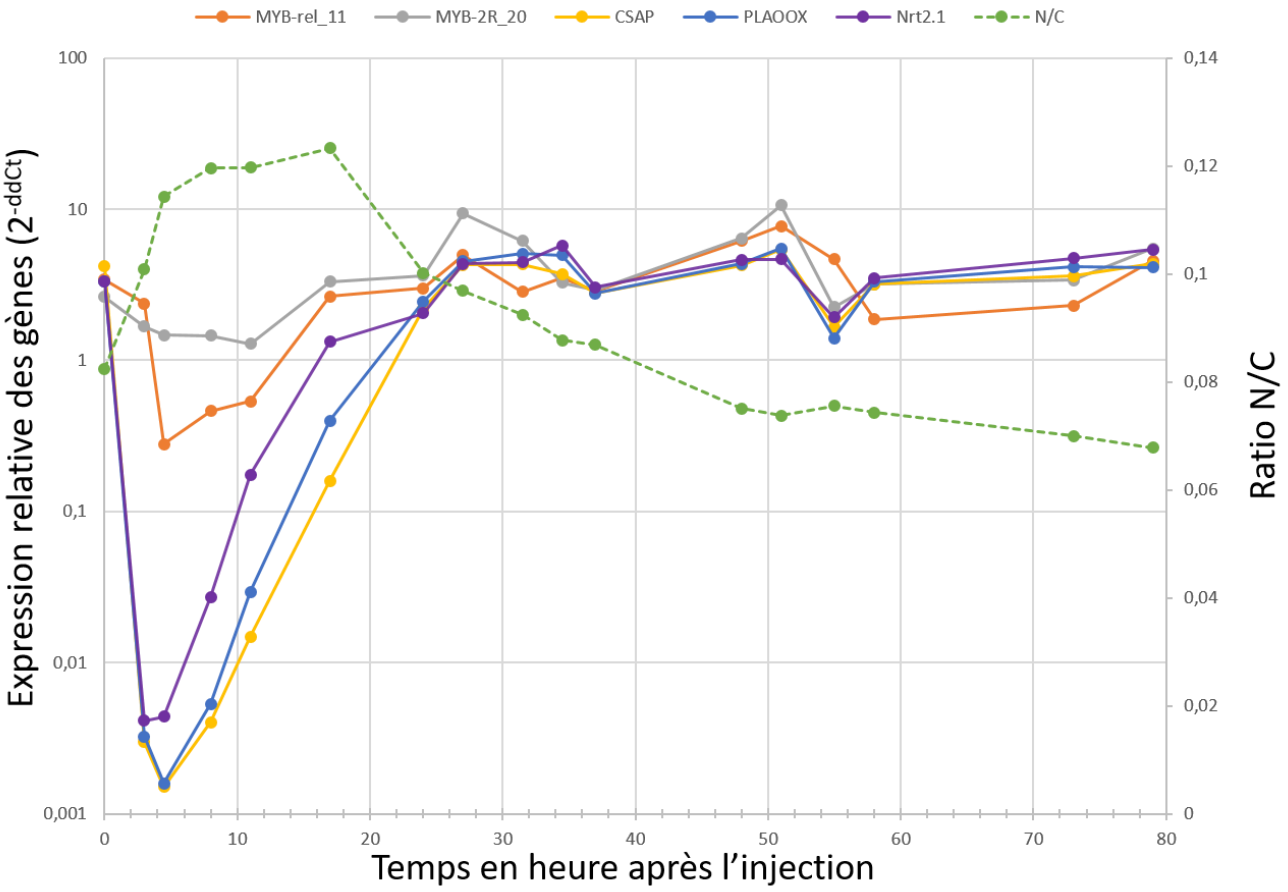

Figure 50 : profil d'expression par q-RT-PCR des gènes codant la CSAP, la PLAAOx, le Nrt2.1 et les FTs MYB-rel_11 et MYB-2R_20 chez la souche 2Xc1 suite à la deuxième injection d'azote. Les valeurs d'expression de gène sont normalisées par deux gènes de ménage. L'expression relative est calculée par rapport à l'échantillon précédent l'injection d'azote (état d'équilibre).

En fonction de celle-ci, elles peuvent agir au niveau des acides-aminés libres intra- ou extracellulaires. Néanmoins, quel que soit leur localisation, le recyclage de l'azote et du carbone pourrait nourrir la synthèse de protéines essentielles ainsi que celle des formes de stockage du carbone telles les lipides et les carbohydrates.

A la lumière de ces résultats, le FT MYB-2R_20 semble donc impliqué dans la co-régulation des gènes de la PLAAOx, de la CSAP et du *Nrt*2.1 chez les deux souches, alors que le FT MYB-rel_11 apparait plutôt impliqué dans la régulation de ces gènes chez la souche mutante spécifiquement. Cette analyse par q-RT-PCR confirme donc l'implication potentielle de ces deux FTs dans la mise en place des mécanismes spécifique à la souche 2Xc1 en réponse à une privation d'azote. A la fois l'analyse des réseaux de gènes et l'analyse q-RT-PCR vont dans le sens d'une implication des FTs MYB-2R_20 et MYB-rel_11 dans le recyclage de l'azote et du carbone des acides-aminés produits par protéolyse chez la souche mutante. La présence d'une UDP-glucose pyrophosphorylase parmi les cibles du FT MYB-rel_11 ainsi que leur enrichissement en gènes priorisés pour leur lien avec la quantité de carbohydrates cellulaires suggère que les éléments recyclés seraient utilisés pour la synthèse de carbohydrates.

IV. Résultats complémentaires : L'analyse de réseaux de co-expression des FTs offre une vue globale de la régulation de la réponse au stress azoté

Ces travaux ont identifié sept régulateurs potentiels de la réponse de la souche 2Xc1 de *T. lutea* à un stress azoté. Le FT Fungal-TRF_8 qui, bien qu'exprimé de façon différentielle, semble réguler les mêmes fonctions chez les deux souches de *T. lutea* (transport, métabolisme des carbohydrates, modification des protéines par phosphorylation, photosynthèse et synthèse de chlorophylle). Le FT HB-other_9_PAS apparaît, lui, être impliqué dans la dégradation, le trafic et la maturation des protéines ainsi que dans la synthèse des TAGs et des carbohydrates. Les FTs NF-YB_2 et MYB-2R_14 qui dirigent la communauté rouge, semblent liés à la photosynthèse ainsi qu'à la réponse au stress oxydatif qu'elle provoque, notamment via la synthèse de TAG. Enfin, les FTs GATA_2, MYB-2R_20 et MYB-rel_11, dirigeant la communauté verte, semblent liés au recyclage de l'azote et du carbone afin d'alimenter la synthèse de lipides de réserve et de carbohydrates. Par souci de

clarté à l'égard du lecteur, le nom des FTs GATA_2, MYB-2R_20 et MYB-rel_11 qui dirigent la communauté verte seront écrits en vert et celui des FTs NF-YB_2 et MYB-2R_14 qui dirigent la communauté rouge seront écrits en rouge.

Au même titre que leurs gènes cibles, le niveau d'expression des FTs est régulé par d'autres FTs. L'identification des FTs régulant l'expression de ces sept régulateurs candidats permettrait d'atteindre un niveau supérieur dans la compréhension de la régulation de la réponse de *T. lutea* à un stress azoté. Dans ce but, un réseau de co-expression de l'ensemble des FTs exprimés a été réalisé pour chacune des deux souches à partir des données RNA-seq. Au sein des deux réseaux, les gènes « hub » et « bottleneck » (goulot d'étranglement) ont été identifiés. Au contraire des gènes « hub » caractérisés par un nombre élevé de connexions et situés au centre des communautés du réseau, les gènes « bottleneck » (BN) sont situés entre les différents communautés et sont caractérisés par une centralité d'intermédiarité (betweenness centrality) élevée. Sommairement, la centralité d'intermédiarité correspond au nombre de fois qu'un gène est traversé par le plus court chemin entre deux nœuds du réseau (Figure 51). Le plus court chemin entre chaque paire de nœuds du réseau est identifié et les gènes constituant un point de passage du plus grand nombre de ces plus courts chemins sont définis comme des « bottleneck » (goulots d'étranglements) du réseau (Figure 52). Bien qu'ayant parfois peu de connexions, ces gènes sont traversés par les flux venant de plusieurs communautés, faisant le lien entre elles (Figure 52). De par cette position clé, de tels gènes sont primordiaux pour le bon fonctionnement des réseaux biologiques, d'autant plus dans le cas de réseaux de régulation des gènes dans lesquels la notion de flux d'information est cruciale. Plusieurs études ont montré que de tels gènes étaient essentiels aux réseaux biologiques, d'autant plus s'ils combinaient une position de « hub » et de BN (Figure 52) (Hahn & Kern, 2005 ; Yu *et al.*, 2007 ; McDermott *et al.*, 2009). Dans le but d'analyser le réseau des deux souches, et conformément à la littérature, les 20% des gènes ayant le plus de connexions ont été considérés comme « hub » et les 20% des gènes ayant une centralité d'intermédiarité la plus élevée comme des gènes BN (Tableau 7) (Yu *et al.*, 2007 ; McDermott *et al.*, 2009). Dans le réseau de la souche WTc1 (Figure 53) deux communautés distinctes sont clairement identifiées. Il est intéressant de noter que trois FTs différentiellement exprimés entre les deux souches sont des BN de ce réseau (HB-other_9_PAS, NF-YB_2 et Fungal-TRF_8). Du fait de la position clé de ces gènes au sein du réseau, leur expression différentielle chez la souche 2Xc1 suggère de profondes modifications structurelles du réseau de la souche mutante.

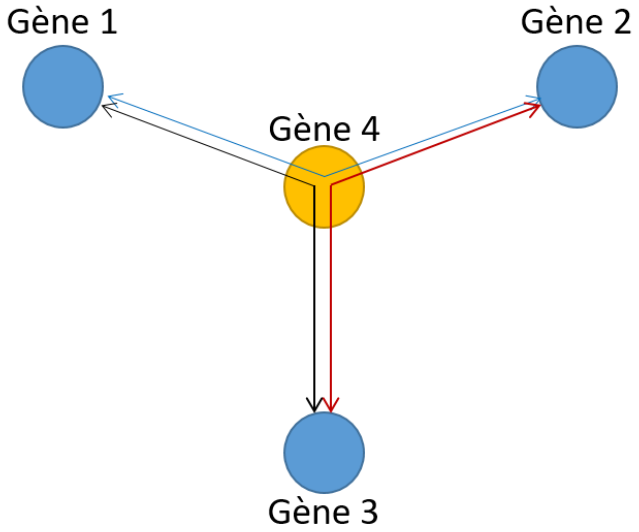

Figure 51 : illustration de la notion de centralité d'intermédiarité. Au sein de ce réseau simplifié, les flèches bleues représentent le plus court chemin entre les gènes 1 et 2, les flèches rouges entre les gènes 2 et 3, et les flèches noires entre les gènes 1 et 3. Les gènes 1, 2 et 3 ne sont que les point de départ et d'arriver de chaque plus court chemin du réseau. Le gène 4, en revanche, est traversé par tous les plus courts chemins. La centralité d'intermédiarité du gène 4 étant égale au nombre de plus courts chemins le traversant, il constitue un goulot d'étranglement du réseau.

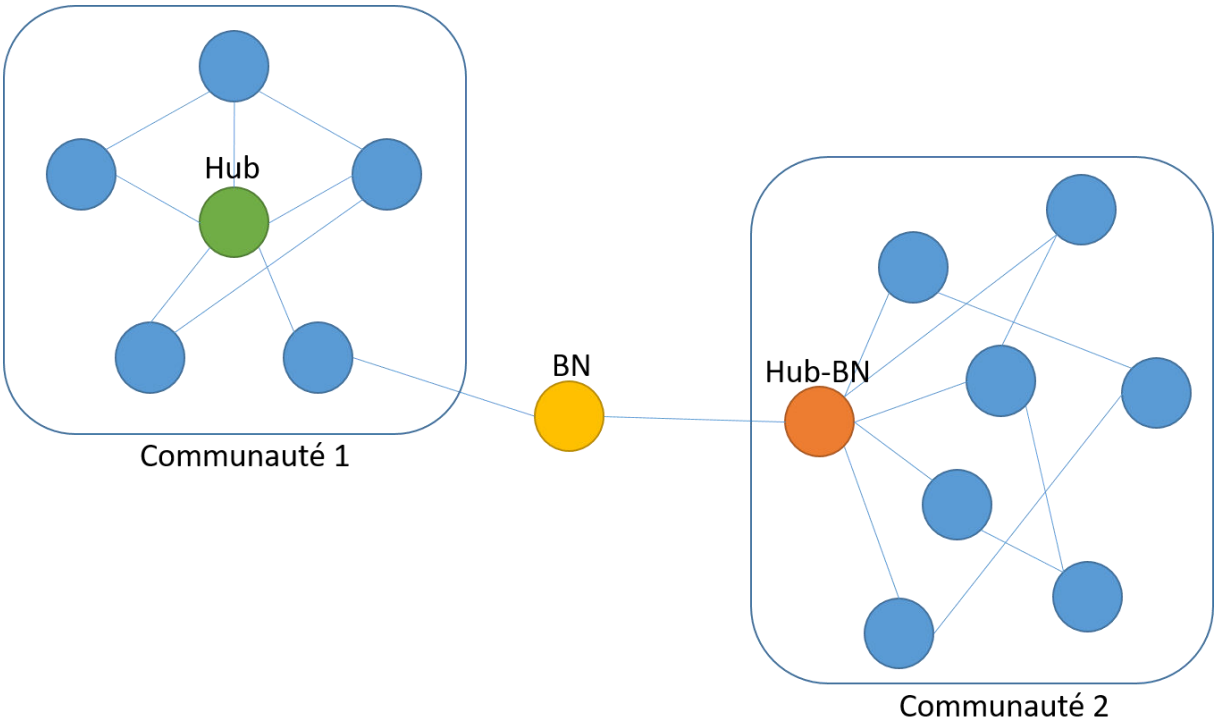

Figure 52 : représentation schématique de la notion de « bottleneck ». Schémas d'un réseau de co-expression des gènes. Le nœud en vert est un « hub » du fait de son nombre élevé de connexions. Le nœud en jaune est un « bottleneck » (BN) ou goulot d'étranglement puisqu'il est au carrefour des flux d'informations entre les deux communautés du réseau. Le nœud en orange cumule les fonctions de « hub » et de BN.

Tableau 7 : liste des FTs constituant des "hub" et des goulots d'étranglements ("bottleneck") au sein des réseaux de co-expression des FTs des deux souches de *T. lutea* (a) souche mutante et b) souche WTc1). Les FTs en rouges ont un profil d'expression corrélé à la dynamique de la quantité de lipides de stockage et les FTs en gras et soulignés sont les FTs exprimés de façon différentielle chez la souche 2Xc1.

a)

2Xc1

bottleneck	hub
HB-other_11_PAS	HB-other_11_PAS
TUB_2	mTERF_5
HB-other_14	**Fungal-TRF_8**
LSD	CSD_1
C2H2_2	HB-other_3
MYB-3R	HSF_5
HB-other_5	**GATA_2**
Fungal-TRF_8	HB-other_14
Fungal-TRF_7	MYB-2R_13
HSF_8_PAS	C3H_12
mTERF_5	MYB-2R_22

b)

WTc1

bottleneck	hub
Sigma-70_2	MYB-rel_22
HB-other_9_PAS	G2-like_3
MYB-rel_13	HSF_7_2PAS
HB-other_11_PAS	Sigma-70_3
HSF_1	MYB-2R_1
mTERF_4	C3H_3
ERF_2	C3H_5
HB-other_2	MYB-2R_17
Fungal-TRF_7	GATA_1
LIM_2	Fungal-TRF_3
CSD_1	MYB-2R_13
NF-YB_2	HSF_2
C3H_13	MYB-2R_21
MYB-rel_16	MYB-rel_4
Fungal-TRF_8	HB-other_12
Fungal-TRF_12	HSF_1
HB-other_6_PAS	**Fungal-TRF_8**

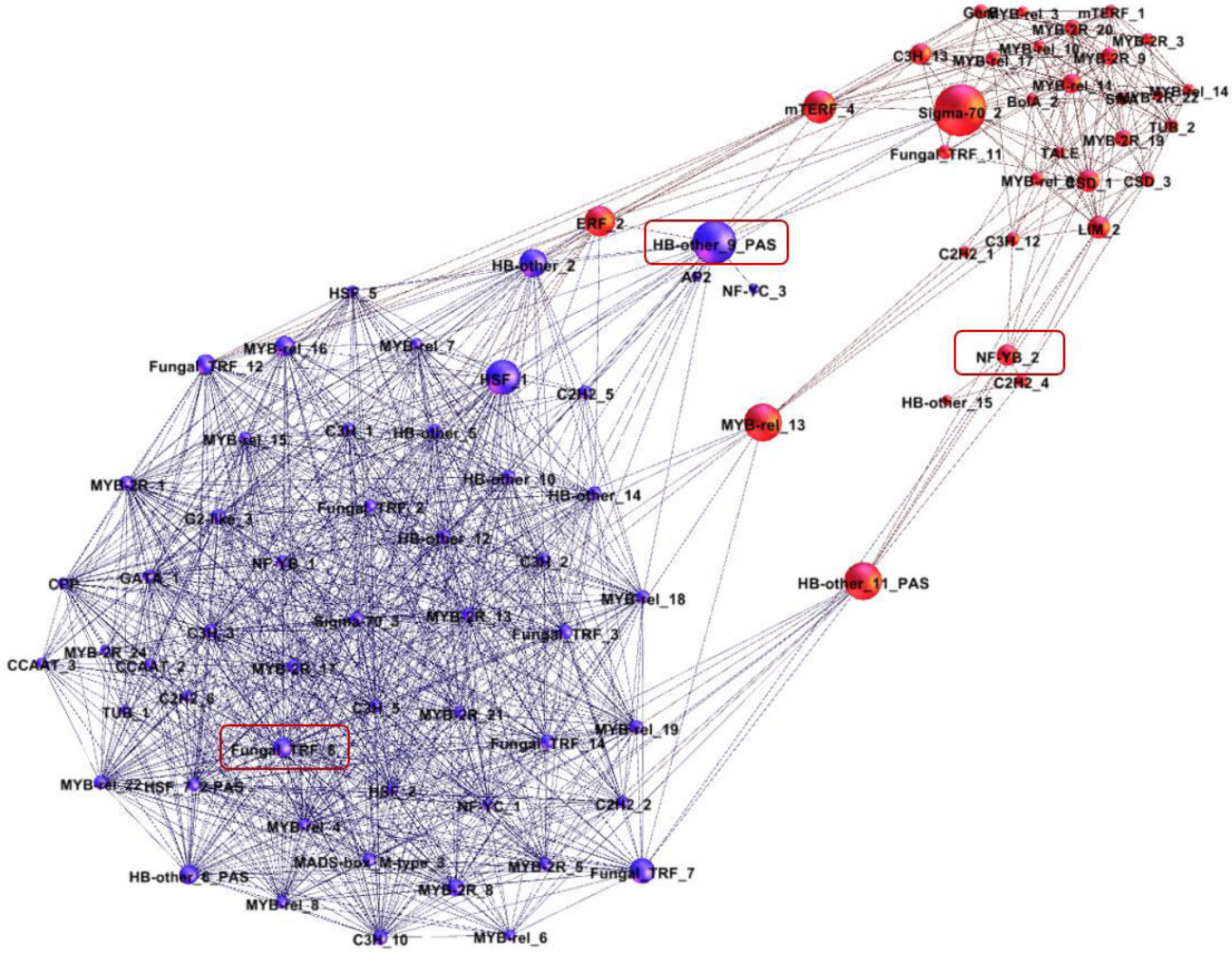

Figure 53: représentation du réseau de co-expression des FTs de la souche WTc1. la taille des nœuds est proportionnelle à leur valeur de centralité d'intermédiarité. Les nœuds les plus gros sont donc les « bottlenecks » ou goulots d'étranglement du réseau. Les trois FTs exprimés de façon différentielle et ayant une fonction de goulot d'étranglement de ce réseau sont encadrés en rouge.

En effet, la topologie du réseau de la souche 2Xc1 (Figure 54, a) est très différente. Celui-ci est divisé en quatre communautés. Précédemment dans ce chapitre (analyse des réseaux de régulation des gènes de la souche 2Xc1), les sept FTs exprimés de façon différentielle étaient également séparés en quatre communautés (Figure 54, b). Cette répartition des sept FTs est retrouvée au sein du réseau de co-expression des FTs de la souche 2Xc1. Les deux FTs NF-YB_2 et MYB-2R_14, qui dirigeaient la communauté rouge liée à la quantité de lipides de réserve, se retrouvent également liés l'un à l'autre dans la communauté rouge du réseau de co-expression des FTs (Figure 54). De manière intéressante, deux autres FTs dont le profil d'expression est lié à la quantité de lipides de réserve (LIM_1 ($R^2 = 0.76$) et MYB-3R ($R^2 = 0.79$)) appartiennent également à cette communauté. Ces deux FTs sont liés à la fois entre eux et avec le FT MYB-2R_14. La co-expression de ces trois FTs liés à la quantité de lipides de réserve n'est retrouvée que dans le réseau de la souche 2Xc1 (Figure 55), suggérant donc un rôle commun de chacun d'eux dans la réponse de la souche 2Xc1 à un stress azoté. De plus, le FT MYB-3R est un BN du réseau de la souche mutante (Tableau 7). Cette caractéristique topologique renforce donc son influence sur la régulation de la réponse de la souche 2Xc1. Une relation similaire est observée pour le FT MYB-rel_12 et les trois FTs dirigeants la communauté verte du réseau de régulation des gènes de la souche 2Xc1 qui semble liée au recyclage de l'azote et à la quantité de carbohydrates (MYB-2R_20, MYB-rel_11 et GATA_2 Figure 54 b). Ces trois derniers sont retrouvés au sein de la communauté verte du réseau de co-expression des FTs de la souche 2Xc1 et y sont co-exprimés à la fois entre eux et avec le FT MYB-rel_12 (Figure 55), lequel présente un profil d'expression corrélé à la quantité de carbohydrates ($R^2 = 0.77$). Cette co-expression spécifique à la souche 2Xc1 suggère également un rôle commun de ces quatre FTs dans la réponse de la souche 2Xc1 à un stress azoté. Le FT GATA_2, outre son implication potentielle dans le recyclage de l'azote et le fait qu'il ne soit exprimé que chez la souche mutante, occupe également une position de « hub » du réseau de co-expression des FTs de la souche 2Xc1. Cette position clé fait de GATA_2 un FT important dans la régulation de la réponse de la souche 2Xc1. Trois autres FTs semblent importants dans la régulation de la réponse de la souche mutante à un stress azoté. Le FT HB-other_11_PAS est un BN du réseau de co-expression des FTs chez les deux souches, suggérant ainsi une conservation de son rôle dans l'orchestration de la réponse à un stress azoté. Toutefois, son impact est d'autant plus important chez la souche 2Xc1 qu'il est à la fois le « hub » et le BN majeur de ce réseau de co-expression des FTs (Tableau 7).

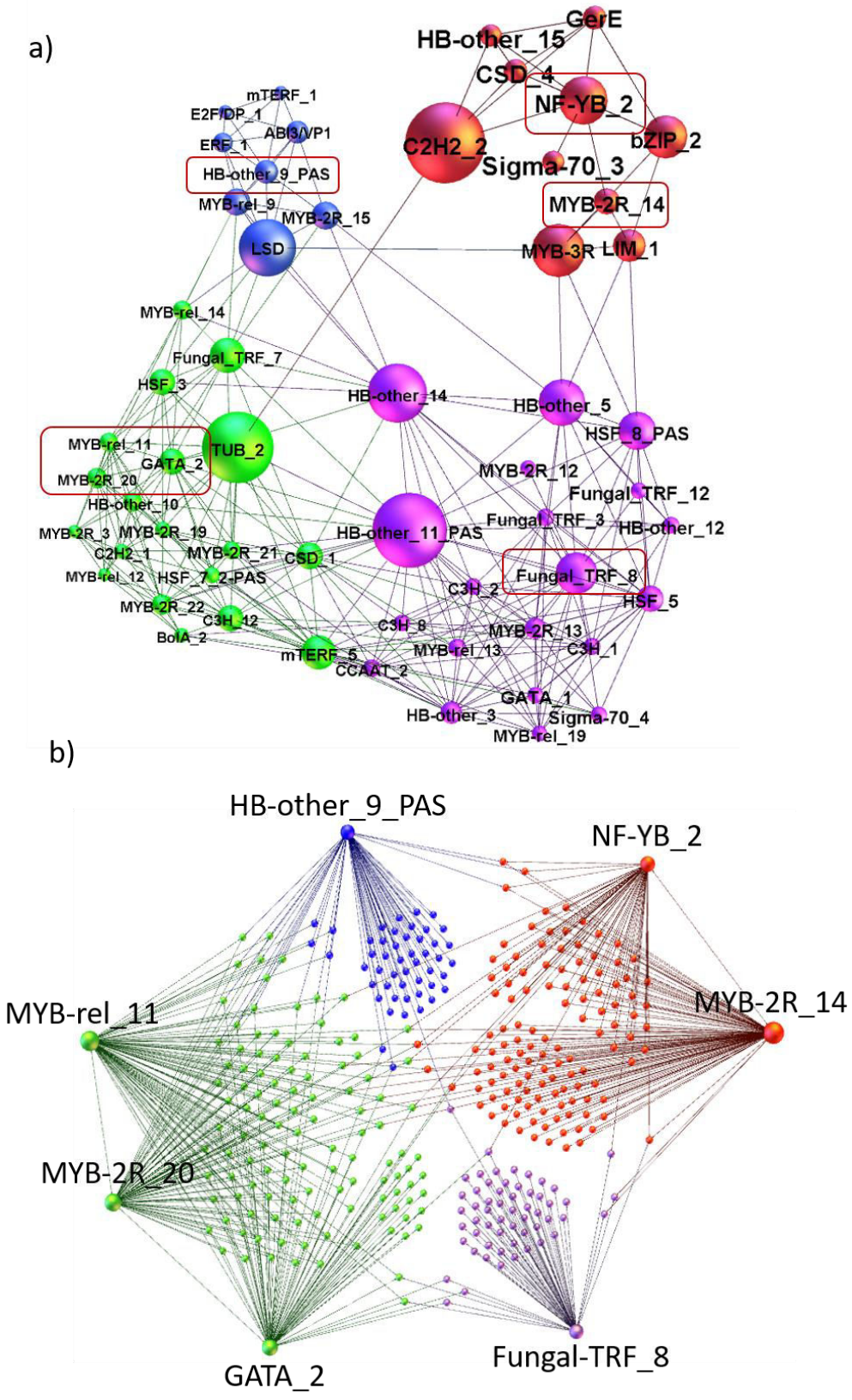

Figure 54 : représentation du réseau de co-expression des FTs de la souche 2Xc1 (a) et du réseau de régulation des gènes exprimés de façon différentielle chez cette même souche (présenté dans la partie 2 de ce chapitre) (b). La structure globale des deux réseaux est conservée avec quatre communautés. Les FTs exprimés de façon différentielle sont encadrés en rouge dans le réseau de co-expression des FTs (a). La répartition des FTs du réseau de régulation des gènes (b) est retrouvée au sein du réseau de co-expression des FTs (a).

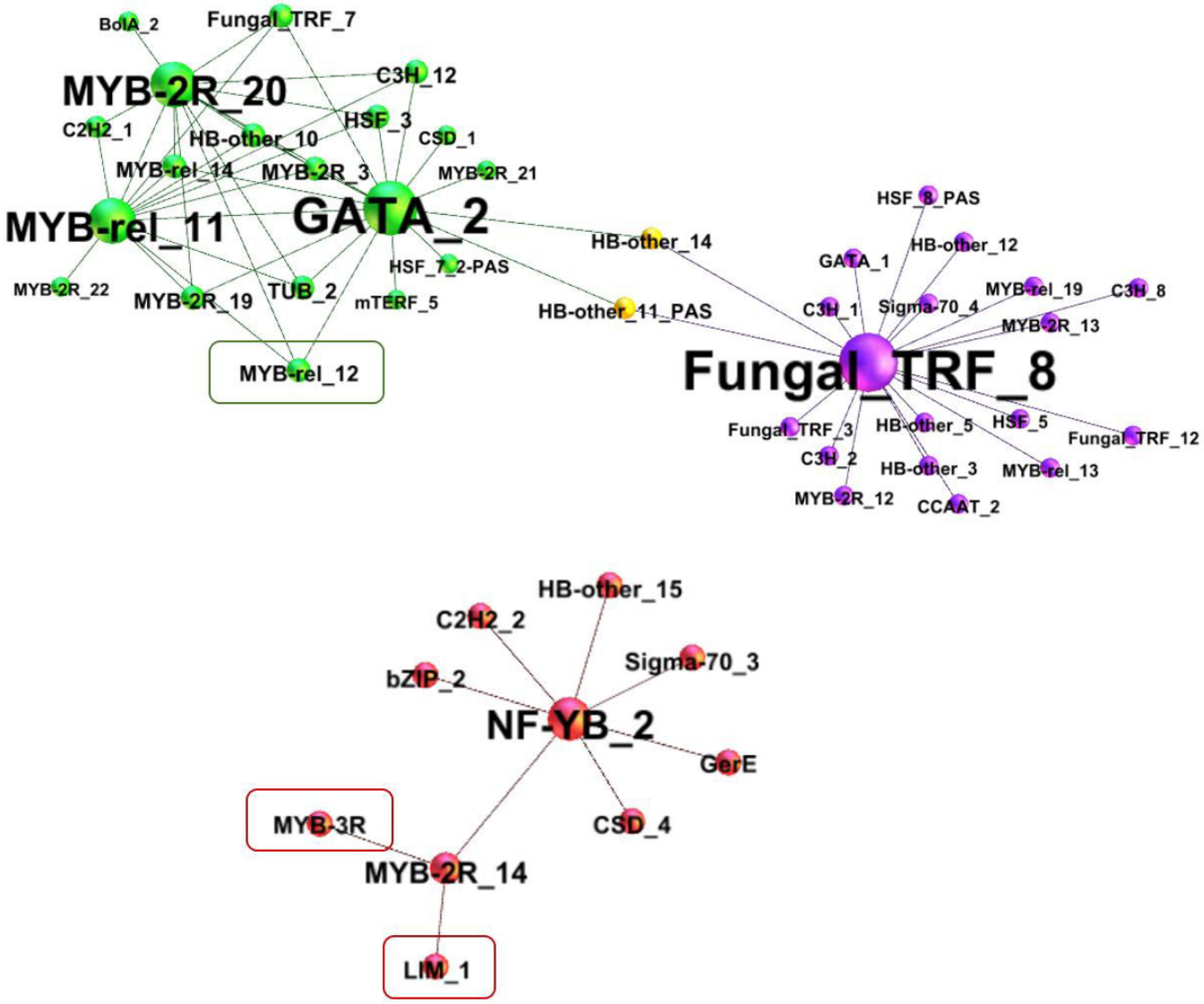

Figure 55 : représentation réduite du réseau de co-expression des FTs de la souche mutante présenté en figure 51 a). Seuls les FTs directement liés aux FTs MYB-rel_11, MYB-2R_20, GATA_2, Fungal-TRF_8, NF-YB_2 et MYB-2R_14, différentiellement exprimés chez la souche 2Xc1, sont représentés. Trois spécificités de la souche mutante sont mises en évidence : la co-expression du FT MYB-rel_11 (liée à la quantité de carbohydrates et encadré en vert) avec les trois FTs MYB-rel_11, MYB-2R_20 et GATA_2; la position de goulot d'étranglement des FTs HB-other_14 et HB-other_11_PAS entre les communautés vertes et violettes (nœuds colorés en jaune); la co-expression des FTs LIM_1 et MYB-3R (liés à la quantité de lipides de réserve et encadrés en rouge) avec le FT MYB-2R_14.

De plus, il constitue un BN entre le FT Fungal-TRF_8 et les trois FTs de la communauté verte GATA_2, MYB-2R_20 et MYB-rel_11 (Figure 55), qui sont exprimés de façon différentielle chez la souche 2Xc1 et constituent des candidats de la réponse spécifique de cette souche. Le FT HB-other_11_PAS partage ce rôle avec le FT HB-other_14, également « hub » et BN chez la souche 2Xc1. Les deux FTs HB-other_11_PAS et HB-other_14 jouent donc un rôle clé dans la régulation des flux d'information au sein du réseau de co-expression des FTs de la souche 2Xc1. Du fait de ce rôle clé, ils semblent donc primordiaux dans la régulation de la réponse de la souche mutante à un stress azoté. Ce rôle est renforcé par leur implication dans la régulation des flux d'information entre les quatre FTs liés à la quantité de carbohydrates chez la souche mutante (GATA_2, MYB-rel_11, MYB-2R_20 et MYB-rel_12) et le FT Fungal-TRF_8, exprimé de façon différentielle chez cette souche. Ce dernier est également un cas intéressant puisque, bien qu'étant exprimé de façon différentielle chez la souche 2Xc1, il occupe une position de « hub » et de BN dans le réseau de co-expression des FTs des deux souches (Tableau 7). Ces deux caractéristiques suggèrent à la fois une implication dans la réponse spécifique de la souche 2Xc1 à un stress azoté, ainsi qu'une conservation de son rôle clé dans la régulation de la réponse des deux souches à un tel stress. De plus, dans la partie précédente de ce chapitre, les fonctions enrichies parmi ses gènes cibles potentiels étaient les mêmes pour les deux souches. Ce FT semble donc réguler des fonctions communes à la réponse des deux souches de *T. lutea* à un stress azoté et occupe une position stratégique au sein du réseau de la régulation de cette réponse chez les deux souches.

Cependant, même si les caractéristiques topologiques de certains FTs sont conservées d'un réseau de co-expression des FTs à l'autre, les deux réseaux sont très différents. Pourtant, les deux souches appartiennent à la même espèce. Une telle différence peut avoir plusieurs explications, parmi lesquelles le fait que trois des BN du réseau de la souche WTc1 soient exprimés de façon différentielle chez la souche mutante. Ces gènes ont une fonction clé dans la régulation des flux au sein du réseau et ont plus tendance que les autres gènes à occuper une fonction clé (Yu *et al.*, 2007 ; McDermott *et al.*, 2009). Leur expression différentielle impacte donc fortement la structure du réseau et, par extension, la réponse de l'organisme orchestrée par ce réseau. De plus, de nombreuses différences sont notables d'un point de vue physiologique et structurel. Outre les quantités de lipides de stockage et de carbohydrates, des différences de profil pigmentaire, un large remaniement du protéome ou encore une différence de la taille des cellules ont été recensées entre les deux souches (Garnier, 2016). Ces différences impliquent une grande diversité de

mécanismes dont la mise en place implique des régulations variées. Une telle variété peut, en partie, expliquer ces différences du programme de régulation de la réponse de la souche 2Xc1 par rapport à la souche WTc1. Il faut également prendre en compte que ces réseaux ne reflètent que la réponse de l'organisme au niveau transcriptionnel. Bien que la transcription soit le premier niveau de l'expression des gènes et indispensable à la production de protéines fonctionnelles, ce sont bien ces dernières qui assurent la réponse de l'organisme à un stress azoté. Or, d'autres niveaux de régulation entrent en jeu. Des régulations post-transcriptionnelles, post-traductionnelles, de maturation des protéines ou de leur activité (Maier et al., 2009 ; Vogel & Marcotte, 2012) ont également un rôle dans la régulation de la réponse de l'algue au stress azoté.

Ces hypothèses ne sont, bien sûr, pas les seules à même d'expliquer les différences entre les deux réseaux de co-expression des FTs. La vérité se cache très probablement derrière l'implication conjointe de plusieurs d'entre elles.

Certains de ces mécanismes, parmi d'autres, sont à l'origine de l'expression différentielle des FTs chez tous les organismes, dans le but de réguler plus finement l'expression des gènes et la réponse des organismes à divers stimuli (cf introduction). Toutefois, les sept FTs identifiés ici (Fungal-TRF_8, HB-othr_9_PAS, NF-YB_2, MYB-2R_14, GATA_2, MYB-2R_20 et MYB-rel_11) présentent une expression différentielle entre deux souches d'une même espèce. Du fait que la souche mutante 2Xc1 résulte d'un procédé de domestication faisant notamment intervenir la mutagénèse UVc, il apparaîtrait donc que les nombreuses différences observées entre ces deux souches soient majoritairement dues aux mutations conséquentes à l'exposition aux UVc. Sans être exhaustif, plusieurs hypothèses, en lien à cette exposition, sont à même d'expliquer l'expression différentielle de ces différents FTs. (i) Des mutations ont pu être générées dans les régions promotrices des gènes codant ces FTs. Bien que l'affinité d'un FT pour son site de fixation permette une variation de la séquence ciblée, une mutation en son sein peut fortement réduire l'action d'un FT sur son gène cible (Morley *et al.*, 2004 ; Wang *et al.*, 2005 ; Pai *et al.*, 2015). (ii) Bien que n'étant pas exprimés de façon différentielle, les FTs régulant l'expression de ces régulateurs candidats peuvent faire l'objet d'autres types de régulations. Des modifications post-transcriptionnelles peuvent, en effet, intervenir telle que, par exemple, l'implication des microARN (Arora *et al.*, 2013). Une régulation de l'activation des FTs peut également être impliquée via des mécanismes impliquant des modifications post-traductionnelles telle que la

phosphorylation (Gao & Stock, 2015). Enfin, (iii) des modifications affectant les co-facteurs impliqués dans la régulation de l'expression des FTs candidats peut aboutir à leur expression différentielle. Ces modifications peuvent agir sur l'expression de ces co-facteurs autant que leur structure protéique ou leur activité enzymatique quand celle-ci est requise. Là encore, plusieurs mécanismes combinent très probablement leurs effets dans l'expression différentielle de ces gènes.

V. Bilan et perspectives

Les travaux présentés dans ce chapitre ont donc permis l'identification de sept FTs exprimés de façon différentielle lors de la réponse à un stress azoté chez la souche 2Xc1 par rapport à la souche sauvage WTc1. Une stratégie couplant l'analyse de réseaux de co-expression des gènes et de régulation des gènes a permis de pallier aux problèmes d'annotation fonctionnelle rencontrés lors de l'étude d'organismes non-modèles. Des mécanismes impliqués dans la réponse spécifique de la souche 2Xc1 de *T. lutea* à un stress azoté ont ainsi pu être mis en évidence. Deux FTs (NF-YB_2 et MYB-2R_14) potentiellement liés à la photosynthèse et à la réponse au stress oxydatif qu'elle induit via la production de TAG. Trois FTs (GATA_2, MYB-2R_20 et MYB-rel_11) semblent impliqués dans le recyclage de l'azote et du carbone afin de nourrir la synthèse de lipides de réserve et de carbohydrates. Plus précisément, le FT MYB-rel_11 semble jouer un rôle important dans la réponse spécifique de la souche mutante, via la régulation de l'expression des gènes de la PLAAOx, de la CSAP et du transporteur de nitrate à haute affinité *Nrt*2.1. Enfin, l'analyse du réseau de co-expression des FTs de chaque souche a permis d'appréhender un niveau supérieur de la régulation de la réponse de la souche mutante face à un stress azoté. Bien que n'étant pas exprimés de façon différentielle, certains FTs pourraient être impliqués dans cette réponse de par leur lien avec la physiologie de l'algue et les candidats exprimés de façon différentielle, ou via leur position topologique clé au sein du réseau de co-expression des FTs. Ainsi, deux FTs (LIM_1 et MYB-3R) liés à la quantité de lipides de réserve ont pu être associés aux FTs NF-YB_2 et MYB-2R_14, et un FT (MYB-rel_12) lié à la quantité de carbohydrates a pu être associé aux trois FTs GATA_2, MYB-2R_20 et MYB-rel_11. De plus, deux FTs (HB-other_11_PAS et HB-other_14) constituant des goulots d'étranglement du réseau de co-expression des FTs de la souche mutante

jouent un rôle déterminant dans la gestion des flux d'informations au sein du réseau et notamment entre les FTs exprimés de façon différentielle chez la souche 2Xc1.

Autant de pistes intéressantes pour comprendre la régulation de la réponse de la souche mutante de *T. lutea* à un stress azoté au niveau de l'expression des gènes. Cette étude est la première de ce type chez une microalgue non-modèle. Malgré les problèmes inhérents à ces organismes (le génome de *T. lutea* et son annotation son encore récents et les données RNA-seq à notre disposition ne concernent que 6 conditions physiologiques par souche), les résultats obtenus sont concluants et prometteurs. Dans les mois et les années à venir le génome sera re-séquencé en utilisant la technologie PacBio (Rhoads & Au, 2015) permettant d'améliorer sa qualité et davantage de données RNA-seq seront générées. Utiliser des données RNA-seq représentant plus de conditions expérimentales permettrait d'augmenter la finesse de la détection des modules par WGCNA. Cette étape étant très importante dans la stratégie développée dans ce chapitre, le nombre de conditions expérimentales est un paramètre crucial. Ces données permettront de profiter du potentiel de la stratégie présentée dans ce chapitre.

Les candidats identifiés dans cette étude demandent à être confirmés d'un point de vue fonctionnel. Tout d'abord, le lien entre les FTs NF-YB_2 et MYB-2R_14 et leurs gènes cibles potentiels doit être confirmé par des techniques telles que le retard de migration sur gel, l'immunoprécipitation de la chromatine (ChIp-seq) ou encore la technique de levure ou bactérie simple hybride. La même démarche doit être entreprise pour les FTs GATA_2, MYB-2R_20 et MYB-rel_11, notamment vis-à-vis des gènes de la PLAAOx, de la CSAP et du *Nrt*2.1. La maîtrise de la transformation génétique de *T. lutea* serait un outil puissant dans cette otique pour valider définitivement l'implication des FTs candidats dans l'établissement du phénotype mutant. En effet, les deux FTs GATA_2 et MYB-2R_14 n'étant exprimés que chez la souche 2Xc1, réprimer leur expression (par génie génétique ou par RNAi) permettrait d'évaluer leur impact sur le phénotype mutant. Utiliser la même approche afin de cibler les FTs HB-other_11_PAS et HB-other_14, constituant des goulots d'étranglement clés au sein du réseau de co-expression des FTs de la souche 2Xc1, permettrait également d'évaluer leur impact sur la régulation de la réponse de cette souche vis à vis d'un stress azoté. Dans l'optique de confirmer et caractériser plus largement l'implication des FTs candidats dans la réponse de la souche mutante à un stress azoté, une approche de ChIp-seq ciblant un de ces FT dans différentes conditions physiologiques consécutives

à une modification de la disponibilité en azote serait intéressante. Une telle approche permettrait ainsi d'identifier l'ensemble des voies métaboliques affectées par le FT étudié au cours de la réponse de l'algue.

VI. Matériels et methodes

Conditions de culture et traitements

La souche sauvage de *Tisochrysis lutea* CCAP 927/14 (WTc1) et la souche mutante précédemment décrite (Bougaran *et al.*, 2012) ont été cultivée pendant 85 jours en chémostats, à un taux d dilution de 0,5 j-1. Les cultures ont été réalisées dans un milieu de culture de Walne modifié contenant un ratio de N : P de 125/125µM dans un photobioréacteur (1000 x 400 x 250 mm) soumis à une lumière continue (150 µmol.m-3.s-1) et maintenu à 27°C et pH 7,3. Le taux de dilution a été régulièrement vérifié par pesée du milieu de culture sortant. Trois injections de 3,5 µmoles de NaNO3 dans les 10 L de culture ont été réalisées aux 20ème, 43ème et 83ème jours.

Une fois que la culture limitée par l'azote a atteint un état d'équilibre, caractérisé par des paramètres physiologiques constants (Figure 44), l'injection de NaNO3 a été réalisée (Garnier *et al.*, acceptée). Cette condition de réplétion azotée provoque une augmentation de la concentration cellulaire et du carbone particulaire. Parallèlement, le rapport N/C augmente puisque les microalgues absorbent l'azote injecté (Figure 44). Ensuite, la concentration cellulaire, le carbone particulaire et le ratio N/C se maintiennent à un niveau élevé. Le manque d'azote disponible dans le milieu de culture induit, par la suite, une diminution de ces paramètres physiologiques, caractérisant une condition de déplétion azotée (Figure 44). Enfin, la culture atteint un nouvel état d'équilibre, dû au taux de dilution du chémostat (Garnier *et al.*, acceptée).

Construction des librairies RNA-seq et analyse des données de séquençage

L'étude transcriptomique par RNA-seq est fondée sur six échantillons biologiques par souche, provenant de cette expérience (Figure 46). Ces douze échantillons ont fait l'objet d'un séquençage des ARN par la technologie Illumina.

Pour chaque échantillon, les ARN totaux ont été extraits en utilisant le TRIZOL (Invitrogen, USA) selon les instructions du fabricant. Un traitement à la DNase (DNase RQ1, Promega) a été réalisé afin d'éliminer d'éventuelles traces d'ADN génomique. La qualité des ARN purifiés a été déterminée par mesure de l'absorbance (260 nm/280 nm) au nanodrop ND-1000 (LabTech, USA). Les ARN messagers polyadénylés ont été isolés grâce à des billes magnétiques (MicroPoly(A)PuristTM kit, Ambion) selon les instructions du fabricant. Les AND complémentaires (ADNc) ont été synthétisés (SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen, USA)) selon les instructions du fabricant. Les 12 librairies ont été fabriquées et séquencées avec un séquenceur Illumina HiSeq 2000 (Illumina Corporation Inc.). Environ 4-5 ng d'ADNc ont été utilisés pour la fabrication des librairies, prise en charge par la plateforme génomique Biogenouest. Le séquençage a été réalisé en paired-end, chaque read étant composé de 100 paires de bases.

Pour chaque échantillon, les reads séquencés ont été filtrés grâce au logiciel Cutadapt (version 1.0) pour éliminer les séquences des adaptateurs Illumina. Reads quality filter (version 1.0.0) a ensuite été utilisé afin d'exclure les reads de faible qualité selon un seuil de qualité de 30 et une longueur minimal de read de 75 bases. La qualité des reads a été évaluée par le logiciel FastQC développé par S. Andrews à l'institut Babraham (www.bioinformatics.bbsrc.ac.uk). Les 352 199 068 paires de reads nettoyées ont été alignées sur le génome de *T. lutea* (données brutes accessibles sur SRA, RUN : SRR3156597) grâce au logiciel Tophat2 (version 0.5) (Trapnell *et al.*, 2009). Les reads alignés sur chaque genes ont été comptés grâce à htseq-count (version 0.3.1) en mode union. Le niveau d'expression des gènes a ensuite été calculé en reads par kilobases par millons de reads alignés (RPKM). Un gène ayant une valeur de RPKM > 1 dans au moins un des 12 échantillons a été considéré comme étant exprimé. Parmi les 20 582 gènes identifiés dans le génome de *T. lutea*, 15 333 étaient exprimé dans cette étude. Les gènes ont été annotés par BLAST contre la base de données swissprot et les domaines fonctionnels ont été identifiés par InterProScan selon les méthodes implémentées dans BLAST2GO (Conesa *et al.*, 2005). Les facteurs de transcription ont été annotés grâce au pipeline mis au point dans le chapitre 1 (Thiriet-Rupert *et al.*, 2016).

Construction et analyse du réseau de co-expression des gènes

A partir des 15 333 gènes exprimés, un réseau de co-expression des gènes non signé a été construit grâce à la méthode implémentée dans le package R WGCNA (Langfelder & Horvath, 2008). Un seuil β (soft threshold power) de 16 a été utilize pour remplir le critère de topologie libre d'échelle (scale-free topology) requise pour une clusterisation optimale. La matrice similarité obtenue a ensuite été transformée en une matrice d'adjacence, puis en matrice de recouvrement topologique (topological overlap matrix, TOM). Cette mesure de dissimilarité fondée sur la TOM a été couplée à une clusterisation hiérarchique grâce à l'algorithme Dynamic Tree Cut, permettant l'identification des modules de gènes co-exprimés. Dans le but de former des modules plus cohérents, les modules similaires ont été fusionnés en se fondant sur la première composante principale de chacun d'eux (l'eigengene) selon un seuil de 0,25. Cet eigengene représente le profil d'expression d'un module donné. Pour identifier les modules significativement associés aux paramètres physiologiques, l'eigengene de chaque module a été corrélé avec la dynamique des quantités cellulaires de lipides de réserve et de carbohydrates. Les gènes hub ont été identifiés dans chacun des trois modules d'intérêt par le calcul de la connectivité intramodulaire (kME). La corrélation du profil d'expression de chaque gène avec les paramètres physiologiques a également été calculée.

Concernant la priorisation de gènes, un gène a été priorisé si le coefficient de la corrélation de son profil d'expression avec la dynamique des quantités cellulaires de lipides de réserve ou de carbohydrates était supérieur ou égale à 0,75. Les gènes dont la valeur de kME était supérieure ou égale à 0,95 ont été considérés comme des gènes hub et ont également été priorisés.

Identification des gènes exprimés de façon différentielle et construction du réseau de régulation des gènes

Les gènes exprimés de manière différentielle entre les deux souches ont été identifiés entre couples d'échantillons correspondant à un même temps d'échantillonnage (par exemple : l'échantillon Pt-27 de la souche WTc1 Vs l'échantillon Pt-27 de la souche 2Xc1). L'expression différentielle a été mesurée grâce au logiciel Gfold V1.1.2, en utilisant la valeur GFOLD ayant

plus de sens d'un point de vue biologique (Feng *et al.*, 2012). Un gène pour lequel la valeur absolue de GFOLD était supérieure à 2 a été considéré comme exprimé de manière différentielle.

Un réseau de régulation des gènes a ensuite été construit pour chaque souche. Le premier à partir des valeurs de RPKM dans les six échantillons WTc1 des 527 gènes exprimés de manière différentielle. Le deuxième, à partir des valeurs de RPKM des mêmes gènes dans les six échantillons 2Xc1. Chaque réseau a été construit grâce u logiciel eLSA (extended local similarity analysis) (Xia *et al.*, 2011, 2013). Ces réseaux dirigés sont fondés sur les corrélations du profil d'expression d'un FT avec celui de ses gènes cibles potentiels. Une telle corrélation a été considérée significative lorsque le coefficient de Spearman était supérieur à 0,8 et la p-value associée inférieure à 0,05. Les réseaux ainsi produits ont été visualisés et analysés grâce au logiciel Gephi (Bastian *et al.*, 2009).

Extraction des ARN totaux et transcription inverse

Les échantillons ont été centrifugés (20 min, 5000 g, 4°C). Le surnageant et l'eau de mer ont été éliminés afin d'enlever le sel. Les culots ont ensuite été re-suspends dans du Trizol (Invitrogen, Carlsbad, CA, USA) et du chlorophorm. Après centrifugation, la phase supérieure a été collectée et 0,5 volume d'éthanol absolu a été ajouté. Les échantillons ont été transférés sur une colonne du mini kit RNeasy Plant (Qiagen, Helden, Germany) puis traités selon les instructions du fabricant. Un traitement a la DNase (RQ1 DNase, Promega, Madison, WI, USA) a été appliqué et les ARN ont été purifiés en utilisant le mini kit RNeasy Plant avec le tampon RLT et l'éthanol. La qualité et la concentration ont été déterminées par au nanodrop (ND-1000; NanoDrop Technologies, Wilmington, DE) aux longueurs d'ondes de 260 et 280 nm. L'amplification PCR des ARN extraits a servi de contrôle afin de vérifier l'absence d'ADN génomique. Les ARN totaux ont été stockés à -80°C.

La transcription inverse a été réalisée en utilisant le kit de transcription inverse High Capacity cDNA Reverse transcription kit (Applied Technologies, Foster, CA, USA) selon les instructions du fabricant. La q-RT-PCR a été réalisée en utilisant la technologie Fluidigm Biomark par la plateforme du Genotoul (http://get.genotoul.fr/). La moyenne géométrique de la normalisation par deux gènes de ménage, EF1 (translation elongation factor 1 alpha) et GAPDH

(Glyceraldehyde-3-Phosphate Dehydrogenase), a été utilisée pour calculer l'expression différentielle relative par la formule du 2^{-ddCt}. L'expression différentielle des gènes codant les FTs *MYB-2R_20*, *MYB-rel_11* et les gènes codant la *PLAAOx*, la *CSAP* and le *Nrt*2.1 a été mesurée au fil de la cinétique de la deuxième injection d'azote. L'état d'équilibre précédent l'injection a été utilisé comme condition de référence. Huit gènes représentant différents niveaux d'expression ont été sélectionnés afin d'évaluer la qualité des données de RNA-seq. La corrélation obtenue entre les données de RNA-seq et celles de q-RT-PCR est de 0,82. Cela indique que les données RNA-seq sont bien représentatives du profil transcriptomique de nos échantillons. Les amorces utilisées sont disponibles en annexe E.

Conclusions générales et perspectives

I. De l'identification des FTs dans le génome de microalgues à l'élucidation de leur histoire évolutive

Cette thèse a débuté par l'élaboration d'un pipeline d'identification et de classification des FTs adapté aux organismes photosynthétiques, et particulièrement aux microalgues. Sa relative exhaustivité permet d'identifier des FTs appartenant à des familles jusqu'ici très peu étudiées chez les microalgues, puisque celles-ci ne sont habituellement pas (ou très rarement) recherchées dans leur génome. Lorsque les FTs sont utilisés pour étudier l'histoire évolutive d'un organisme ou d'un groupe d'organismes, répertorier les familles de FTs présentes dans leur génome avec la plus grande exhaustivité possible est une première étape cruciale. Ces études comparatives sont, en effet, en grande partie fondée sur la présence et l'absence de familles de FTs entre les différents organismes étudiés. Il est donc clair que fonder une telle étude sur une identification incomplète des FTs ne peut qu'aboutir à des résultats eux-mêmes incomplets, voire parfois trompeurs.

L'utilisation de ce pipeline dans une étude comparant quatre lignées de microalgues (haptophytes, straménopiles, rhodophytes et chlorophytes) nous a permis de mettre en évidence des spécificités de lignées. Certaines d'entre elles avaient précédemment été identifiées, telle la présence des auréochromes couplant une fonction de facteur de transcription et de récepteur de la lumière bleue chez les straménopiles photosynthétiques (Ishikawa *et al.*, 2009). D'autres, ont pu être confirmées ou découvertes, telles l'absence des familles GATA, MADS-box et LIM chez les straménopiles (Rayko *et al.*, 2010). D'autres, enfin, ont pu être infirmées et affinées, telles les familles de FTs décrites comme spécifiques de la lignée verte (Lang *et al.*, 2010 ; Sharma *et al.*, 2013). La spécificité de ces familles à la lignée verte n'était fondée que sur leur absence chez deux espèces de microalgues rouges : *Galdieria sulfuraria* et *Cyanidioschyzon merolae*. Or, ces deux espèces sont adaptées à un milieu de vie extrême, les sources chaudes du parc Yellowstone (Schönknecht *et al.*, 2013). Par conséquent, ces deux espèces ne sont pas représentatives de la lignée rouge. Fonder une absence spécifique de cette lignée à partir de seulement deux espèces extrêmophiles introduit donc un biais. L'utilisation dans notre étude de plusieurs lignées de microalgues ainsi que de *Porphyridium purpureum*, une microalgue rouge mésophile, a permis d'affiner les spécificités de lignée établies précédemment. Ces résultats étant dus à la fois à l'exhaustivité du pipeline et à la couverture des lignées étudiées, ce dernier critère apparait également crucial. Utiliser un grand nombre d'espèces permet, en effet, de mettre en évidence des spécificités beaucoup plus précises,

lesquelles aboutissent à des conclusions permettant de mieux comprendre l'histoire évolutive des organismes étudiés. L'ajout du génome de *C. tobin* aux données de la publication (Thiriet-Rupert et al 2016) illustre parfaitement ce propos. L'analyse des FTs de la famille des bHLH chez les haptophytes qui en découle révèle une distribution particulièrement intéressante. Les données génomiques et transcriptomiques disponibles à ce jour semblent aller dans le sens du gain de la famille des bHLH chez les Phaeocystale et les Prymnesiales. Ces deux ordres sont, en effet, les seuls dont les membres comportent au moins un FT de cette famille (Figure 41).

Il serait intéressant de poursuivre cette étude comparative (Thiriet-Rupert et al 2016) en y incluant les FTs identifiés dans le génome de microalgues appartenant à d'autres lignées (Tableau 1). Augmenter le nombre de lignées représentées ainsi que le nombre d'individus représentant chaque lignée permettrait de mettre en lumière des points communs et des spécificités de lignées penchant en faveur d'une des hypothèses établies concernant leur histoire évolutive. Quant à la répartition des bHLH chez les haptophytes, la taille de l'échantillonnage est, là encore, cruciale. En effet, outre les Isochrysidales dont quatre genres sont représentés parmi les données transcriptomiques disponibles, les Coccolithales, Prymnesiales et Phaeocystales n'en comptent que deux, et les Pavlovales qu'un seul (Tableau 3). De plus, aucune donnée n'est disponible concernant des espèces représentant les ordres des Zygodiscales et des Syracosphaerales. Analyser des données transcriptomiques (plus faciles à obtenir que le génome de l'organisme) provenant d'espèces représentant chaque ordre des haptophytes permettrait d'affiner cette répartition atypique des bHLH. Enfin, des données génomiques sont indispensables à la confirmation de l'absence de bHLH chez les Isochrysidales, Coccolithales et Pavlovales. Cette analyse participerait à clarifier l'histoire évolutive de cette lignée de microalgues, peu connue malgré leur importance écologique.

II. Identification de régulateurs de la réponse de *T. lutea* 2Xc1 à un stress azoté : comprendre la production de lipides de réserve dans la perspective de futures approches de bio-engineering

Dans un contexte physiologique particulier, l'étude des mécanismes de régulation de l'expression des gènes requiert un inventaire le plus précis possible des FTs présents dans le génome de l'organisme étudié. Grâce au pipeline élaboré dans la partie précédente, l'expression de l'ensemble des FTs des deux souches de *T. lutea* a pu être suivie au cours de différentes conditions expérimentales liées à une variation de la disponibilité en azote. Une stratégie couplant l'analyse de réseaux de co-expression et de régulation des gènes a permis d'identifier des régulateurs potentiels de la réponse spécifique de la souche 2Xc1, tout en palliant au manque d'annotation fonctionnelle inhérente aux organismes non-modèles. Plus particulièrement, deux FTs (NF-YB_2 et MYB-2R_14) semblent impliqués dans la régulation de mécanismes liés à la photosynthèse ainsi qu'à la réponse au stress oxydatif qu'elle provoque. Des mécanismes liés à la synthèse de lipides de réserve dans le but de prévenir ce stress oxydatif seraient également régulés par ces deux FTs. Trois autres FTs (GATA_2, MYB-2R_20 et MYB-rel_11) semblent impliqués dans l'absorption de l'azote, la protéolyse, le recyclage de l'azote des acides aminés libres, et la synthèse de carbohydrates sous forme de chrysolaminarine. Une analyse q-RT-PCR a précisé le rôle potentiel des deux FTs MYB-2R_20 et MYB-rel_11 dans la réponse spécifique de la souche mutante. Ils sont tous deux impliqués dans l'absorption de l'azote via la régulation de l'expression du gène codant pour le transporteur de nitrate Nrt2.1, ainsi que le recyclage de l'azote et du carbone produit par protéolyse via la régulation de l'expression de la PLAAOx et de la CSAP. La fonction de ces deux protéines demande cependant à être précisée. De plus, alors que le FT MYB-2R_20 semble réguler l'expression de ces trois gènes chez les deux souches, le FT MYB-rel_11 ne semble les réguler que chez la souche 2Xc1. Ce dernier apparait donc comme un candidat sérieux de la réponse spécifique de la souche 2Xc1 à un stress azoté via son implication dans les mécanismes clés que sont le recyclage de l'azote et du carbone ainsi que l'absorption d'azote. Enfin, des régulateurs potentiels de ces FTs candidats ont pu être identifiés par l'analyse de réseaux de co-expression de FTs.

Toutefois, ces régulateurs de la réponse de la souche mutante à un stress azoté ont été identifiés à partir de données transcriptomiques. Bien que la transcription soit indispensable à l'expression des gènes, la compréhension de sa régulation n'est pas suffisante à expliquer l'expression différentielle des gènes dans son ensemble. D'autres niveaux de régulation interviennent entre la synthèse de l'ARNm et la modification du métabolisme. L'étude de ces mécanismes permettrait la construction d'une carte plus complète de la réponse de *T. lutea* à un stress azoté. Principalement, peu de correspondances sont retrouvées entre le niveau de transcription des gènes et la quantité des protéines correspondantes (Garnier, 2016). De faibles corrélations entre ces deux niveaux de l'expression des gènes ont également été observées chez les microalgues straménopiles *Aureococcus anophagefferens* et *Thalassiosira pseudonana* (Wurch *et al.*, 2011 ; Dyhrman *et al.*, 2012). Identifier les mécanismes à même d'expliquer cette différence aboutirait à une meilleure compréhension de la régulation de la réponse de *T. lutea*. Dans cette optique, l'étude des ARN régulateurs constituerait une première étape. Parmi ces ARN régulateurs, les micro ARN (miRNA) et les petits ARN interférant (siRNA pour short interfering RNA) sont les mieux caractérisés (Eggleston, 2009 ; Ghildiyal & Zamore, 2009 ; Molnar *et al.*, 2011). Ces molécules jouent un rôle régulateur important via la dégradation des ARNm ciblés ou par l'inhibition de leur traduction (Carthew & Sontheimer, 2009 ; Voinnet, 2009). A ce jour, peu d'études sur les ARN régulateurs ont été menées chez les microalgues (Molnár *et al.*, 2007 ; Rogato *et al.*, 2014 ; Li *et al.*, 2014). Réaliser le séquençage haut débit des ARN régulateurs chez *T. lutea* permettrait, suite à la prédiction de leurs cibles, de compléter le réseau de régulation de la réponse face à un stress azoté.

De plus, une validation fonctionnelle de nos FTs candidats est nécessaire. La validation de leurs gènes cibles putatifs peut être entreprise via des techniques de retard de migration sur gel ou encore de bactérie ou levure simple hybride, validant ainsi les fonctions auxquelles ces FTs sont associés. Cela permettrait également d'identifier une séquence consensus de leur site de fixation qui pourrait, par la suite, être utilisée dans des approches *in silico*. En effet, grâce au re-séquençage du génome, les séquences promotrices et intergéniques, partiellement séquencées jusqu'ici, seront exploitables. Rechercher ainsi l'emplacement des sites de fixation potentiels des FTs candidats permettrait d'affiner le réseau de régulation des deux souches de *T. lutea*. Coupler ces informations aux données d'expression de gènes déjà disponibles, caractériserait plus largement la fonction de ces FTs à l'échelle du génome.

Une approche de transformation génétique ou de RNAi ciblant l'un de ces FTs afin de réprimer son expression peut également être envisagée. Evaluer l'impact de cette répression sur le phénotype de la souche 2Xc1 ainsi que sur l'expression des gènes cibles potentielles du FT ciblé permettrait de valider l'implication des FTs candidats dans l'établissement du phénotype mutant suite au stress azoté. Malheureusement, des approches de génie génétique (transformation génétique, utilisation de RNAi …) n'ont, à ce jour, pas encore été mises au point chez *T. lutea*. Toutefois, dans l'optique de l'utilisation d'approches de génie génétique chez cette espèce, l'identification de gènes tels la PLAAOx, la CSAP ou le Nrt2.1 présente un grand intérêt. En effet, leur expression dépendante de la disponibilité en azote est conférée par la structure de leur séquence promotrice. L'utilisation de tels promoteurs inductibles, peut constituer un outil très utile pour maitriser l'expression d'un transgène. Le re-séquençage du génome permettra d'identifier et analyser ces séquences promotrices et fournira du matériel à la future mise au point de la transformation génétique de *T. lutea*.

Une approche couplant l'analyse de données d'immunoprécipitation (Chip-seq) de la chromatine et de RNA-seq peut également être envisagée dans l'optique d'une caractérisation fonctionnelle de ces FTs. L'utilisation de Chip-seq permet d'identifier, à l'échelle du génome, les localisations génomiques auxquelles le FT étudié est lié à son site de fixation. L'analyse de ces données permettrait d'identifier les cibles potentielles de ce FT. De plus, identifier les cibles potentielles exprimées et, surtout, exprimées de façon différentielle entre les deux souches grâce à des données RNA-seq permettrait de caractériser le rôle de ce FT chez chacune des deux souches. Afin d'évaluer le niveau d'expression des gènes de manière plus fidèle à la réalité, un séquençage des ARN de manière brin spécifique peut être utilisé. Cette approche, récemment mise au point, permet de quantifier plus efficacement l'expression des gènes dont les loci respectif se chevauchent mais sont transcrits à partir de brins opposés (Zhao *et al.*, 2015). La possible implication de transcrits anti-sens peut également être évaluée. Classiquement, un ARNm est le produit de la transcription du brin anti-sens. Toutefois, un transcrit peut également être produit à partir du brin sens. Celui-ci aura une séquence complémentaire de l'ARNm et est appelé transcrit anti-sens. Ces transcrits pouvant impacter l'expression des gènes à différents niveaux, leur détection permet de mieux comprendre et analyser le niveau d'expression des gènes dans les conditions étudiées (Mills *et al.*, 2013). Cette approche, plus fidèle à la réalité biologique, devrait être de plus en plus utilisée à l'avenir, évitant une surestimation de l'expression des gènes étudiés.

L'identification de FTs candidats ainsi que l'estimation de l'expression de leurs gènes cibles n'en sera que plus efficace. Toutefois, une telle approche n'impliquerait que l'expression des gènes au niveau transcriptionnel. Coupler cette analyse à des données protéomiques et métabolomiques, lesquelles illustrent à la fois le produit final de l'expression des gènes et son impact sur le métabolisme, permettrait de caractériser de manière plus complète le rôle des FTs dans la régulation du phénotype. De plus, mener cette étude sur une population de cellules synchronisées permettrait de réduire les variations inter-individuelles de la réponse au stress appliqué. Chez de nombreux organismes, et particulièrement chez les microalgues qui sont des organismes photosynthétiques, les processus biologiques sont synchronisés par l'alternance de l'exposition à la lumière et à l'obscurité (Farinas *et al.*, 2006 ; Imoto *et al.*, 2011 ; Miyagishima *et al.*, 2012 ; Noordally & Millar, 2015 ; Suzuki *et al.*, 2016). En l'absence de synchronisation, chaque cellule suit son propre rythme, ce qui peut créer un bruit de fond lors de l'analyse des données. Synchroniser la population étudiée au préalable permet de réduire ce bruit de fond.

De telles approches intégratives sont, de nos jours, de plus en plus utilisées et permettent de mettre en évidence des mécanismes réellement impactant à l'échelle de la cellule. Mener cette analyse intégrative chez les deux souches et à différents moments de la cinétique de leur réponse à la disponibilité en azote aboutirait à une identification plus précise du rôle de ces FTs et des mécanismes impliqués dans la régulation de cette réponse. Elucider les mécanismes sous-jacents à la réponse spécifique de la souche 2Xc1 mènera à une meilleure compréhension de l'établissement de son phénotype, lequel présente un clair intérêt biotechnologique. Une pleine compréhension de ces processus permettrait d'identifier les mécanismes et cibles moléculaires stratégiques dans le but de manipuler et favoriser la production de lipides par bio-engineering.

Enfin, l'ensemble de la stratégie utilisée dans cette thèse peut être appliquée à l'identification de régulateurs de la production d'autres composés d'intérêt. Par exemple, *T. lutea* étant connue pour produire du DHA, les régulateurs de la production de ces acides gras polyinsaturés à longue chaine pourraient être recherchés. L'augmentation de la production de composés d'intérêt par les microalgues grâce à des approches de bio-engineering ciblant ces régulateurs pourrait ensuite être envisagée.

Bibliographie

Adams C, Godfrey V, Wahlen B, Seefeldt L, Bugbee B. 2013. Understanding precision nitrogen stress to optimize the growth and lipid content tradeoff in oleaginous green microalgae. *Bioresource Technology* **131**: 188–194.

Adl SM, Simpson AGB, Farmer MA, Andersen RA, Anderson OR, Barta JR, Bowser SS, Brugerolle G, Fensome RA, Fredericq S, *et al.* 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *The Journal of Eukaryotic Microbiology* **52**: 399–451.

Agalioti T, Lomvardas S, Parekh B, Yie J, Maniatis T, Thanos D. 2000. Ordered Recruitment of Chromatin Modifying and General Transcription Factors to the IFN-β Promoter. *Cell* **103**: 667–678.

Aksoy M, Pootakham W, Grossman AR. 2014. Critical function of a Chlamydomonas reinhardtii putative polyphosphate polymerase subunit during nutrient deprivation. *The Plant Cell* **26**: 4214–4229.

Aksu Z. 1998. Biosorption of Heavy Metals by Microalgae in Batch and Continuous Systems. In: Wong Y-S,, In: Tam NFY, eds. Biotechnology Intelligence Unit. Wastewater Treatment with Algae. Springer Berlin Heidelberg, 37–53.

Alipanah L, Rohloff J, Winge P, Bones AM, Brembu T. 2015. Whole-cell response to nitrogen deprivation in the diatom Phaeodactylum tricornutum. *Journal of Experimental Botany*: erv340.

Allen AE, Dupont CL, Oborník M, Horák A, Nunes-Nesi A, McCrow JP, Zheng H, Johnson DA, Hu H, Fernie AR, *et al.* 2011. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* **473**: 203–207.

Amoutzias GD, Robertson DL, Van de Peer Y, Oliver SG. 2008. Choose your partners: dimerization in eukaryotic transcription factors. *Trends in Biochemical Sciences* **33**: 220–229.

Amoutzias GD, Veron AS, Weiner J, Robinson-Rechavi M, Bornberg-Bauer E, Oliver SG, Robertson DL. 2007. One billion years of bZIP transcription factor evolution: conservation and change in dimerization and DNA-binding site specificity. *Molecular Biology and Evolution* **24**: 827–835.

Andersen RA. 1992. Diversity of eukaryotic algae. *Biodiversity & Conservation* **1**: 267–292.

Andersen RA. 2004. Biology and systematics of heterokont and haptophyte algae. *American journal of botany* **91**: 1508–1522.

Andersson JO. 2005. Lateral gene transfer in eukaryotes. *Cellular and molecular life sciences: CMLS* **62**: 1182–1197.

Aoki K, Ogata Y, Shibata D. 2007. Approaches for Extracting Practical Information from Gene Co-expression Networks in Plant Biology. *Plant and Cell Physiology* **48**: 381–390.

Aravind L, Koonin EV. **1999**. DNA-binding proteins and evolution of transcription regulation in the archaea. *Nucleic Acids Research* **27**: 4658–4670.

Archibald JM. **2007**. Nucleomorph genomes: structure, function, origin and evolution. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* **29**: 392–402.

Arora S, Rana R, Chhabra A, Jaiswal A, Rani V. **2013**. miRNA–transcription factor interactions: a combinatorial regulation of gene expression. *Molecular Genetics and Genomics* **288**: 77–87.

Arsovski AA, Pradinuk J, Guo XQ, Wang S, Adams KL. **2015**. Evolution of Cis-Regulatory Elements and Regulatory Networks in Duplicated Genes of Arabidopsis1[OPEN]. *Plant Physiology* **169**: 2982–2991.

Austin RW, Petzold TJ. **1986**. Spectral Dependence Of The Diffuse Attenuation Coefficient Of Light In Ocean Waters. *Optical Engineering* **25**: 253471-253471-.

Avila J, González C, Brito N, Machín F, Pérez MD, Siverio JM. **2002**. A second Zn(II)(2)Cys(6) transcriptional factor encoded by the YNA2 gene is indispensable for the transcriptional activation of the genes involved in nitrate assimilation in the yeast Hansenula polymorpha. *Yeast (Chichester, England)* **19**: 537–544.

Avila J, González C, Brito N, Siverio JM. **1998**. Clustering of the YNA1 gene encoding a Zn(II)2Cys6 transcriptional factor in the yeast Hansenula polymorpha with the nitrate assimilation genes YNT1, YNI1 and YNR1, and its involvement in their transcriptional activation. *The Biochemical journal* **335 (Pt 3)**: 647–652.

Baroukh C, Muñoz-Tamayo R, Steyer J-P, Bernard O. **2015**. A state of the art of metabolic networks of unicellular microalgae and cyanobacteria for biofuel production. *Metabolic Engineering* **30**: 49–60.

Basehoar AD, Zanton SJ, Pugh BF. **2004**. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116**: 699–709.

Bastian M, Heymann S, Jacomy M. **2009**. Gephi: An Open Source Software for Exploring and Manipulating Networks. Third International AAAI Conference on Weblogs and Social Media.

Baurain D, Brinkmann H, Petersen J, Rodríguez-Ezpeleta N, Stechmann A, Demoulin V, Roger AJ, Burger G, Lang BF, Philippe H. **2010**. Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. *Molecular biology and evolution* **27**: 1698–1709.

Beck A, Divakar PK, Zhang N, Molina MC, Struwe L. **2014**. Evidence of ancient horizontal gene transfer between fungi and the terrestrial alga Trebouxia. *Organisms Diversity & Evolution* **15**: 235–248.

Becker W. **2003**. Microalgae in Human and Animal Nutrition. In: Richmond A, ed. Handbook of Microalgal Culture. Blackwell Publishing Ltd, 312–351.

Becker B, Hoef-Emden K, Melkonian M. **2008**. Chlamydial genes shed light on the evolution of photoautotrophic eukaryotes. *BMC Evolutionary Biology* **8**: 203.

Becker A, Winter K-U, Meyer B, Saedler H, Theißen G. **2000**. MADS-Box Gene Diversity in Seed Plants 300 Million Years Ago. *Molecular Biology and Evolution* **17**: 1425–1434.

Bendif EM, Probert I, Schroeder DC, Vargas C de. **2013**. On the description of Tisochrysis lutea gen. nov. sp. nov. and Isochrysis nuda sp. nov. in the Isochrysidales, and the transfer of Dicrateria to the Prymnesiales (Haptophyta). *Journal of Applied Phycology* **25**: 1763–1776.

Benemann JR. Microalgae aquaculture feeds. *Journal of Applied Phycology* **4**: 233–245.

Bertsch C, Beuve M, Dolja VV, Wirth M, Pelsy F, Herrbach E, Lemaire O. **2009**. Retention of the virus-derived sequences in the nuclear genome of grapevine as a potential pathway to virus resistance. *Biology Direct* **4**: 21.

Bhattacharya D, Helmchen T, Melkonian M. **1995**. Molecular Evolutionary Analyses of Nuclear-Encoded Small Subunit Ribosomal RNA Identify an Independent Rhizopod Lineage Containing the Euglyphina and the Chlorarachniophyta. *Journal of Eukaryotic Microbiology* **42**: 65–69.

Bhattacharya D, Price DC, Xin Chan C, Qiu H, Rose N, Ball S, Weber APM, Cecilia Arias M, Henrissat B, Coutinho PM, *et al.* 2013. Genome of the red alga Porphyridium purpureum. *Nature Communications* **4**.

Bhattacharya D, Qiu H, Price DC, Yoon HS. **2015**. Why we need more algal genomes. *Journal of Phycology* **51**: 1–5.

Blackwood EM, Kadonaga JT. **1998**. Going the distance: a current view of enhancer action. *Science (New York, N.Y.)* **281**: 60–63.

Blanc G, Wolfe KH. **2004**. Functional Divergence of Duplicated Genes Formed by Polyploidy during Arabidopsis Evolution. *The Plant Cell Online* **16**: 1679–1691.

Borowitzka MA. **1997**. Microalgae as sources of pharmaceuticals and other biologically active compounds. *Journal of Applied Phycology* **7**: 3–15.

Bougaran G, Bernard O, Sciandra A. **2010**. Modeling continuous cultures of microalgae colimited by nitrogen and phosphorus. *Journal of Theoretical Biology* **265**: 443–454.

Bougaran G, Le Déan L, Lukomska E, Kaas R, Baron R. **2003**. Transient initial phase in continuous culture of Isochrysis galbana affinis Tahiti. *Aquatic Living Resources* **16**: 389–394.

Bougaran G, Rouxel C, Dubois N, Kaas R, Grouas S, Lukomska E, Le Coz J-R, Cadoret J-P. **2012**. Enhancement of neutral lipid productivity in the microalga Isochrysis affinis Galbana (T-Iso) by a mutation-selection procedure. *Biotechnology and bioengineering* **109**: 2737–2745.

Bourbon H-M. **2008**. Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex. *Nucleic Acids Research* **36**: 3993–4008.

Bowler C, Karl DM, Colwell RR. **2009**. Microbial oceanography in a sea of opportunity. *Nature* **459**: 180–184.

Boyle NR, Page MD, Liu B, Blaby IK, Casero D, Kropat J, Cokus SJ, Hong-Hermesdorf A, Shaw J, Karpowicz SJ, *et al.* 2012. Three acyltransferases and nitrogen-responsive regulator are implicated in nitrogen starvation-induced triacylglycerol accumulation in Chlamydomonas. *The Journal of biological chemistry* **287**: 15811–15825.

Brennan L, Owende P. **2013**. Biofuels from Microalgae: Towards Meeting Advanced Fuel Standards. In: Lee JW, ed. Advanced Biofuels and Bioproducts. Springer New York, 553–599.

Brivanlou AH, Darnell JE. **2002**. Signal Transduction and the Control of Gene Expression. *Science* **295**: 813–818.

Brkljacic J, Grotewold E. **2016**. Combinatorial control of plant gene expression. *Biochimica Et Biophysica Acta*.

Brown MR, Jeffrey SW, Volkman JK, Dunstan GA. **1997**. Fish Nutrition and Feeding Proceedings of the Sixth International Symposium on Feeding and Nutrition in FishNutritional properties of microalgae for mariculture. *Aquaculture* **151**: 315–331.

Brownlee C, Wheeler GL, Taylor AR. **2015**. Coccolithophore biomineralization: New questions, new answers. *Seminars in Cell & Developmental Biology* **46**: 11–16.

Buitrago-Flórez FJ, Restrepo S, Riaño-Pachón DM. **2014**. Identification of transcription factor genes and their correlation with the high diversity of stramenopiles. *PloS One* **9**: e111841.

Bunyavanich S, Schadt EE, Himes BE, Lasky-Su J, Qiu W, Lazarus R, Ziniti JP, Cohain A, Linderman M, Torgerson DG, *et al.* 2014. Integrated genome-wide association, coexpression network, and expression single nucleotide polymorphism analysis identifies novel pathway in allergic rhinitis. *BMC Medical Genomics* **7**: 48.

Burki F. **2014**. The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harbor Perspectives in Biology* **6**: a016147.

Burki F, Kudryavtsev A, Matz MV, Aglyamova GV, Bulman S, Fiers M, Keeling PJ, Pawlowski J. **2010**. Evolution of Rhizaria: new insights from phylogenomic analysis of uncultivated protists. *BMC evolutionary biology* **10**: 377.

Burki F, Okamoto N, Pombert J-F, Keeling PJ. **2012**. The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. *Proceedings. Biological sciences / The Royal Society* **279**: 2246–2254.

Burki F, Shalchian-Tabrizi K, Pawlowski J. **2008**. Phylogenomics reveals a new 'megagroup' including most photosynthetic eukaryotes. *Biology Letters* **4**: 366–369.

Cadoret J-P, Bernard O. **2008**. La production de biocarburant lipidique avec des microalgues : promesses et défis. *Journal de la Société de Biologie* **202**: 201–211.

Cadoret J-P, Bougaran G, Bérard J-B, Carrier G, Charrier A, Coulombier N, Garnier M, Kaas R, Le Déan L, Lukomska E, *et al.* 2014. Microalgae and Biotechnology. In: Monaco A,, In: Prouzet P, eds. Development of Marine Resources. John Wiley & Sons, Inc., 57–115.

Cadoret J-P, Garnier M, Saint-Jean B. 2012. Chapter Eight - Microalgae, Functional Genomics and Biotechnology. In: Piganeau G, ed. Genomic Insights into the Biology of Algae. Advances in Botanical Research. Academic Press, 285–341.

Cannons AC, Shiflett SD. 2001. Transcriptional regulation of the nitrate reductase gene in Chlorella vulgaris: identification of regulatory elements controlling expression. *Current Genetics* **40**: 128–135.

Capell T, Christou P. 2004. Progress in plant metabolic engineering. *Current Opinion in Biotechnology* **15**: 148–154.

Carretero-Paulet L, Galstyan A, Roig-Villanova I, Martínez-García JF, Bilbao-Castro JR, Robertson DL. 2010. Genome-Wide Classification and Evolutionary Analysis of the bHLH Family of Transcription Factors in Arabidopsis, Poplar, Rice, Moss, and Algae. *Plant Physiology* **153**: 1398–1412.

Carrier G, Garnier M, Le Cunff L, Bougaran G, Probert I, De Vargas C, Corre E, Cadoret J-P, Saint-Jean B. 2014. Comparative Transcriptome of Wild Type and Selected Strains of the Microalgae Tisochrysis lutea Provides Insights into the Genetic Basis, Lipid Metabolism and the Life Cycle. *PloS one* **9**: e86889.

Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, Eeckhoute J, Shao W, Hestermann EV, Geistlinger TR, *et al.* 2005. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122**: 33–43.

Carter SL, Brechbühler CM, Griffin M, Bond AT. 2004a. Gene co-expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics* **20**: 2242–2250.

Carter SL, Brechbühler CM, Griffin M, Bond AT. 2004b. Gene co-expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics* **20**: 2242–2250.

Carthew RW, Sontheimer EJ. 2009. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**: 642–655.

Cavalier-Smith T. 1982. The origins of plastids. *Biological Journal of the Linnean Society* **17**: 289–306.

Cavalier-Smith T. 1999. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *The Journal of eukaryotic microbiology* **46**: 347–366.

Chan CX, Reyes-Prieto A, Bhattacharya D. **2011**. Red and green algal origin of diatom membrane transporters: insights into environmental adaptation and cell evolution. *PloS One* **6**: e29138.

Chang RL, Ghamsari L, Manichaikul A, Hom EFY, Balaji S, Fu W, Shen Y, Hao T, Palsson BØ, Salehi-Ashtiani K, *et al.* 2011. Metabolic network reconstruction of Chlamydomonas offers insight into light-driven algal metabolism. *Molecular Systems Biology* **7**: 518.

Chang M, Jaehning JA. **1997**. A multiplicity of mediators: alternative forms of transcription complexes communicate with transcriptional regulators. *Nucleic Acids Research* **25**: 4861–4865.

Charitou T, Bryan K, Lynn DJ. **2016**. Using biological networks to integrate, visualize and analyze genomics data. *Genetics Selection Evolution* **48**: 27.

Charoensawan V, Wilson D, Teichmann SA. **2010a**. Lineage-specific expansion of DNA-binding transcription factor families. *Trends in Genetics* **26**: 388–393.

Charoensawan V, Wilson D, Teichmann SA. **2010b**. Genomic repertoires of DNA-binding transcription factors across the tree of life. *Nucleic acids research* **38**: 7364–7377.

Charrier A, Bérard J-B, Bougaran G, Carrier G, Lukomska E, Schreiber N, Fournier F, Charrier AF, Rouxel C, Garnier M, *et al.* 2015. High-affinity nitrate/nitrite transporter genes (Nrt2) in Tisochrysis lutea: identification and expression analyses reveal some interesting specificities of Haptophyta microalgae. *Physiologia Plantarum* **154**: 572–590.

de Château M, Björck L. **1994**. Protein PAB, a mosaic albumin-binding bacterial protein representing the first contemporary example of module shuffling. *The Journal of Biological Chemistry* **269**: 12147–12151.

Chen B, Huang Q, Lin X, Shi Q, Wu S. **1998**. Accumulation of Ag, Cd, Co, Cu, Hg, Ni and Pb in Pavlova viridis Tseng (Haptophyceae). *Journal of Applied Phycology* **10**: 371–376.

Chen D-S, Wu Y-Q, Zhang W, Jiang S-J, Chen S-Z. **2016**. Horizontal gene transfer events reshape the global landscape of arm race between viruses and homo sapiens. *Scientific Reports* **6**.

Chiang CM, Roeder RG. **1995**. Cloning of an intrinsic human TFIID subunit that interacts with multiple transcriptional activators. *Science (New York, N.Y.)* **267**: 531–536.

Chisti Y. **2013**. Constraints to commercialization of algal fuels. *Journal of Biotechnology* **167**: 201–214.

Choy B, Green MR. **1993**. Eukaryotic activators function during multiple steps of preinitiation complex assembly. *Nature* **366**: 531–536.

Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS. **2002**. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Molecular Cell* **9**: 279–289.

Cirillo LA, Zaret KS. **1999**. An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Molecular Cell* **4**: 961–969.

Clark KL, Halay ED, Lai E, Burley SK. **1993**. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* **364**: 412–420.

Conaway RC, Brower CS, Conaway JW. **2002**. Emerging Roles of Ubiquitin in Transcription Regulation. *Science* **296**: 1254–1258.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. **2005**. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics (Oxford, England)* **21**: 3674–3676.

Cook PR. **1999**. The Organization of Replication and Transcription. *Science* **284**: 1790–1795.

Cooper TF, Morby AP, Gunn A, Schneider D. **2006**. Effect of random and hub gene disruptions on environmental and mutational robustness in Escherichia coli. *Bmc Genomics* **7**: 1.

Courchesne NMD, Parisien A, Wang B, Lan CQ. **2009**. Enhancement of lipid production using biochemical, genetic and transcription factor engineering approaches. *Journal of biotechnology* **141**: 31–41.

Courey AJ, Tjian R. **1988**. Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. *Cell* **55**: 887–898.

Coutteau P, Sorgeloos P. **1992**. The use of algal substitutes and the requirement for live algae in the hatchery and nursery rearing of bivalve molluscs: an international survey. *Journal of Shellfish Research*.

Czemmel S, Stracke R, Weisshaar B, Cordon N, Harris NN, Walker AR, Robinson SP, Bogs J. **2009**. The grapevine R2R3-MYB transcription factor VvMYBF1 regulates flavonol synthesis in developing grape berries. *Plant Physiology* **151**: 1513–1530.

von Dassow P, Ogata H, Probert I, Wincker P, Da Silva C, Audic S, Claverie J-M, de Vargas C. **2009**. Transcriptome analysis of functional differentiation between haploid and diploid cells of Emiliania huxleyi, a globally significant photosynthetic calcifying cell. *Genome Biology* **10**: R114.

Deato MDE, Tjian R. **2008**. An unexpected role of TAFs and TRFs in skeletal muscle differentiation: switching core promoter complexes. *Cold Spring Harbor Symposia on Quantitative Biology* **73**: 217–225.

DeFalco J, Childs G. **1996**. The embryonic transcription factor stage specific activator protein contains a potent bipartite activation domain that interacts with several RNA polymerase II basal transcription factors. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 5802–5807.

Delwiche. **1999**. Tracing the Thread of Plastid Diversity through the Tapestry of Life. *The American naturalist* **154**: S164–S177.

Deschamps P, Moreira D. **2012**. Reevaluating the green contribution to diatom genomes. *Genome Biology and Evolution* **4**: 683–688.

Dong H-P, Williams E, Wang D, Xie Z-X, Hsia R, Jenck A, Halden R, Li J, Chen F, Place AR. **2013**. Responses of Nannochloropsis oceanica IMET1 to Long-Term Nitrogen Starvation and Recovery. *Plant Physiology* **162**: 1110–1126.

Doolittle WF. **1998**. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends in genetics: TIG* **14**: 307–311.

Dorrell RG, Smith AG. **2011**. Do red and green make brown?: perspectives on plastid acquisitions within chromalveolates. *Eukaryotic Cell* **10**: 856–868.

Dyhrman ST, Jenkins BD, Rynearson TA, Saito MA, Mercier ML, Alexander H, Whitney LP, Drzewianowski A, Bulygin VV, Bertrand EM, *et al.* 2012. The Transcriptome and Proteome of the Diatom Thalassiosira pseudonana Reveal a Diverse Phosphorus Stress Response. *PLOS ONE* **7**: e33768.

Edger PP, Pires JC. **2009**. Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology* **17**: 699–717.

Eggleston AK. **2009**. RNA silencing. *Nature* **457**: 395.

Eglinton TI, Eglinton G. **2008**. Molecular proxies for paleoclimatology. *Earth and Planetary Science Letters* **275**: 1–16.

Elias M, Patron NJ, Keeling PJ. **2009**. The RAB family GTPase Rab1A from Plasmodium falciparum defines a unique paralog shared by chromalveolates and rhizaria. *The Journal of Eukaryotic Microbiology* **56**: 348–356.

El-Sharkawy I, Liang D, Xu K. **2015**. Transcriptome analysis of an apple (Malus × domestica) yellow fruit somatic mutation identifies a gene network module highly associated with anthocyanin and epigenetic regulation. *Journal of Experimental Botany* **66**: 7359–7376.

Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield O, Taylor FJR. **2004**. The evolution of modern eukaryotic phytoplankton. *Science (New York, N.Y.)* **305**: 354–360.

Farinas B, Mary C, de O Manes C-L, Bhaud Y, Peaucellier G, Moreau H. **2006**. Natural synchronisation for the study of cell division in the green unicellular alga Ostreococcus tauri. *Plant Molecular Biology* **60**: 277–292.

Feller A, Machemer K, Braun EL, Grotewold E. **2011**. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *The Plant journal: for cell and molecular biology* **66**: 94–116.

Felsenfeld G, Groudine M. **2003**. Controlling the double helix. *Nature* **421**: 448–453.

Feng J, Meyer CA, Wang Q, Liu JS, Shirley Liu X, Zhang Y. 2012. GFOLD: a generalized fold change for ranking differentially expressed genes from RNA-seq data. *Bioinformatics (Oxford, England)* **28**: 2782–2788.

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. 1998. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science* **281**: 237–240.

Fields MW, Hise A, Lohman EJ, Bell T, Gardner RD, Corredor L, Moll K, Peyton BM, Characklis GW, Gerlach R. 2014. Sources and resources: importance of nutrients, resource allocation, and ecology in microalgal cultivation for lipid accumulation. *Applied microbiology and biotechnology*.

Filée J. 2014. Multiple occurrences of giant virus core genes acquired by eukaryotic genomes: the visible part of the iceberg? *Virology* **466–467**: 53–59.

Fortin I. 2005. Domaines protéiques du complexe histone acétyltransférase NuA4 impliqués dans la transciption et le maintien de l'intégrité du génome. Thèse de doctorat.

Fouilland E. 2012. Biodiversity as a tool for waste phycoremediation and biomass production. *Reviews in Environmental Science and Bio/Technology* **11**: 1–4.

Franco-Zorrilla JM, Solano R. 2017. Identification of plant transcription factor target sequences. *Biochimica Et Biophysica Acta* **1860**: 21–30.

Franklin SE, Mayfield SP. 2005. Recent developments in the production of human therapeutic proteins in eukaryotic algae. *Expert Opinion on Biological Therapy* **5**: 225–235.

Fraser P, Bickmore W. 2007. Nuclear organization of the genome and the potential for gene regulation. *Nature* **447**: 413–417.

Frommolt R, Werner S, Paulsen H, Goss R, Wilhelm C, Zauner S, Maier UG, Grossman AR, Bhattacharya D, Lohr M. 2008. Ancient recruitment by chromists of green algal genes encoding enzymes for carotenoid biosynthesis. *Molecular Biology and Evolution* **25**: 2653–2667.

Frontini M, Imbriano C, Manni I, Mantovani R. 2004. Cell cycle regulation of NF-YC nuclear localization. *Cell Cycle (Georgetown, Tex.)* **3**: 217–222.

Gao C, Ren X, Mason AS, Liu H, Xiao M, Li J, Fu D. 2014. Horizontal gene transfer in plants. *Functional & Integrative Genomics* **14**: 23–29.

Gao R, Stock AM. 2015. Temporal hierarchy of gene expression mediated by transcription factor binding affinity and activation dynamics. *mBio* **6**: e00686-615.

Gargouri M, Park J-J, Holguin FO, Kim M-J, Wang H, Deshpande RR, Shachar-Hill Y, Hicks LM, Gang DR. 2015. Identification of regulatory network hubs that control lipid metabolism in Chlamydomonas reinhardtii. *Journal of Experimental Botany* **66**: 4551–4566.

Garnier M. **2016**. Recherche et étude de protéines candidates impliquées dans le stress azoté et l'accumulation lipidique chez l'haptophyte Isochrysis aff galbana. Thèse de doctorat, ISBN.

Garnier M, Carrier G, Rogniaux H, Nicolau E, Bougaran G, Saint-Jean B, Cadoret JP. **2014**. Comparative proteomics reveals proteins impacted by nitrogen deprivation in wild-type and high lipid-accumulating mutant strains of Tisochrysis lutea. *Journal of Proteomics* **105**: 107–120.

Ge Z, Liu C, Björkholm M, Gruber A, Xu D. **2006**. Mitogen-Activated Protein Kinase Cascade-Mediated Histone H3 Phosphorylation Is Critical for Telomerase Reverse Transcriptase Expression/Telomerase Activation Induced by Proliferation. *Molecular and Cellular Biology* **26**: 230–237.

Geider RJ, Delucia EH, Falkowski PG, Finzi AC, Grime JP, Grace J, Kana TM, La Roche J, Long SP, Osborne BA, *et al.* 2001. Primary productivity of planet earth: biological determinants and physical constraints in terrestrial and aquatic habitats. *Global Change Biology* **7**: 849–882.

Gerber HP, Seipel K, Georgiev O, Höfferer M, Hug M, Rusconi S, Schaffner W. **1994**. Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science (New York, N.Y.)* **263**: 808–811.

Ghildiyal M, Zamore PD. **2009**. Small silencing RNAs: an expanding universe. *Nature reviews. Genetics* **10**: 94–108.

Gilbert C, Chateigner A, Ernenwein L, Barbe V, Bézier A, Herniou EA, Cordaux R. **2014**. Population genomics supports baculoviruses as vectors of horizontal transfer of insect transposons. *Nature Communications* **5**: 3348.

Gill G. **2005**. Something about SUMO inhibits transcription. *Current Opinion in Genetics & Development* **15**: 536–541.

Gill G, Ptashne M. **1987**. Mutants of GAL4 protein altered in an activation function. *Cell* **51**: 121–126.

Gilson PR, Su V, Slamovits CH, Reith ME, Keeling PJ, McFadden GI. **2006**. Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 9566–9571.

Godos I de, Blanco S, García-Encina PA, Becares E, Muñoz R. **2009**. Long-term operation of high rate algal ponds for the bioremediation of piggery wastewaters at high loading rates. *Bioresource Technology* **100**: 4332–4339.

Goncalves EC, Wilkie AC, Kirst M, Rathinasabapathi B. **2016**. Metabolic regulation of triacylglycerol accumulation in the green algae: identification of potential targets for engineering to improve oil yield. *Plant Biotechnology Journal*.

Gould SB. **2012**. Evolutionary genomics: Algae's complex origins. *Nature* **492**: 46–48.

Green BR. **2004**. The chloroplast genome of dinoflagellates--a reduced instruction set? *Protist* **155**: 23–31.

Gualdi R, Bossard P, Zheng M, Hamada Y, Coleman JR, Zaret KS. **1996**. Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. *Genes & Development* **10**: 1670–1682.

Guarnieri MT, Nag A, Yang S, Pienkos PT. **2013**. Proteomic analysis of Chlorella vulgaris: potential targets for enhanced lipid accumulation. *Journal of proteomics* **93**: 245–253.

Guertin MJ, Lis JT. **2010**. Chromatin landscape dictates HSF binding to target DNA elements. *PLoS genetics* **6**: e1001114.

Guiry MD. **2012**. How Many Species of Algae Are There? *Journal of Phycology* **48**: 1057–1063.

Gutiérrez RA, Green PJ, Keegstra K, Ohlrogge JB. **2004**. Phylogenetic profiling of the Arabidopsis thaliana proteome: what proteins distinguish plants from other organisms? *Genome biology* **5**: R53.

Haag JR, Pikaard CS. **2011**. Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nature Reviews. Molecular Cell Biology* **12**: 483–492.

Habib B, Parvin H, Hasan M. **2008**. A review on culture, production and use of spirulina as food for humans and feeds for domestic animals. FAO Fisheries and Aquaculture Department.

Hackett JD, Yoon HS, Li S, Reyes-Prieto A, Rümmele SE, Bhattacharya D. **2007**. Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates. *Molecular Biology and Evolution* **24**: 1702–1713.

Hagopian JC, Reis M, Kitajima JP, Bhattacharya D, de Oliveira MC. **2004**. Comparative analysis of the complete plastid genome sequence of the red alga Gracilaria tenuistipitata var. liui provides insights into the evolution of rhodoplasts and their relationship to other plastids. *Journal of Molecular Evolution* **59**: 464–477.

Hahn S. **1993**. Structure(?) and function of acidic transcription activators. *Cell* **72**: 481–483.

Hahn MW, Kern AD. **2005**. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Molecular Biology and Evolution* **22**: 803–806.

Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, Shiu S-H. **2008**. Importance of Lineage-Specific Expansion of Plant Tandem Duplicates in the Adaptive Response to Environmental Stimuli. *Plant Physiology* **148**: 993–1003.

Harper JT, Waanders E, Keeling PJ. **2005**. On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes. *International Journal of Systematic and Evolutionary Microbiology* **55**: 487–496.

Hedges SB, Blair JE, Venturi ML, Shoe JL. **2004**. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evolutionary Biology* **4**: 2.

Hemaiswarya S, Raja R, Ravi Kumar R, Ganesan V, Anbazhagan C. 2011. Microalgae: a sustainable feed source for aquaculture. *World Journal of Microbiology and Biotechnology* **27**: 1737–1746.

Hengartner CJ, Thompson CM, Zhang J, Chao DM, Liao SM, Koleske AJ, Okamura S, Young RA. 1995. Association of an activator with an RNA polymerase II holoenzyme. *Genes & Development* **9**: 897–910.

Herr AJ. 2005. Pathways through the small RNA world of plants. *FEBS Letters* **579**: 5879–5888.

Hildebrand M, Dahlin K. 2000. NITRATE TRANSPORTER GENES FROM THE DIATOM CYLINDROTHECA FUSIFORMIS (BACILLARIOPHYCEAE): mRNA LEVELS CONTROLLED BY NITROGEN SOURCE AND BY THE CELL CYCLE. *Journal of Phycology* **36**: 702–713.

Ho L, Crabtree GR. 2010. Chromatin remodelling during development. *Nature* **463**: 474–484.

Hochberg M, Kohen R, Enk CD. 2006. Role of antioxidants in prevention of pyrimidine dimer formation in UVB irradiated human HaCaT keratinocytes. *Biomedicine & Pharmacotherapy = Biomédecine & Pharmacothérapie* **60**: 233–237.

Hollender CA, Kang C, Darwish O, Geretz A, Matthews BF, Slovin J, Alkharouf N, Liu Z. 2014. Floral Transcriptomes in Woodland Strawberry Uncover Developing Receptacle and Anther Gene Networks1[W][OPEN]. *Plant Physiology* **165**: 1062–1075.

Hope IA, Struhl K. 1986. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell* **46**: 885–894.

Horikoshi M, Carey MF, Kakidani H, Roeder RG. 1988. Mechanism of action of a yeast activator: Direct effect of GAL4 derivatives on mammalian TFIID-promoter interactions. *Cell* **54**: 665–669.

Hovde BT, Deodato CR, Hunsperger HM, Ryken SA, Yost W, Jha RK, Patterson J, Jr RJM, Barlow SB, Starkenburg SR, *et al.* 2015. Genome Sequence and Transcriptome Analyses of Chrysochromulina tobin : Metabolic Tools for Enhanced Algal Fitness in the Prominent Order Prymnesiales (Haptophyceae). *PLOS Genet* **11**: e1005469.

Hsieh H-J, Su C-H, Chien L-J. 2012. Accumulation of lipid production in Chlorella minutissima by triacylglycerol biosynthesis-related genes cloned from Saccharomyces cerevisiae and Yarrowia lipolytica. *Journal of Microbiology (Seoul, Korea)* **50**: 526–534.

Hsu H-T, Chen H-M, Yang Z, Wang J, Lee NK, Burger A, Zaret K, Liu T, Levine E, Mango SE. 2015. Recruitment of RNA polymerase II by the pioneer transcription factor PHA-4. *Science* **348**: 1372–1376.

Hu J, Wang D, Li J, Jing G, Ning K, Xu J. 2014. Genome-wide identification of transcription factors and transcription-factor binding sites in oleaginous microalgae Nannochloropsis. *Scientific Reports* **4**.

Huang J, Gogarten JP. **2007**. Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? *Genome Biology* **8**: R99.

Huang J, Mullapudi N, Lancto CA, Scott M, Abrahamsen MS, Kissinger JC. **2004**. Phylogenomic evidence supports past endosymbiosis, intracellular and horizontal gene transfer in Cryptosporidium parvum. *Genome Biology* **5**: R88.

Hunsperger HM, Randhawa T, Cattolico RA. **2015**. Extensive horizontal gene transfer, duplication, and loss of chlorophyll synthesis genes in the algae. *BMC Evolutionary Biology* **15**.

Huysman MJJ, Fortunato AE, Matthijs M, Costa BS, Vanderhaeghen R, Van den Daele H, Sachse M, Inzé D, Bowler C, Kroth PG, *et al.* 2013. AUREOCHROME1a-mediated induction of the diatom-specific cyclin dsCYC2 controls the onset of cell division in diatoms (Phaeodactylum tricornutum). *The Plant cell* **25**: 215–228.

Ibáñez-Salazar A, Rosales-Mendoza S, Rocha-Uribe A, Ramírez-Alonso JI, Lara-Hernández I, Hernández-Torres A, Paz-Maldonado LMT, Silva-Ramírez AS, Bañuelos-Hernández B, Martínez-Salgado JL, *et al.* 2014. Over-expression of Dof-type transcription factor increases lipid production in Chlamydomonas reinhardtii. *Journal of biotechnology*.

Iborra FJ, Pombo A, Jackson DA, Cook PR. **1996**. Active RNA polymerases are localized within discrete transcription "factories' in human nuclei. *Journal of Cell Science* **109 (Pt 6)**: 1427–1436.

Imamura S, Kanesaki Y, Ohnuma M, Inouye T, Sekine Y, Fujiwara T, Kuroiwa T, Tanaka K. **2009**. R2R3-type MYB transcription factor, CmMYB1, is a central nitrogen assimilation regulator in Cyanidioschyzon merolae. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 12548–12553.

Imoto Y, Yoshida Y, Yagisawa F, Kuroiwa H, Kuroiwa T. **2011**. The cell cycle, including the mitotic cycle and organelle division cycles, as revealed by cytological observations. *Journal of Electron Microscopy* **60 Suppl 1**: S117-136.

Ishikawa M, Takahashi F, Nozaki H, Nagasato C, Motomura T, Kataoka H. **2009**. Distribution and phylogeny of the blue light receptors aureochromes in eukaryotes. *Planta* **230**: 543–552.

Iyer LM, Koonin EV, Aravind L. **2002**. Extensive domain shuffling in transcription regulators of DNA viruses and implications for the origin of fungal APSES transcription factors. *Genome Biology* **3**: RESEARCH0012.

Janouškovec J, Horák A, Oborník M, Lukeš J, Keeling PJ. **2010**. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 10949–10954.

Jardillier L, Zubkov MV, Pearman J, Scanlan DJ. **2010**. Significant CO2 fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean. *The ISME Journal* **4**: 1180–1192.

Jeong H, Mason SP, Barabási A-L, Oltvai ZN. **2001**. Lethality and centrality in protein networks. *Nature* **411**: 41–42.

de Jesus Raposo MF, de Morais RMSC, de Morais AMMB. **2013**. Health applications of bioactive compounds from marine microalgae. *Life Sciences* **93**: 479–486.

Jin J, Zhang H, Kong L, Gao G, Luo J. **2014**. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic acids research* **42**: D1182-1187.

Jinkerson RE, Jonikas MC. **2015**. Molecular techniques to interrogate and edit the Chlamydomonas nuclear genome. *The Plant Journal* **82**: 393–412.

Johnson PF, Sterneck E, Williams SC. **1993**. Activation domains of transcriptional regulatory proteins. *The Journal of Nutritional Biochemistry* **4**: 386–398.

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. *Bioinformatics (Oxford, England)* **30**: 1236–1240.

Juergens MT, Deshpande RR, Lucker BF, Park J-J, Wang H, Gargouri M, Holguin FO, Disbrow B, Schaub T, Skepper JN, *et al.* 2015. The regulation of photosynthetic structure and function during nitrogen deprivation in Chlamydomonas reinhardtii. *Plant Physiology* **167**: 558–573.

Jung CH, O'Brien M, Singh MB, Bhalla PL. **2015**. Epigenetic landscape of germline specific genes in the sporophyte cells of Arabidopsis thaliana. *Frontiers in Plant Science* **6**: 328.

Kahle J, Baake M, Doenecke D, Albig W. **2005**. Subunits of the Heterotrimeric Transcription Factor NF-Y Are Imported into the Nucleus by Distinct Pathways Involving Importin β and Importin 13. *Molecular and Cellular Biology* **25**: 5339–5354.

Kang L-K, Hwang S-PL, Gong G-C, Lin H-J, Chen P-C, Chang J. **2007**. Influences of nitrogen deficiency on the transcript levels of ammonium transporter, nitrate transporter and glutamine synthetase genes in Isochrysis galbana (Isochrysidales, Haptophyta). *Phycologia* **46**: 521–533.

Kang NK, Jeon S, Kwon S, Koh HG, Shin S-E, Lee B, Choi G-G, Yang J-W, Jeong B-R, Chang YK. **2015**. Effects of overexpression of a bHLH transcription factor on biomass and lipid production in Nannochloropsis salina. *Biotechnology for Biofuels* **8**: 200.

Kang J-Y, Ryu SH, Park S-H, Cha GS, Kim D-H, Kim K-H, Hong AW, Ahn T, Pan J-G, Joung YH, *et al.* 2014. Chimeric cytochromes P450 engineered by domain swapping and random mutagenesis for producing human metabolites of drugs. *Biotechnology and Bioengineering* **111**: 1313–1322.

Kawashima T, Kawashima S, Tanaka C, Murai M, Yoneda M, Putnam NH, Rokhsar DS, Kanehisa M, Satoh N, Wada H. **2009**. Domain shuffling and the evolution of vertebrates. *Genome Research* **19**: 1393–1403.

Kazantseva J, Palm K. **2014**. Diversity in TAF proteomics: consequences for cellular differentiation and migration. *International Journal of Molecular Sciences* **15**: 16680–16697.

Keeling PJ. **2001**. Foraminifera and Cercozoa are related in actin phylogeny: two orphans find a home? *Molecular Biology and Evolution* **18**: 1551–1557.

Keeling PJ. **2013**. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annual Review of Plant Biology* **64**: 583–607.

Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, Armbrust EV, Archibald JM, Bharti AK, Bell CJ, *et al.* 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. *PLOS Biol* **12**: e1001889.

Keeling PJ, Palmer JD. **2008**. Horizontal gene transfer in eukaryotic evolution. *Nature Reviews. Genetics* **9**: 605–618.

Kim Y-J, Björklund S, Li Y, Sayre MH, Kornberg RD. **1994**. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**: 599–608.

Kim J, Iyer VR. **2004**. Global role of TATA box-binding protein recruitment to promoters in mediating gene expression profiles. *Molecular and Cellular Biology* **24**: 8104–8112.

Kim HJ, Kieber JJ, Schaller GE. **2012**. Overlapping and lineage-specific roles for the type-B response regulators of monocots and dicots. *Plant Signaling &Behavior* **7**: 1110–1113.

Kim KM, Park J-H, Bhattacharya D, Yoon HS. **2014**. Applications of next-generation sequencing to unravelling the evolutionary history of algae. *International journal of systematic and evolutionary microbiology* **64**: 333–345.

Kim TK, Roeder RG. **1994**. Proline-rich activator CTF1 targets the TFIIB assembly step during transcriptional activation. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 4170–4174.

Kitano H. **2002**. Computational systems biology. *Nature* **420**: 206–210.

Kogelman LJA, Cirera S, Zhernakova DV, Fredholm M, Franke L, Kadarmideen HN. **2014**. Identification of co-expression gene networks, regulatory genes and pathways for obesity based on adipose tissue RNA Sequencing in a porcine model. *BMC Medical Genomics* **7**: 57.

Koleske AJ, Young RA. **1994**. An RNA polymerase II holoenzyme responsive to activators. *Nature* **368**: 466–469.

Koller M, Muhr A, Braunegg G. **2014**. Microalgae as versatile cellular factories for valued products. *Algal Research* **6**, **Part A**: 52–63.

Koonin EV, Galperin MY. **2003**. Genomes and the Protein Universe.

Kornberg R. 2007. The Molecular Basis of Eukaryotic Transcription (Nobel Lecture). *Angewandte Chemie International Edition* **46**: 6956–6965.

Kouzarides T. 2000. Acetylation: a regulatory modification to rival phosphorylation? *The EMBO journal* **19**: 1176–1179.

Kouzarides T. 2007. Chromatin modifications and their function. *Cell* **128**: 693–705.

Kröger N, Poulsen N. 2008. Diatoms-from cell wall biogenesis to nanotechnology. *Annual Review of Genetics* **42**: 83–107.

Kumazaki T, Hori H, Osawa S. 1983. Phylogeny of protozoa deduced from 5S rRNA sequences. *Journal of Molecular Evolution* **19**: 411–419.

Lacour T, Sciandra A, Talec A, Mayzaud P, Bernard O. 2012. NEUTRAL LIPID AND CARBOHYDRATE PRODUCTIVITIES AS A RESPONSE TO NITROGEN STATUS IN ISOCHRYSIS SP. (T-ISO; HAPTOPHYCEAE): STARVATION VERSUS LIMITATION(1). *Journal of Phycology* **48**: 647–656.

Lacroix B, Citovsky V. 2016. Transfer of DNA from Bacteria to Eukaryotes. *mBio* **7**.

Landick R. 2009. Functional divergence in the growing family of RNA polymerases. *Structure (London, England: 1993)* **17**: 323–325.

Lang D, Weiche B, Timmerhaus G, Richardt S, Riano-Pachon DM, Correa LGG, Reski R, Mueller-Roeber B, Rensing SA. 2010. Genome-Wide Phylogenetic Comparative Analysis of Plant Transcriptional Regulation: A Timeline of Loss, Gain, Expansion, and Correlation with Complexity. *Genome Biology and Evolution* **2**: 488–503.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* **9**: 559.

Lang-Unnasch N, Reith ME, Munholland J, Barta JR. 1998. Plastids are widespread and ancient in parasites of the phylum Apicomplexa. *International Journal for Parasitology* **28**: 1743–1754.

Lee J, Kim KM, Yang EC, Miller KA, Boo SM, Bhattacharya D, Yoon HS. 2016. Reconstructing the complex evolutionary history of mobile plasmids in red algal genomes. *Scientific Reports* **6**: 23744.

Lemon B, Tjian R. 2000. Orchestrated response: a symphony of transcription factors for gene control. *Genes & development* **14**: 2551–2569.

Lenhard B, Sandelin A, Carninci P. 2012. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nature Reviews. Genetics* **13**: 233–245.

Lespinet O, Wolf YI, Koonin EV, Aravind L. 2002. The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Research* **12**: 1048–1059.

Levine M. **2011**. Paused RNA Polymerase II as a Developmental Checkpoint. *Cell* **145**: 502–511.

Levitan O, Dinamarca J, Zelzion E, Gorbunov MY, Falkowski PG. **2015**. An RNAi knock-down of nitrate reductase enhances lipid biosynthesis in the diatom Phaeodactylum tricornutum. *The Plant Journal: For Cell and Molecular Biology*.

Li B, Carey M, Workman JL. **2007**. The role of chromatin during transcription. *Cell* **128**: 707–719.

Li J, Wu Y, Qi Y. **2014**. MicroRNAs in a multicellular green alga Volvox carteri. *Science China. Life Sciences* **57**: 36–45.

Li C, Yang G, Pan J, Zhang H. **2010**. Experimental studies on dimethylsulfide (DMS) and dimethylsulfoniopropionate (DMSP) production by four marine microalgae. *Acta Oceanologica Sinica* **29**: 78–87.

Lim S-L, Chu W-L, Phang S-M. **2010**. Use of Chlorella vulgaris for bioremediation of textile wastewater. *Bioresource Technology* **101**: 7314–7322.

Liu W, Huang Z, Li P, Xia J, Chen B. **2012**. Formation of triacylglycerol in Nitzschia closterium f. minutissima under nitrogen limitation and possible physiological and biochemical mechanisms. *Journal of Experimental Marine Biology and Ecology* **418–419**: 24–29.

Liu H, Probert I, Uitz J, Claustre H, Aris-Brosou S, Frada M, Not F, de Vargas C. **2009**. Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 12803–12808.

Liu L, Ramsay T, Zinkgraf M, Sundell D, Street NR, Filkov V, Groover A. **2015**. A resource for characterizing genome-wide binding and putative target genes of transcription factors expressed during secondary growth and wood formation in Populus. *The Plant Journal: For Cell and Molecular Biology* **82**: 887–898.

López García de Lomana A, Schäuble S, Valenzuela J, Imam S, Carter W, Bilgin DD, Yohn CB, Turkarslan S, Reiss DJ, Orellana MV, *et al.* 2015. Transcriptional program for nitrogen starvation-induced lipid accumulation in Chlamydomonas reinhardtii. *Biotechnology for Biofuels* **8**: 207.

Luscombe NM, Austin SE, Berman HM, Thornton JM. **2000**. An overview of the structures of protein-DNA complexes. *Genome biology* **1**: REVIEWS001.

Lv H, Qu G, Qi X, Lu L, Tian C, Ma Y. **2013**. Transcriptome analysis of Chlamydomonas reinhardtii during the process of lipid accumulation. *Genomics* **101**: 229–237.

Lynch M, Conery JS. **2000**. The evolutionary fate and consequences of duplicate genes. *Science (New York, N.Y.)* **290**: 1151–1155.

Ma J. **2005**. Crossing the line between activation and repression. *Trends in genetics: TIG* **21**: 54–59.

Ma Y-H, Wang X, Niu Y-F, Yang Z-K, Zhang M-H, Wang Z-M, Yang W-D, Liu J-S, Li H-Y. 2014. Antisense knockdown of pyruvate dehydrogenase kinase promotes the neutral lipid accumulation in the diatom Phaeodactylum tricornutum. *Microbial Cell Factories* **13**: 100.

Mackiewicz P, Bodył A, Moszczyński K. 2013. The case of horizontal gene transfer from bacteria to the peculiar dinoflagellate plastid genome. *Mobile Genetic Elements* **3**: e25845.

MacPherson S, Larochelle M, Turcotte B. 2006. A Fungal Family of Transcriptional Regulators: the Zinc Cluster Proteins. *Microbiology and Molecular Biology Reviews* **70**: 583–604.

Maier T, Güell M, Serrano L. 2009. Correlation of mRNA and protein in complex biological samples. *FEBS letters* **583**: 3966–3973.

Mairet F, Bernard O, Masci P, Lacour T, Sciandra A. 2011. Modelling neutral lipid production by the microalga Isochrysis aff. galbana under nitrogen limitation. *Bioresource Technology* **102**: 142–149.

Marchetti J, Bougaran G, Jauffrais T, Lefebvre S, Rouxel C, Saint-Jean B, Lukomska E, Robert R, Cadoret JP. 2012. Effects of blue light on the biochemical composition and photosynthetic activity of Isochrysis sp. (T-iso). *Journal of Applied Phycology* **25**: 109–119.

Markou G, Nerantzis E. 2013. Microalgae for high-value compounds and biofuels production: a review with focus on cultivation under stress conditions. *Biotechnology Advances* **31**: 1532–1542.

Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D. 2002. Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 12246–12251.

Martínez-Bueno M, Molina-Henares AJ, Pareja E, Ramos JL, Tobes R. 2004. BacTregulators: a database of transcriptional regulators in bacteria and archaea. *Bioinformatics (Oxford, England)* **20**: 2787–2791.

Martínez ME, Sánchez S, Jiménez JM, El Yousfi F, Muñoz L. 2000. Nitrogen and phosphorus removal from urban wastewater by the microalga Scenedesmus obliquus. *Bioresource Technology* **73**: 263–272.

Maruyama S, Suzaki T, Weber AP, Archibald JM, Nozaki H. 2011. Eukaryote-to-eukaryote gene transfer gives rise to genome mosaicism in euglenids. *BMC Evolutionary Biology* **11**: 105.

Marzluf GA. 1997. Genetic regulation of nitrogen metabolism in the fungi. *Microbiology and molecular biology reviews: MMBR* **61**: 17–32.

Massari ME, Murre C. 2000. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Molecular and Cellular Biology* **20**: 429–440.

Mata TM, Martins AA, Caetano NS. **2010**. Microalgae for biodiesel production and other applications: A review. *Renewable and Sustainable Energy Reviews* **14**: 217–232.

Matthijs M, Fabris M, Broos S, Vyverman W, Goossens A. **2016**. Profiling of the Early Nitrogen Stress Response in the Diatom Phaeodactylum tricornutum Reveals a Novel Family of RING-Domain Transcription Factors. *Plant Physiology* **170**: 489–498.

Mavrich TN, Jiang C, Ioshikhes IP, Li X, Venters BJ, Zanton SJ, Tomsho LP, Qi J, Glaser RL, Schuster SC, *et al.* 2008. Nucleosome organization in the Drosophila genome. *Nature* **453**: 358–362.

McDermott JE, Taylor RC, Yoon H, Heffron F. **2009**. Bottlenecks and hubs in inferred networks are important for virulence in Salmonella typhimurium. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* **16**: 169–180.

McFadden GI. **1999**. Plastids and protein targeting. *The Journal of Eukaryotic Microbiology* **46**: 339–346.

Medlin LK, Sáez AG, Young JR. **2008**. A molecular clock for coccolithophores and implications for selectivity of phytoplankton extinctions across the K/T boundary. *Marine Micropaleontology* **67**: 69–86.

Meireles LA, Guedes AC, Malcata FX. **2003**. Increase of the yields of eicosapentaenoic and docosahexaenoic acids by the microalga Pavlova lutheri following random mutagenesis. *Biotechnology and Bioengineering* **81**: 50–55.

Mendes A, Reis A, Vasconcelos R, Guerra P, Silva TL da. **2008**. Crypthecodinium cohnii with emphasis on DHA production: a review. *Journal of Applied Phycology* **21**: 199–214.

Metting B, Pyne JW. **1986**. Biologically active compounds from microalgae. *Enzyme and Microbial Technology* **8**: 386–394.

Miller R, Wu G, Deshpande RR, Vieler A, Gärtner K, Li X, Moellering ER, Zäuner S, Cornish AJ, Liu B, *et al.* 2010. Changes in transcript abundance in Chlamydomonas reinhardtii following nitrogen deprivation predict diversion of metabolism. *Plant physiology* **154**: 1737–1752.

Milliman JD. **1993**. Production and accumulation of calcium carbonate in the ocean: Budget of a nonsteady state. *Global Biogeochemical Cycles* **7**: 927–957.

Mills JD, Kawahara Y, Janitz M. **2013**. Strand-Specific RNA-Seq Provides Greater Resolution of Transcriptome Profiling. *Current Genomics* **14**: 173–181.

Mimouni V, Ulmann L, Pasquet V, Mathieu M, Picot L, Bougaran G, Cadoret J-P, Morant-Manceau A, Schoefs B. **2012**. The potential of microalgae for the production of bioactive molecules of pharmaceutical interest. *Current Pharmaceutical Biotechnology* **13**: 2733–2750.

Miyagishima S, Suzuki K, Okazaki K, Kabeya Y. **2012**. Expression of the Nucleus-Encoded Chloroplast Division Genes and Proteins Regulated by the Algal Cell Cycle. *Molecular Biology and Evolution* **29**: 2957–2970.

Molnar A, Melnyk C, Baulcombe DC. **2011**. Silencing signals in plants: a long journey for small RNAs. *Genome Biology* **12**: 215.

Molnár A, Schwach F, Studholme DJ, Thuenemann EC, Baulcombe DC. **2007**. miRNAs control gene expression in the single-cell alga Chlamydomonas reinhardtii. *Nature* **447**: 1126–1129.

Monier A, Pagarete A, de Vargas C, Allen MJ, Read B, Claverie J-M, Ogata H. **2009**. Horizontal gene transfer of an entire metabolic pathway between a eukaryotic alga and its DNA virus. *Genome Research* **19**: 1441–1449.

Moon-van der Staay SY, van der Staay GW, Guillou L, Vaulot D, Claustre H, Medlin LK. **2000**. Abundance and diversity of prymnesiophytes in the picoplankton coμmunity from the equatorial Pacific Ocean inferred from 18S rDNA sequences. *Limnology and Oceanography* **45**: 98–109.

Moore RB, Oborník M, Janouškovec J, Chrudimský T, Vancová M, Green DH, Wright SW, Davies NW, Bolch CJS, Heimann K, *et al.* 2008. A photosynthetic alveolate closely related to apicomplexan parasites. *Nature* **451**: 959–963.

de Morais MG, Vaz B da S, de Morais EG, Costa JAV. **2015**. Biologically Active Metabolites Synthesized by Microalgae. *BioMed Research International* **2015**: 835761.

Morgenstern B, Atchley WR. **1999**. Evolution of bHLH transcription factors: modular evolution by domain shuffling? *Molecular Biology and Evolution* **16**: 1654–1663.

Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG. **2004**. Genetic analysis of genome-wide variation in human gene expression. *Nature* **430**: 743–747.

Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D. **2009**. Genomic Footprints of a Cryptic Plastid Endosymbiosis in Diatoms. *Science* **324**: 1724–1726.

Msanne J, Xu D, Konda AR, Casas-Mollano JA, Awada T, Cahoon EB, Cerutti H. **2012**. Metabolic and gene expression changes triggered by nitrogen deprivation in the photoautotrophically grown microalgae Chlamydomonas reinhardtii and Coccomyxa sp. C-169. *Phytochemistry* **75**: 50–59.

Nassoury N, Cappadocia M, Morse D. **2003**. Plastid ultrastructure defines the protein import pathway in dinoflagellates. *Journal of Cell Science* **116**: 2867–2874.

Nathan D, Ingvarsdottir K, Sterner DE, Bylebyl GR, Dokmanovic M, Dorsey JA, Whelan KA, Krsmanovic M, Lane WS, Meluh PB, *et al.* 2006. Histone sumoylation is a negative regulator in Saccharomyces cerevisiae and shows dynamic interplay with positive-acting histone modifications. *Genes & Development* **20**: 966–976.

Ng FS, Schütte J, Ruau D, Diamanti E, Hannah R, Kinston SJ, Göttgens B. **2014**. Constrained transcription factor spacing is prevalent and important for transcriptional control of mouse blood cells. *Nucleic Acids Research* **42**: 13513–13524.

Noordally ZB, Millar AJ. **2015**. Clocks in Algae. *Biochemistry* **54**: 171–183.

Nosenko T, Bhattacharya D. **2007**. Horizontal gene transfer in chromalveolates. *BMC Evolutionary Biology* **7**: 173.

Not F, Siano R, Kooistra WHCF, Simon N, Vaulot D, Probert I. **2012**. Diversity and Ecology of Eukaryotic Marine Phytoplankton. Advances in Botanical Research. Elsevier, 1–53.

Nowick K, Stubbs L. **2010**. Lineage-specific transcription factors and the evolution of gene regulatory networks. *Briefings in Functional Genomics* **9**: 65–78.

Nurse P. **2003**. Systems biology: Understanding cells. *Nature* **424**: 883–883.

Oborník M, Modrý D, Lukeš M, Cernotíková-Stříbrná E, Cihlář J, Tesařová M, Kotabová E, Vancová M, Prášil O, Lukeš J. **2012**. Morphology, ultrastructure and life cycle of Vitrella brassicaformis n. sp., n. gen., a novel chromerid from the Great Barrier Reef. *Protist* **163**: 306–323.

Ochman H, Lawrence JG, Groisman EA. **2000**. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299–304.

Ohkawa H, Hashimoto N, Furukawa S, Kadono T, Kawano T. **2011**. Forced symbiosis between Synechocystis spp. PCC 6803 and apo-symbiotic Paramecium bursaria as an experimental model for evolutionary emergence of primitive photosynthetic eukaryotes. *Plant Signaling & Behavior* **6**: 773–776.

Okauchi M. **1990**. Food value of isochrysis aff. galbana for the growth of pearl oyster spat. *Nippon Suisan Gakkaishi* **56**.

Orphanides G, Lagrange T, Reinberg D. **1996**. The general transcription factors of RNA polymerase II. *Genes & Development* **10**: 2657–2683.

Pai AA, Pritchard JK, Gilad Y. **2015**. The Genetic and Mechanistic Basis for Variation in Gene Regulation. *PLOS Genet* **11**: e1004857.

Pani L, Overdier DG, Porcella A, Qian X, Lai E, Costa RH. **1992**. Hepatocyte nuclear factor 3 beta contains two transcriptional activation domains, one of which is novel and conserved with the Drosophila fork head protein. *Molecular and Cellular Biology* **12**: 3723–3732.

Park J, Park J, Jang S, Kim S, Kong S, Choi J, Ahn K, Kim J, Lee S, Kim S, *et al.* 2008. FTFD: an informatics pipeline supporting phylogenomic analysis of fungal transcription factors. *Bioinformatics (Oxford, England)* **24**: 1024–1025.

Parsons M, Karnataki A, Feagin JE, DeRocher A. 2007. Protein Trafficking to the Apicoplast: Deciphering the Apicomplexan Solution to Secondary Endosymbiosis. *Eukaryotic Cell* **6**: 1081–1088.

Paterson AH, Chapman BA, Kissinger JC, Bowers JE, Feltus FA, Estill JC. 2006. Many gene and domain families have convergent fates following independent whole-genome duplication events in Arabidopsis, Oryza, Saccharomyces and Tetraodon. *Trends in genetics: TIG* **22**: 597–602.

Patthy L. 1996. Exon shuffling and other ways of module exchange. *Matrix Biology* **15**: 301–310.

Perez JC, Fordyce PM, Lohse MB, Hanson-Smith V, DeRisi JL, Johnson AD. 2014. How duplicated transcription regulators can diversify to govern the expression of nonoverlapping sets of genes. *Genes & Development* **28**: 1272–1277.

Pérez-Rodríguez P, Riaño-Pachón DM, Corrêa LGG, Rensing SA, Kersten B, Mueller-Roeber B. 2010. PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic acids research* **38**: D822-827.

Petersen J, Ludewig A-K, Michael V, Bunk B, Jarek M, Baurain D, Brinkmann H. 2014. Chromera velia, endosymbioses and the rhodoplex hypothesis--plastid evolution in cryptophytes, alveolates, stramenopiles, and haptophytes (CASH lineages). *Genome Biology and Evolution* **6**: 666–684.

Prahl FG, Wakeham SG. 1987. Calibration of unsaturation patterns in long-chain ketone compositions for palaeotemperature assessment. *Nature* **330**: 367–369.

Provart NJ, Alonso J, Assmann SM, Bergmann D, Brady SM, Brkljacic J, Browse J, Chapple C, Colot V, Cutler S, *et al.* 2016. 50 years of Arabidopsis research: highlights and future directions. *The New Phytologist* **209**: 921–944.

Pugh BF, Tjian R. 1991. Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes & Development* **5**: 1935–1945.

Puranik S, Acajjaoui S, Conn S, Costa L, Conn V, Vial A, Marcellin R, Melzer R, Brown E, Hart D, *et al.* 2014. Structural basis for the oligomerization of the MADS domain transcription factor SEPALLATA3 in Arabidopsis. *The Plant Cell* **26**: 3603–3615.

Qiu Y, Fakas S, Han G-S, Barbosa AD, Siniossoglou S, Carman GM. 2013a. Transcription factor Reb1p regulates DGK1-encoded diacylglycerol kinase and lipid metabolism in Saccharomyces cerevisiae. *The Journal of biological chemistry* **288**: 29124–29133.

Qiu H, Yoon HS, Bhattacharya D. 2013b. Algal endosymbionts as vectors of horizontal gene transfer in photosynthetic eukaryotes. *Plant Physiology* **4**: 366.

Rabara RC, Tripathi P, Rushton PJ. 2014. The potential of transcription factor-based genetic engineering in improving crop tolerance to drought. *Omics: A Journal of Integrative Biology* **18**: 601–614.

Racanelli AC, Turner FB, Xie L-Y, Taylor SM, Moran RG. **2008**. A mouse gene that coordinates epigenetic controls and transcriptional interference to achieve tissue-specific expression. *Molecular and Cellular Biology* **28**: 836–848.

Rayko E, Maumus F, Maheswari U, Jabbari K, Bowler C. **2010**. Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. *New Phytologist* **188**: 52–66.

Read BA, Kegel J, Klute MJ, Kuo A, Lefebvre SC, Maumus F, Mayer C, Miller J, Monier A, Salamov A, *et al.* 2013. Pan genome of the phytoplankton Emiliania underpins its global distribution. *Nature* **499**: 209–213.

Reményi A, Schöler HR, Wilmanns M. **2004**. Combinatorial control of gene expression. *Nature Structural & Molecular Biology* **11**: 812–815.

Renaud SM, Zhou HC, Parry DL, Thinh L-V, Woo KC. Effect of temperature on the growth, total lipid content and fatty acid composition of recently isolated tropical microalgae Isochrysis sp., Nitzschia closterium, Nitzschia paleacea, and commercial species Isochrysis sp. (clone T.ISO). *Journal of Applied Phycology* **7**: 595–602.

Reyes-Prieto A, Moustafa A, Bhattacharya D. **2008**. Multiple genes of apparent algal origin suggest ciliates may once have been photosynthetic. *Current biology: CB* **18**: 956–962.

Rhoads A, Au KF. **2015**. PacBio Sequencing and Its Applications. *Genomics, Proteomics & Bioinformatics* **13**: 278–289.

Richardt S, Lang D, Reski R, Frank W, Rensing SA. **2007**. PlanTAPDB, a phylogeny-based resource of plant transcription-associated proteins. *Plant physiology* **143**: 1452–1466.

Richmond TJ, Finch JT, Rushton B, Rhodes D, Klug A. **1984**. Structure of the nucleosome core particle at 7 A resolution. *Nature* **311**: 532–537.

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, *et al.* 2000. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science (New York, N.Y.)* **290**: 2105–2110.

Roberts SG, Green MR. **1994**. Activator-induced conformational change in general transcription factor TFIIB. *Nature* **371**: 717–720.

Rogato A, Richard H, Sarazin A, Voss B, Cheminant Navarro S, Champeimont R, Navarro L, Carbone A, Hess WR, Falciatore A. **2014**. The diversity of small non-coding RNAs in the diatom Phaeodactylum tricornutum. *BMC genomics* **15**: 698.

Rogers MB, Gilson PR, Su V, McFadden GI, Keeling PJ. **2007**. The complete chloroplast genome of the chlorarachniophyte Bigelowiella natans: evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. *Molecular Biology and Evolution* **24**: 54–62.

Roh T-Y, Cuddapah S, Zhao K. 2005. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes & Development* **19**: 542–552.

Rokitta SD, de Nooijer LJ, Trimborn S, de Vargas C, Rost B, John U. 2011. Transcriptome Analyses Reveal Differential Gene Expression Patterns Between the Life-Cycle Stages of Emiliania Huxleyi (haptophyta) and Reflect Specialization to Different Ecological Niches1. *Journal of Phycology* **47**: 829–838.

Ryckebosch E, Bruneel C, Termote-Verhalle R, Goiris K, Muylaert K, Foubert I. 2014. Nutritional evaluation of microalgae oils rich in omega-3 long chain polyunsaturated fatty acids as an alternative for fish oil. *Food Chemistry* **160**: 393–400.

Sadovskaya I, Souissi A, Souissi S, Grard T, Lencel P, Greene CM, Duin S, Dmitrenok PS, Chizhov AO, Shashkov AS, *et al.* 2014. Chemical structure and biological activity of a highly branched (1 → 3,1 → 6)-β-D-glucan from Isochrysis galbana. *Carbohydrate Polymers* **111**: 139–148.

Sagan L. 1967. On the origin of mitosing cells. *Journal of Theoretical Biology* **14**: 255–274.

Sailsbery JK, Dean RA. 2012. Accurate discrimination of bHLH domains in plants, animals, and fungi using biologically meaningful sites. *BMC evolutionary biology* **12**: 154.

Sakuma S, Pourkheirandish M, Hensel G, Kumlehn J, Stein N, Tagiri A, Yamaji N, Ma JF, Sassa H, Koba T, *et al.* 2013. Divergence of expression pattern contributed to neofunctionalization of duplicated HD-Zip I transcription factor in barley. *The New Phytologist* **197**: 939–948.

Salomon M, Christie JM, Knieb E, Lempert U, Briggs WR. 2000. Photochemical and mutational analysis of the FMN-binding domains of the plant blue light receptor, phototropin. *Biochemistry* **39**: 9401–9410.

Sánchez Á, Maceiras R, Cancela Á, Pérez A. 2013. Culture aspects of Isochrysis galbana for biodiesel production. *Applied Energy* **101**: 192–197.

Sanchez-Puerta MV, Lippmeier JC, Apt KE, Delwiche CF. 2007. Plastid genes in a non-photosynthetic dinoflagellate. *Protist* **158**: 105–117.

Sanz-Luque E, Chamizo-Ampudia A, Llamas A, Galvan A, Fernandez E. 2015. Understanding nitrate assimilation and its regulation in microalgae. *Frontiers in Plant Science* **6**.

Saoudi-Helis L, Dubacq J-P, Marty Y, Samain J-F, Gudin C. 1994. Influence of growth rate on pigment and lipid composition of the microalgaIsochrysis aff.galbana clone T.iso. *Journal of Applied Phycology* **6**: 315–322.

Scaife MA, Smith AG. 2016. Towards developing algal synthetic biology. *Biochemical Society Transactions* **44**: 716–722.

Schadt EE. 2009. Molecular networks as sensors and drivers of common human diseases. *Nature* **461**: 218–223.

Schellenberger Costa B, Sachse M, Jungandreas A, Bartulos CR, Gruber A, Jakob T, Kroth PG, Wilhelm C. 2013. Aureochrome 1a Is Involved in the Photoacclimation of the Diatom Phaeodactylum tricornutum. *PLoS ONE* **8**: e74451.

Schmollinger S, Mühlhaus T, Boyle NR, Blaby IK, Casero D, Mettler T, Moseley JL, Kropat J, Sommer F, Strenkert D, *et al.* 2014. Nitrogen-Sparing Mechanisms in Chlamydomonas Affect the Transcriptome, the Proteome, and Photosynthetic Metabolism. *The Plant Cell* **26**: 1410–1435.

Schönknecht G, Chen W-H, Ternes CM, Barbier GG, Shrestha RP, Stanke M, Bräutigam A, Baker BJ, Banfield JF, Garavito RM, *et al.* 2013. Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science (New York, N.Y.)* **339**: 1207–1210.

Seizl M, Hartmann H, Hoeg F, Kurth F, Martin DE, Söding J, Cramer P. 2011. A conserved GA element in TATA-less RNA polymerase II promoters. *PloS One* **6**: e27595.

Selby CP, Sancar A. 2006. A cryptochrome/photolyase class of enzymes with single-stranded DNA-specific photolyase activity. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 17696–17700.

Serin EAR, Nijveen H, Hilhorst HWM, Ligterink W. 2016. Learning from Co-expression Networks: Possibilities and Challenges. *Frontiers in Plant Science* **7**: 444.

Serna L, Martin C. 2006. Trichomes: different regulatory networks lead to convergent structures. *Trends in Plant Science* **11**: 274–280.

Ševčíková T, Horák A, Klimeš V, Zbránková V, Demir-Hilton E, Sudek S, Jenkins J, Schmutz J, Přibyl P, Fousek J, *et al.* 2015. Updating algal evolutionary relationships through plastid genome sequencing: did alveolate plastids emerge through endosymbiosis of an ochrophyte? *Scientific Reports* **5**: 10134.

Shah MMR, Liang Y, Cheng JJ, Daroch M. 2016. Astaxanthin-Producing Green Microalga Haematococcus pluvialis: From Single Cell to High Value Commercial Products. *Plant Biotechnology*: 531.

Sharma N, Bhalla PL, Singh MB. 2013. Transcriptome-wide profiling and expression analysis of transcription factor families in a liverwort, Marchantia polymorpha. *BMC Genomics* **14**: 915.

Sharon I, Alperovitch A, Rohwer F, Haynes M, Glaser F, Atamna-Ismaeel N, Pinter RY, Partensky F, Koonin EV, Wolf YI, *et al.* 2009. Photosystem I gene cassettes are present in marine virus genomes. *Nature* **461**: 258–262.

Shiraiwa Y. 2003. Physiological regulation of carbon fixation in the photosynthesis and calcification of coccolithophorids. *Comparative Biochemistry and Physiology. Part B, Biochemistry & Molecular Biology* **136**: 775–783.

Shiu S-H, Shih M-C, Li W-H. 2005. Transcription Factor Families Have Much Higher Expansion Rates in Plants than in Animals. *Plant Physiology* **139**: 18–26.

Short SM. **2012**. The ecology of viruses that infect eukaryotic algae. *Environmental Microbiology* **14**: 2253–2271.

Sialve B, Bernet N, Bernard O. **2009**. Anaerobic digestion of microalgae as a necessary step to make microalgal biodiesel sustainable. *Biotechnology Advances* **27**: 409–416.

Sikorski TW, Buratowski S. **2009**. The basal initiation machinery: beyond the general transcription factors. *Current Opinion in Cell Biology* **21**: 344–351.

da Silva AF, Lourenço SO, Chaloub RM. **2009**. Effects of nitrogen starvation on the photosynthetic physiology of a tropical marine microalga Rhodomonas sp. (Cryptophyceae). *Aquatic Botany* **91**: 291–297.

Slattery M, Zhou T, Yang L, Dantas Machado AC, Gordân R, Rohs R. **2014**. Absence of a simple code: how transcription factors read the genome. *Trends in Biochemical Sciences* **39**: 381–399.

Solovchenko AE. **2012**. Physiological role of neutral lipid accumulation in eukaryotic microalgae under stresses. *Russian Journal of Plant Physiology* **59**: 167–176.

Song B, Ward BB. **2007**. MOLECULAR CLONING AND CHARACTERIZATION OF HIGH-AFFINITY NITRATE TRANSPORTERS IN MARINE PHYTOPLANKTON. *Journal of Phycology* **43**: 542–552.

Sørensen MES, Cameron DD, Brockhurst MA, Wood AJ. **2016**. Metabolic constraints for a novel symbiosis. *Royal Society Open Science* **3**: 150708.

Soufi A, Garcia MF, Jaroszewicz A, Osman N, Pellegrini M, Zaret KS. **2015**. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**: 555–568.

Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA. **2005**. Interchromosomal associations between alternatively expressed loci. *Nature* **435**: 637–645.

Stiller JW, Schreiber J, Yue J, Guo H, Ding Q, Huang J. **2014**. The evolution of photosynthesis in chromist algae through serial endosymbioses. *Nature Communications* **5**.

St-Pierre F. **2009**. Mécanismes de régulation transcriptionnelle du gène de l'alpha-foetoprotéine.

Sulli C, Fang Z, Muchhal U, Schwartzbach SD. **1999**. Topology of Euglena chloroplast protein precursors within endoplasmic reticulum to Golgi to chloroplast transport vesicles. *The Journal of Biological Chemistry* **274**: 457–463.

Sutherland JA, Cook A, Bannister AJ, Kouzarides T. **1992**. Conserved motifs in Fos and Jun define a new class of activation domain. *Genes & development* **6**: 1810–1819.

Suzuki S, Ishida K-I, Hirakawa Y. 2016. Diurnal Transcriptional Regulation of Endosymbiotically Derived Genes in the Chlorarachniophyte Bigelowiella natans. *Genome Biology and Evolution* **8**: 2672–2682.

Suzuki S, Shirato S, Hirakawa Y, Ishida K. 2015. Nucleomorph genome sequences of two chlorarachniophytes, Amorphochlora amoebiformis and Lotharella vacuolata. *Genome Biology and Evolution*: evv096.

Taelman V, Van Wayenbergh R, Sölter M, Pichon B, Pieler T, Christophe D, Bellefroid EJ. 2004. Sequences downstream of the bHLH domain of the Xenopus hairy-related transcription factor-1 act as an extended dimerization domain that contributes to the selection of the partners. *Developmental Biology* **276**: 47–63.

Takahashi F, Yamagata D, Ishikawa M, Fukamatsu Y, Ogura Y, Kasahara M, Kiyosue T, Kikuyama M, Wada M, Kataoka H. 2007. AUREOCHROME, a photoreceptor required for photomorphogenesis in stramenopiles. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 19625–19630.

Taylor-Teeples M, Lin L, de Lucas M, Turco G, Toal TW, Gaudinier A, Young NF, Trabucco GM, Veling MT, Lamothe R, *et al.* 2015. An Arabidopsis gene regulatory network for secondary cell wall synthesis. *Nature* **517**: 571–575.

Thiriet-Rupert S, Carrier G, Chénais B, Trottier C, Bougaran G, Cadoret J-P, Schoefs B, Saint-Jean B. 2016. Transcription factors in microalgae: genome-wide prediction and comparative analysis. *BMC genomics* **17**: 282.

Tillmann U. 1998. Phagotrophy by a plastidic haptophyte, Prymnesium patelliferum. *Aquatic Microbial Ecology* **14**: 155–160.

Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics* **5**: 123–135.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics (Oxford, England)* **25**: 1105–1111.

Trentacoste EM, Shrestha RP, Smith SR, Glé C, Hartmann AC, Hildebrand M, Gerwick WH. 2013. Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 19748–19753.

Tyler BM, Tripathy S, Zhang X, Dehal P, Jiang RHY, Aerts A, Arredondo FD, Baxter L, Bensasson D, Beynon JL, *et al.* 2006. Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science (New York, N.Y.)* **313**: 1261–1266.

Valenzuela J, Mazurie A, Carlson RP, Gerlach R, Cooksey KE, Peyton BM, Fields MW. 2012. Potential role of multiple carbon fixation pathways during lipid accumulation in Phaeodactylum tricornutum. *Biotechnology for Biofuels* **5**: 40.

Valledor L, Furuhashi T, Recuenco-Muñoz L, Wienkoop S, Weckwerth W. **2014**. System-level network analysis of nitrogen starvation and recovery in Chlamydomonas reinhardtii reveals potential new targets for increased lipid accumulation. *Biotechnology for Biofuels* **7**: 171.

Van de Peer Y, Van der Auwera G, De Wachter R. **1996**. The evolution of stramenopiles and alveolates as derived by 'substitution rate calibration' of small ribosomal subunit RNA. *Journal of Molecular Evolution* **42**: 201–210.

Vandermeers F, Neelature Sriramareddy S, Costa C, Hubaux R, Cosse J-P, Willems L. **2013**. The role of epigenetics in malignant pleural mesothelioma. *Lung Cancer (Amsterdam, Netherlands)* **81**: 311–318.

de Vargas C, Aubry MP, Probert I, Young J. **2007**. The origin and evolution of coccolithophores: from coastal hunters to oceanic farmers.

de Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, Lara E, Berney C, Le Bescot N, Probert I, *et al.* 2015. Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. *Science (New York, N.Y.)* **348**: 1261605.

Verrijzer CP. **2001**. Transcription Factor IID--Not So Basal After All. *Science* **293**: 2010–2011.

Vieler A, Wu G, Tsai C-H, Bullard B, Cornish AJ, Harvey C, Reca I-B, 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.

Vogel C, Marcotte EM. **2012**. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews. Genetics* **13**: 227–232.

Voinnet O. **2009**. Origin, biogenesis, and activity of plant microRNAs. *Cell* **136**: 669–687.

Vom Endt D, Kijne JW, Memelink J. **2002**. Transcription factors controlling plant secondary metabolism: what regulates the regulators? *Phytochemistry* **61**: 107–114.

Walker TL, Collet C, Purton S. **2005**. ALGAL TRANSGENICS IN THE GENOMIC ERA1: GENETIC ENGINEERING OF ALGAE. *Journal of Phycology* **41**: 1077–1093.

Wang YXR, Huang H. **2014**. Review on statistical methods for gene network reconstruction using expression data. *Journal of Theoretical Biology* **362**: 53–61.

Wang X, Tomso DJ, Liu X, Bell DA. **2005**. Single nucleotide polymorphism in transcriptional regulatory regions and expression of environmentally responsive genes. *Toxicology and Applied Pharmacology* **207**: 84–90.

Wang H-T, Yao C-H, Ai J-N, Cao X-P, Xue S, Wang W. **2014**. Identification of carbohydrates as the major carbon sink of the marine microalga Isochrysis zhangjiangensis (Haptophyta) and optimization of its productivity by nitrogen manipulation. *Bioresource Technology* **171**: 298–304.

Wang H, Zhai L, Xu J, Joo H-Y, Jackson S, Erdjument-Bromage H, Tempst P, Xiong Y, Zhang Y. 2006. Histone H3 and H4 ubiquitylation by the CUL4-DDB-ROC1 ubiquitin ligase facilitates cellular response to DNA damage. *Molecular Cell* **22**: 383–394.

Wase N, Black PN, Stanley BA, DiRusso CC. 2014. Integrated quantitative analysis of nitrogen stress response in Chlamydomonas reinhardtii using metabolite and protein profiling. *Journal of Proteome Research* **13**: 1373–1396.

Wasserman WW, Sandelin A. 2004. Applied bioinformatics for the identification of regulatory elements. *Nature Reviews Genetics* **5**: 276–287.

Watts JA, Zhang C, Klein-Szanto AJ, Kormish JD, Fu J, Zhang MQ, Zaret KS. 2011. Study of FoxA Pioneer Factor at Silent Genes Reveals Rfx-Repressed Enhancer at Cdx2 and a Potential Indicator of Esophageal Adenocarcinoma Development. *PLOS Genet* **7**: e1002277.

Wendel JF. 2000. Genome evolution in polyploids. *Plant Molecular Biology* **42**: 225–249.

Wenkel S, Turck F, Singer K, Gissot L, Gourrierec JL, Samach A, Coupland G. 2006. CONSTANS and the CCAAT Box Binding Complex Share a Functionally Important Domain and Interact to Regulate Flowering of Arabidopsis. *The Plant Cell Online* **18**: 2971–2984.

Wilson RJ, Denny PW, Preiser PR, Rangachari K, Roberts K, Roy A, Whyte A, Strath M, Moore DJ, Moore PW, *et al.* 1996. Complete gene map of the plastid-like DNA of the malaria parasite Plasmodium falciparum. *Journal of Molecular Biology* **261**: 155–172.

Woehle C, Dagan T, Martin WF, Gould SB. 2011. Red and problematic green phylogenetic signals among thousands of nuclear genes from the photosynthetic and apicomplexa-related Chromera velia. *Genome Biology and Evolution* **3**: 1220–1230.

Wu J, Zhao F, Wang S, Deng G, Wang J, Bai J, Lu J, Qu J, Bao Q. 2007. cTFbase: a database for comparative genomics of transcription factors in cyanobacteria. *BMC genomics* **8**: 104.

Wunderlich Z, Mirny LA. 2009. Different gene regulation strategies revealed by analysis of binding motifs. *Trends in genetics: TIG* **25**: 434–440.

Wurch LL, Bertrand EM, Saito MA, Mooy BASV, Dyhrman ST. 2011. Proteome Changes Driven by Phosphorus Deficiency and Recovery in the Brown Tide-Forming Alga Aureococcus anophagefferens. *PLOS ONE* **6**: e28949.

Xia LC, Ai D, Cram J, Fuhrman JA, Sun F. 2013. Efficient statistical significance approximation for local similarity analysis of high-throughput time series data. *Bioinformatics (Oxford, England)* **29**: 230–237.

Xia LC, Steele JA, Cram JA, Cardon ZG, Simmons SL, Vallino JJ, Fuhrman JA, Sun F. 2011. Extended local similarity analysis (eLSA) of microbial community and other time series data with replicates. *BMC systems biology* **5 Suppl 2**: S15.

Xue J, Niu Y-F, Huang T, Yang W-D, Liu J-S, Li H-Y. **2015**. Genetic improvement of the microalga Phaeodactylum tricornutum for boosting neutral lipid accumulation. *Metabolic Engineering* **27**: 1–9.

Yaakob Z, Ali E, Zainal A, Mohamad M, Takriff MS. **2014**. An overview: biomolecules from microalgae for animal feed and aquaculture. *Journal of Biological Research (Thessalonikē, Greece)* **21**: 6.

Yamamoto A, Kagaya Y, Toyoshima R, Kagaya M, Takeda S, Hattori T. **2009a**. Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. *The Plant Journal* **58**: 843–856.

Yamamoto N, Takemori Y, Sakurai M, Sugiyama K, Sakurai H. **2009b**. Differential recognition of heat shock elements by members of the heat shock transcription factor family. *The FEBS journal* **276**: 1962–1974.

Yang C, Bolotin E, Jiang T, Sladek FM, Martinez E. **2007**. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* **389**: 52–65.

Yang Y, Matsuzaki M, Takahashi F, Qu L, Nozaki H. **2014**. Phylogenomic analysis of 'red' genes from two divergent species of the 'green' secondary phototrophs, the chlorarachniophytes, suggests multiple horizontal gene transfers from the red lineage before the divergence of extant chlorarachniophytes. *PloS One* **9**: e101158.

Yang T-H, Wu W-S. **2012**. Identifying biologically interpretable transcription factor knockout targets by jointly analyzing the transcription factor knockout microarray and the ChIP-chip data. *BMC systems biology* **6**: 102.

Yilancioglu K, Cokol M, Pastirmaci I, Erman B, Cetiner S. **2014**. Oxidative Stress Is a Mediator for Increased Lipid Accumulation in a Newly Isolated Dunaliella salina Strain. *PLoS ONE* **9**.

Yu W-L, Ansari W, Schoepp NG, Hannon MJ, Mayfield SP, Burkart MD. **2011**. Modifications of the metabolic pathways of lipid and triacylglycerol production in microalgae. *Microbial cell factories* **10**: 91.

Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. **2007**. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS computational biology* **3**: e59.

Yuan G-C, Liu Y-J, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ. **2005**. Genome-scale identification of nucleosome positions in S. cerevisiae. *Science (New York, N.Y.)* **309**: 626–630.

Zabet NR, Adryan B. **2013**. The effects of transcription factor competition on gene regulation. *Frontiers in Genetics* **4**: 197.

Zaret KS, Carroll JS. **2011**. Pioneer transcription factors: establishing competence for gene expression. *Genes & Development* **25**: 2227–2241.

Zaret KS, Mango SE. **2016**. Pioneer transcription factors, chromatin dynamics, and cell fate control. *Current Opinion in Genetics & Development* **37**: 76–81.

Zekavati A, Nasir A, Alcaraz A, Aldrovandi M, Marsh P, Norton JD, Murphy JJ. **2014**. Post-transcriptional regulation of BCL2 mRNA by the RNA-binding protein ZFP36L1 in malignant B cells. *PloS One* **9**: e102625.

Zhang Y-M, Chen H, He C-L, Wang Q. **2013**. Nitrogen starvation induced oxidative stress in an oil-producing green alga Chlorella sorokiniana C3. *PloS One* **8**: e69225.

Zhang H-M, Chen H, Liu W, Liu H, Gong J, Wang H, Guo A-Y. **2012**. AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic acids research* **40**: D144-149.

Zhang Z, Green BR, Cavalier-Smith T. **1999**. Single gene circles in dinoflagellate chloroplast genomes. *Nature* **400**: 155–159.

Zhang J, Hao Q, Bai L, Xu J, Yin W, Song L, Xu L, Guo X, Fan C, Chen Y, *et al.* 2014a. Overexpression of the soybean transcription factor GmDof4 significantly enhances the lipid content of Chlorella ellipsoidea. *Biotechnology for Biofuels* **7**: 128.

Zhang L, Wang Y, Sun M, Wang J, Kawabata S, Li Y. **2014b**. BrMYB4, a suppressor of genes for phenylpropanoid and anthocyanin biosynthesis, is downregulated by UV-B but not by pigment-inducing sunlight in turnip cv. Tsuda. *Plant & Cell Physiology*.

Zhang T, Zhang W, Jiang J. **2015**. Genome-Wide Nucleosome Occupancy and Positioning and Their Impact on Gene Expression and Evolution in Plants1[OPEN]. *Plant Physiology* **168**: 1406–1416.

Zhao S, Zhang Y, Gordon W, Quan J, Xi H, Du S, von Schack D, Zhang B. **2015**. Comparison of stranded and non-stranded RNA-seq transcriptome profiling and investigation of gene overlap. *BMC Genomics* **16**: 675.

Zhu B-H, Shi H-P, Yang G-P, Lv N-N, Yang M, Pan K-H. **2016**. Silencing UDP-glucose pyrophosphorylase gene in Phaeodactylum tricornutum affects carbon allocation. *New Biotechnology* **33**: 237–244.

Zhu B, Zheng Y, Pham A-D, Mandal SS, Erdjument-Bromage H, Tempst P, Reinberg D. **2005**. Monoubiquitination of human histone H2B: the factors involved and their roles in HOX gene regulation. *Molecular Cell* **20**: 601–611.

Zolfaghari Emameh R, Barker HR, Tolvanen MEE, Parkkila S, Hytönen VP. **2016**. Horizontal transfer of β-carbonic anhydrase genes from prokaryotes to protozoans, insects, and nematodes. *Parasites & Vectors* **9**: 152.

Annexes

Annexe A : enrichissement en fonctions GO des trois modules d'intérêt identifiés par WGCNA

Module WGCNA	GO-ID	GO Term	Category	P-Value
module lié à la quantité de lipides de réserve	GO:0004842	ubiquitin-protein transferase activity	F	2.697988E-3
	GO:0005779	integral component of peroxisomal membrane	C	3.453023E-3
	GO:0016558	protein import into peroxisome matrix	P	1.899324E-2
	GO:0008022	protein C-terminus binding	F	1.899324E-2
	GO:0005506	iron ion binding	F	1.482247E-2
	GO:0007094	mitotic spindle assembly checkpoint	P	1.899324E-2
	GO:0004150	dihydroneopterin aldolase activity	F	1.899324E-2
	GO:0006457	protein folding	P	2.000786E-2
	GO:0051087	chaperone binding	F	3.613894E-2
	GO:0000139	Golgi membrane	C	2.048355E-2
	GO:0003950	NAD+ ADP-ribosyltransferase activity	F	2.437749E-2
	GO:0006561	proline biosynthetic process	P	2.809057E-2
	GO:0004834	tryptophan synthase activity	F	3.762791E-2
	GO:0009765	photosynthesis, light harvesting	P	3.142191E-2
	GO:0004045	aminoacyl-tRNA hydrolase activity	F	3.762791E-2
	GO:0004408	holocytochrome-c synthase activity	F	3.762791E-2
	GO:0004484	mRNA guanylyltransferase activity	F	3.762791E-2
	GO:0006370	7-methylguanosine mRNA capping	P	3.762791E-2
	GO:0042586	peptide deformylase activity	F	3.762791E-2
	GO:0004499	N,N-dimethylaniline monooxygenase activity	F	3.762791E-2
1er module lié à la quantité de carbohydrates	GO:0006511	ubiquitin-dependent protein catabolic process	P	2.556185E-4
	GO:0034450	ubiquitin-ubiquitin ligase activity	F	1.668757E-2
	GO:0016567	protein ubiquitination	P	1.894726E-2
	GO:0000151	ubiquitin ligase complex	C	2.218926E-2
	GO:0031625	ubiquitin protein ligase binding	F	2.218926E-2
	GO:0019773	proteasome core complex, alpha-subunit complex	C	4.924876E-2
	GO:0004222	metalloendopeptidase activity	F	2.822471E-2
	GO:0006515	misfolded or incompletely synthesized protein catabolic process	P	1.115557E-2
	GO:0004252	serine-type endopeptidase activity	F	4.282976E-2
	GO:0045300	acyl-[acyl-carrier-protein] desaturase activity	F	5.593102E-3
	GO:0070985	TFIIK complex	C	5.593102E-3
	GO:0016538	cyclin-dependent protein serine/threonine kinase regulator activity	F	5.593102E-3
	GO:0016992	lipoate synthase activity	F	1.115557E-2
	GO:0009107	lipoate biosynthetic process	P	2.218926E-2
	GO:0046835	carbohydrate phosphorylation	P	3.310236E-2
	GO:0006012	galactose metabolic process	P	4.924876E-2
	GO:0004335	galactokinase activity	F	2.218926E-2
	GO:0007034	vacuolar transport	P	2.76608E-2
2ème module lié à la quantité de carbohydrates	GO:0042555	MCM complex	C	4.361613E-6
	GO:0006270	DNA replication initiation	P	6.788884E-6
	GO:0003678	DNA helicase activity	F	2.27774E-4
	GO:0003688	DNA replication origin binding	F	2.455173E-2
	GO:0000808	origin recognition complex	C	3.660272E-2
	GO:0070481	nuclear-transcribed mRNA catabolic process, non-stop decay	P	1.235143E-2
	GO:0070966	nuclear-transcribed mRNA catabolic process, no-go decay	P	1.235143E-2
	GO:0071025	RNA surveillance	P	1.235143E-2
	GO:0051499	D-aminoacyl-tRNA deacylase activity	F	1.235143E-2
	GO:0006788	heme oxidation	P	1.235143E-2
	GO:0004392	heme oxygenase (decyclizing) activity	F	1.235143E-2
	GO:0004852	uroporphyrinogen-III synthase activity	F	2.455173E-2
	GO:0005096	GTPase activator activity	F	1.428923E-2
	GO:0004668	protein-arginine deiminase activity	F	2.455173E-2
	GO:0009446	putrescine biosynthetic process	P	2.455173E-2
	GO:0034551	mitochondrial respiratory chain complex III assembly	P	2.455173E-2
	GO:0004451	isocitrate lyase activity	F	2.455173E-2
	GO:0004563	beta-N-acetylhexosaminidase activity	F	2.455173E-2
	GO:0015204	urea transmembrane transporter activity	F	2.455173E-2
	GO:0071918	urea transmembrane transport	P	2.455173E-2
	GO:0043087	regulation of GTPase activity	P	2.455173E-2
	GO:0003868	4-hydroxyphenylpyruvate dioxygenase activity	F	2.455173E-2
	GO:0003779	actin binding	F	3.769563E-2
	GO:0006914	autophagy	P	4.850622E-2

Annexe B : enrichissement en fonctions GO des gènes priorisés grâce aux données générées par WGCNA

GO-ID	GO Term	Category	P-Value
GO:0051087	chaperone binding	F	9,12E-04
GO:0001671	ATPase activator activity	F	4,03E-03
GO:0006788	heme oxidation	P	1,24E-02
GO:0004392	heme oxygenase (decyclizing) activity	F	1,24E-02
GO:0051499	D-aminoacyl-tRNA deacylase activity	F	1,24E-02
GO:0000172	ribonuclease MRP complex	C	1,24E-02
GO:0030677	ribonuclease P complex	C	1,24E-02
GO:0006379	mRNA cleavage	P	2,46E-02
GO:0006370	7-methylguanosine mRNA capping	P	2,46E-02
GO:0004484	mRNA guanylyltransferase activity	F	2,46E-02
GO:0004668	protein-arginine deiminase activity	F	2,46E-02
GO:0009446	putrescine biosynthetic process	P	2,46E-02
GO:0004563	beta-N-acetylhexosaminidase activity	F	2,46E-02
GO:0051536	iron-sulfur cluster binding	F	4,97E-02
GO:0016992	lipoate synthase activity	F	2,46E-02
GO:0009107	lipoate biosynthetic process	P	4,85E-02
GO:0006511	ubiquitin-dependent protein catabolic process	P	3,02E-02
GO:0034450	ubiquitin-ubiquitin ligase activity	F	3,66E-02
GO:0000151	ubiquitin ligase complex	C	4,85E-02
GO:0046914	transition metal ion binding	F	4,79E-02
GO:0035639	purine ribonucleoside triphosphate binding	F	4,90E-02

Annexe C : enrichissement en fonctions GO des quatre communautés du réseau de régulation de la souche WTc1

Communauté	GO-ID	GO Term	Category	P-Value
communauté violette	GO:0016763	transferase activity, transferring pentosyl groups	F	2,11E-03
	GO:0016757	transferase activity, transferring glycosyl groups	F	3,81E-02
	GO:0003950	NAD+ ADP-ribosyltransferase activity	F	2,11E-03
	GO:0016810	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds	F	4,78E-03
	GO:0030904	retromer complex	C	5,36E-03
	GO:0042147	retrograde transport, endosome to Golgi	P	1,07E-02
	GO:0016197	endosomal transport	P	1,07E-02
	GO:0046914	transition metal ion binding	F	9,97E-03
	GO:0008270	zinc ion binding	F	1,91E-02
	GO:0004476	mannose-6-phosphate isomerase activity	F	5,36E-03
	GO:0009298	GDP-mannose biosynthetic process	P	1,60E-02
	GO:0009226	nucleotide-sugar biosynthetic process	P	2,13E-02
	GO:0019673	GDP-mannose metabolic process	P	2,65E-02
	GO:0009225	nucleotide-sugar metabolic process	P	4,72E-02
	GO:0055114	oxidation-reduction process	P	3,93E-02
	GO:0005956	protein kinase CK2 complex	C	5,36E-03
	GO:0019207	kinase regulator activity	F	1,60E-02
	GO:0019887	protein kinase regulator activity	F	1,60E-02
	GO:0016708	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of two atoms of oxygen into one donor	F	5,36E-03
	GO:0009055	electron carrier activity	F	2,82E-02
	GO:0009767	photosynthetic electron transport chain	P	1,07E-02
	GO:0009772	photosynthetic electron transport in photosystem II	P	1,07E-02
	GO:0045156	electron transporter, transferring electrons within the cyclic electron transport pathway of photosynthesis activity	F	1,07E-02
	GO:0008131	primary amine oxidase activity	F	1,07E-02
	GO:0016641	oxidoreductase activity, acting on the CH-NH2 group of donors, oxygen as acceptor	F	1,60E-02
	GO:0048038	quinone binding	F	3,17E-02
	GO:0005507	copper ion binding	F	4,72E-02
	GO:0016811	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amides	F	1,26E-02
	GO:0016851	magnesium chelatase activity	F	2,65E-02
	GO:0051002	ligase activity, forming nitrogen-metal bonds	F	2,65E-02
	GO:0051003	ligase activity, forming nitrogen-metal bonds, forming coordination complexes	F	2,65E-02
	GO:0022834	ligand-gated channel activity	F	4,72E-02
	GO:0015276	ligand-gated ion channel activity	F	4,72E-02
	GO:0005230	extracellular ligand-gated ion channel activity	F	3,69E-02
	GO:0030532	small nuclear ribonucleoprotein complex	C	3,69E-02
	GO:0022610	biological adhesion	P	4,21E-02
	GO:0007155	cell adhesion	P	4,21E-02
communauté bleue	GO:0046983	protein dimerization activity	F	8,53E-04
	GO:0004022	alcohol dehydrogenase (NAD) activity	F	5,35E-03
	GO:0008774	acetaldehyde dehydrogenase (acetylating) activity	F	5,35E-03
	GO:0006066	alcohol metabolic process	P	1,60E-02
	GO:0015976	carbon utilization	P	2,12E-02
	GO:0003983	UTP:glucose-1-phosphate uridylyltransferase activity	F	5,35E-03
	GO:0051748	UTP-monosaccharide-1-phosphate uridylyltransferase activity	F	5,35E-03
	GO:0006011	UDP-glucose metabolic process	P	5,35E-03
	GO:0009225	nucleotide-sugar metabolic process	P	2,39E-02
	GO:0070569	uridylyltransferase activity	F	3,43E-02
	GO:0006606	protein import into nucleus	P	8,02E-03
	GO:1902593	single-organism nuclear import	P	8,02E-03
	GO:0051170	nuclear import	P	8,02E-03
	GO:0044744	protein targeting to nucleus	P	8,02E-03
	GO:0034504	protein localization to nucleus	P	8,02E-03
	GO:0017038	protein import	P	2,65E-02

	GO:0008565	protein transporter activity	F	3,69E-02
	GO:0072593	reactive oxygen species metabolic process	P	8,02E-03
	GO:0006801	superoxide metabolic process	P	8,02E-03
	GO:0004784	superoxide dismutase activity	F	1,33E-02
	GO:0016721	oxidoreductase activity, acting on superoxide radicals as acceptor	F	1,33E-02
	GO:0035556	intracellular signal transduction	P	2,95E-02
	GO:0048015	phosphatidylinositol-mediated signaling	P	1,33E-02
	GO:0048017	inositol lipid-mediated signaling	P	1,33E-02
	GO:0046854	phosphatidylinositol phosphorylation	P	3,95E-02
	GO:0046834	lipid phosphorylation	P	3,95E-02
	GO:0000160	phosphorelay signal transduction system	P	4,98E-02
	GO:0002161	aminoacyl-tRNA editing activity	F	3,69E-02
communauté rouge	GO:0016709	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen	F	2,56E-03
	GO:0004499	N,N-dimethylaniline monooxygenase activity	F	2,56E-03
	GO:0004497	monooxygenase activity	F	1,65E-02
	GO:0050661	NADP binding	F	3,53E-02
	GO:0015696	ammonium transport	P	5,12E-03
	GO:0008519	ammonium transmembrane transporter activity	F	5,12E-03
	GO:0072488	ammonium transmembrane transport	P	5,12E-03
	GO:0071705	nitrogen compound transport	P	2,16E-02
	GO:0004198	calcium-dependent cysteine-type endopeptidase activity	F	2,41E-02
	GO:0004197	cysteine-type endopeptidase activity	F	2,54E-02
	GO:0008234	cysteine-type peptidase activity	F	4,64E-02
	GO:0016567	protein ubiquitination	P	4,77E-02
	GO:0046983	protein dimerization activity	F	3,54E-03
	GO:0046982	protein heterodimerization activity	F	2,66E-02
communauté verte	GO:0016624	oxidoreductase activity, acting on the aldehyde or oxo group of donors, disulfide as acceptor	F	1,78E-02
	GO:0003950	NAD+ ADP-ribosyltransferase activity	F	3,28E-02
	GO:0004177	aminopeptidase activity	F	3,53E-02
	GO:0006468	protein phosphorylation	P	3,75E-02
	GO:0004672	protein kinase activity	F	3,95E-02

Annexe D : enrichissement en fonctions GO des quatre communautés du réseau de régulation de la souche 2Xc1

commnaté	GO-ID	GO Term	Category	P-Value
communauté violette	GO:0016763	transferase activity, transferring pentosyl groups	F	2,11E-03
	GO:0016757	transferase activity, transferring glycosyl groups	F	3,81E-02
	GO:0003950	NAD+ ADP-ribosyltransferase activity	F	2,11E-03
	GO:0016810	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds	F	4,78E-03
	GO:0016811	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amides	F	1,26E-02
	GO:0030904	retromer complex	C	5,36E-03
	GO:0042147	retrograde transport, endosome to Golgi	P	1,07E-02
	GO:0016197	endosomal transport	P	1,07E-02
	GO:0005956	protein kinase CK2 complex	C	5,36E-03
	GO:0019207	kinase regulator activity	F	1,60E-02
	GO:0019887	protein kinase regulator activity	F	1,60E-02
	GO:0046914	transition metal ion binding	F	9,97E-03
	GO:0008270	zinc ion binding	F	1,91E-02
	GO:0004476	mannose-6-phosphate isomerase activity	F	5,36E-03
	GO:0009225	nucleotide-sugar metabolic process	P	4,72E-02
	GO:0009298	GDP-mannose biosynthetic process	P	1,60E-02
	GO:0009226	nucleotide-sugar biosynthetic process	P	2,13E-02
	GO:0019673	GDP-mannose metabolic process	P	2,65E-02
	GO:0009055	electron carrier activity	F	2,82E-02
	GO:0009767	photosynthetic electron transport chain	P	1,07E-02
	GO:0009772	photosynthetic electron transport in photosystem II	P	1,07E-02
	GO:0045156	electron transporter, transferring electrons within the cyclic electron transport pathway of photosynthesis activity	F	1,07E-02
	GO:0048038	quinone binding	F	3,17E-02
	GO:0016641	oxidoreductase activity, acting on the CH-NH2 group of donors, oxygen as acceptor	F	1,60E-02
	GO:0005507	copper ion binding	F	4,72E-02
	GO:0008131	primary amine oxidase activity	F	1,07E-02
	GO:0016851	magnesium chelatase activity	F	2,65E-02
	GO:0051002	ligase activity, forming nitrogen-metal bonds	F	2,65E-02
	GO:0051003	ligase activity, forming nitrogen-metal bonds, forming coordination complexes	F	2,65E-02
	GO:0055114	oxidation-reduction process	P	3,93E-02
	GO:0016708	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of two atoms of oxygen into one donor	F	5,36E-03
	GO:0015276	ligand-gated ion channel activity	F	4,72E-02
	GO:0005230	extracellular ligand-gated ion channel activity	F	3,69E-02
	GO:0022834	ligand-gated channel activity	F	4,72E-02
	GO:0030532	small nuclear ribonucleoprotein complex	C	3,69E-02
	GO:0022610	biological adhesion	P	4,21E-02
	GO:0007155	cell adhesion	P	4,21E-02
communauté bleue	GO:0022892	substrate-specific transporter activity	F	1,55E-02
	GO:0005215	transporter activity	F	3,57E-02
	GO:0072509	divalent inorganic cation transmembrane transporter activity	F	4,34E-02
	GO:0072511	divalent inorganic cation transport	P	4,34E-02
	GO:0070838	divalent metal ion transport	P	4,34E-02
	GO:0015693	magnesium ion transport	P	2,99E-02
	GO:0015095	magnesium ion transmembrane transporter activity	F	2,99E-02
	GO:0008565	protein transporter activity	F	4,62E-04
	GO:1902582	single-organism intracellular transport	P	1,36E-02
	GO:0008104	protein localization	P	4,81E-02
	GO:0015031	protein transport	P	4,36E-02
	GO:0045184	establishment of protein localization	P	4,69E-02
	GO:0046907	intracellular transport	P	4,75E-02
	GO:0030904	retromer complex	C	2,33E-03
	GO:0042147	retrograde transport, endosome to Golgi	P	4,66E-03
	GO:0016197	endosomal transport	P	4,66E-03
	GO:0017038	protein import	P	2,31E-02
	GO:0006606	protein import into nucleus	P	6,98E-03
	GO:1902593	single-organism nuclear import	P	6,98E-03
	GO:0051170	nuclear import	P	6,98E-03
	GO:0044744	protein targeting to nucleus	P	6,98E-03
	GO:0034504	protein localization to nucleus	P	6,98E-03
	GO:0046486	glycerolipid metabolic process	P	5,02E-03
	GO:0044255	cellular lipid metabolic process	P	4,58E-02
	GO:0006641	triglyceride metabolic process	P	2,33E-03
	GO:0006638	neutral lipid metabolic process	P	2,33E-03
	GO:0006639	acylglycerol metabolic process	P	2,33E-03
	GO:0004144	diacylglycerol O-acyltransferase activity	F	2,33E-03
	GO:0046460	neutral lipid biosynthetic process	P	2,33E-03
	GO:0046463	acylglycerol biosynthetic process	P	2,33E-03
	GO:0019432	triglyceride biosynthetic process	P	2,33E-03
	GO:0016411	acylglycerol O-acyltransferase activity	F	2,33E-03
	GO:0045017	glycerolipid biosynthetic process	P	2,08E-02
	GO:0008374	O-acyltransferase activity	F	2,08E-02

	GO:0048015	phosphatidylinositol-mediated signaling	P	1,16E-02
	GO:0048017	inositol lipid-mediated signaling	P	1,16E-02
	GO:0046854	phosphatidylinositol phosphorylation	P	3,44E-02
	GO:0046834	lipid phosphorylation	P	3,44E-02
	GO:1901135	carbohydrate derivative metabolic process	P	1,13E-02
	GO:0005975	carbohydrate metabolic process	P	3,37E-02
	GO:0044723	single-organism carbohydrate metabolic process	P	4,81E-02
	GO:1901137	carbohydrate derivative biosynthetic process	P	2,57E-02
	GO:0004476	mannose-6-phosphate isomerase activity	F	2,33E-03
	GO:0009225	nucleotide-sugar metabolic process	P	2,08E-02
	GO:0009298	GDP-mannose biosynthetic process	P	6,98E-03
	GO:0009226	nucleotide-sugar biosynthetic process	P	9,29E-03
	GO:0019673	GDP-mannose metabolic process	P	1,16E-02
	GO:0016861	intramolecular oxidoreductase activity, interconverting aldoses and ketoses	F	3,22E-02
	GO:0016860	intramolecular oxidoreductase activity	F	3,89E-02
	GO:0003980	UDP-glucose:glycoprotein glucosyltransferase activity	F	4,66E-03
	GO:0046527	glucosyltransferase activity	F	1,62E-02
	GO:0035251	UDP-glucosyltransferase activity	F	1,39E-02
	GO:0008194	UDP-glycosyltransferase activity	F	2,54E-02
	GO:0044445	cytosolic part	C	2,54E-02
	GO:0005829	cytosol	C	3,22E-02
	GO:0000015	phosphopyruvate hydratase complex	C	1,62E-02
	GO:0004634	phosphopyruvate hydratase activity	F	1,62E-02
	GO:0006313	transposition, DNA-mediated	P	4,66E-03
	GO:0004803	transposase activity	F	4,66E-03
	GO:0032196	transposition	P	4,66E-03
	GO:0004198	calcium-dependent cysteine-type endopeptidase activity	F	4,34E-02
	GO:0004197	cysteine-type endopeptidase activity	F	4,56E-02
communauté rouge	GO:0009765	photosynthesis, light harvesting	P	2,87E-04
	GO:0019684	photosynthesis, light reaction	P	4,80E-04
	GO:0015979	photosynthesis	P	1,11E-03
	GO:0006091	generation of precursor metabolites and energy	P	5,15E-03
	GO:0003913	DNA photolyase activity	F	7,05E-04
	GO:0016830	carbon-carbon lyase activity	F	2,59E-02
	GO:0003904	deoxyribodipyrimidine photo-lyase activity	F	1,53E-02
	GO:0003950	NAD+ ADP-ribosyltransferase activity	F	1,93E-03
	GO:0016763	transferase activity, transferring pentosyl groups	F	2,50E-02
	GO:0016709	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen	F	1,02E-02
	GO:0004499	N,N-dimethylaniline monooxygenase activity	F	1,02E-02
	GO:0000041	transition metal ion transport	P	2,54E-02
	GO:0046915	transition metal ion transmembrane transporter activity	F	2,54E-02
	GO:0006826	iron ion transport	P	1,53E-02
	GO:0006827	high-affinity iron ion transmembrane transport	P	1,02E-02
	GO:0033573	high-affinity iron permease complex	C	1,02E-02
	GO:0005381	iron ion transmembrane transporter activity	F	1,53E-02
	GO:0034755	iron ion transmembrane transport	P	1,53E-02
	GO:0004842	ubiquitin-protein transferase activity	F	2,79E-02
	GO:0019787	ubiquitin-like protein transferase activity	F	2,79E-02
	GO:0016567	protein ubiquitination	P	1,61E-02
	GO:0032446	protein modification by small protein conjugation	P	1,94E-02
	GO:0070647	protein modification by small protein conjugation or removal	P	4,36E-02
	GO:0022834	ligand-gated channel activity	F	4,52E-02
	GO:0015276	ligand-gated ion channel activity	F	4,52E-02
	GO:0005230	extracellular ligand-gated ion channel activity	F	3,54E-02
	GO:0022610	biological adhesion	P	4,03E-02
	GO:0007155	cell adhesion	P	4,03E-02
communauté verte	GO:0022857	transmembrane transporter activity	F	3,27E-02
	GO:0022892	substrate-specific transporter activity	F	2,54E-02
	GO:0022891	substrate-specific transmembrane transporter activity	F	1,63E-02
	GO:0005215	transporter activity	F	1,85E-02
	GO:0055085	transmembrane transport	P	1,30E-02
	GO:0071705	nitrogen compound transport	P	3,04E-03
	GO:0015696	ammonium transport	P	1,94E-02
	GO:0008519	ammonium transmembrane transporter activity	F	1,94E-02
	GO:0072488	ammonium transmembrane transport	P	1,94E-02
	GO:0019755	one-carbon compound transport	P	9,76E-03
	GO:0015204	urea transmembrane transporter activity	F	9,76E-03
	GO:0042887	amide transmembrane transporter activity	F	9,76E-03
	GO:0071918	urea transmembrane transport	P	9,76E-03
	GO:0015840	urea transport	P	9,76E-03
	GO:0042886	amide transport	P	1,46E-02
	GO:0005230	extracellular ligand-gated ion channel activity	F	3,38E-02
	GO:0022834	ligand-gated channel activity	F	4,32E-02
	GO:0015276	ligand-gated ion channel activity	F	4,32E-02
	GO:0009767	photosynthetic electron transport chain	P	9,76E-03
	GO:0009772	photosynthetic electron transport in photosystem II	P	9,76E-03
	GO:0045156	electron transporter, transferring electrons within the cyclic electron transport pathway of photosynthesis activity	F	9,76E-03

	GO:0003983	UTP:glucose-1-phosphate uridylyltransferase activity	F	9,76E-03
	GO:0051748	UTP-monosaccharide-1-phosphate uridylyltransferase activity	F	9,76E-03
	GO:0009225	nucleotide-sugar metabolic process	P	4,32E-02
	GO:0006011	UDP-glucose metabolic process	P	9,76E-03
	GO:0004871	signal transducer activity	F	3,15E-02
	GO:0060089	molecular transducer activity	F	3,15E-02
	GO:0004872	receptor activity	F	1,19E-02
	GO:0038023	signaling receptor activity	F	1,19E-02
	GO:0004673	protein histidine kinase activity	F	3,38E-02
	GO:0016775	phosphotransferase activity, nitrogenous group as acceptor	F	3,38E-02
	GO:0000155	phosphorelay sensor kinase activity	F	3,38E-02
	GO:0072593	reactive oxygen species metabolic process	P	1,46E-02
	GO:0006801	superoxide metabolic process	P	1,46E-02
	GO:0004784	superoxide dismutase activity	F	2,42E-02
	GO:0016721	oxidoreductase activity, acting on superoxide radicals as acceptor	F	2,42E-02
	GO:0016624	oxidoreductase activity, acting on the aldehyde or oxo group of donors, disulfide as acceptor	F	3,38E-02
	GO:0008237	metallopeptidase activity	F	4,94E-02

Annexe E : liste des amorces utilisées pour l'analyse par q-RT-PCR

Gene	Forward	Revers
MYB-rel_11	CGAGCAATACGAAGACTACGG	GGGACTGCTACGACTATACCG
Myb-2R_20	CTCCTTCTGAGTCGGAATCG	TTCAAGATGCTCGTCACACC
PLAAOx	GACACACGATTTGGAGGACG	TGAGATTGTGTCGAACGTGC
CSAP	CTGCCAACCATCACTTTCCC	ATGAACTTGTGCCACGTGAC
Nrt 2.1	TCTCCCCAATCGTCTCGC	CAACCCTGCAACGCTCTCAC
18872	TCGCAGGTTGTAAGGTACGG	CCGTTTTCTCCGAGTAGGG
19975	CGCTTCTGGAGTTCATTTGC	TGTAGATGGCTTCCCTTTGC
14258	TATCACCGCTACCACTGTGC	GCATCTCCTCATACCCAACC
16415	CCTCTTCTCCTTCCCAAAGC	ATCGCCTCGTCTTTCTCTCC
17273	CCTTCTTTGCGTCCTTACCG	GCGATTGGTATATCAAGTACCG
10371	TTCAGAAGGCTCGGACTACC	CGAGCATTCTTCCTCTTTGG
15474	CCTCATATCCGAGCACATCC	CTCCTGTGGGCTCAACTAGC
4051	GTCGAAGCAGGGAAGAAGC	GAAAAGCGGTAGCACAGAGC
EF1	GTAGCGTGCCTTCTTGTAGC	TAACTTCACCACTGCCATCC
GAPDH	CGGTGCTCAATGTAGTGGTT	TAGTGATCATGCCCTTCTCG

UNIVERSITE BRETAGNE LOIRE

Université du Maine

Thèse de Doctorat

Stanisals THIRIET-RUPERT

Etude des facteurs de transcription impliqués dans l'accumulation lipidique en réponse à un stress azoté chez la microalgue *Isochrysis* affinis *galbana*

Study of transcription factors involved in lipid accumulation induced by nitrogen stress in the microalgae *Isochrysis* affinis *galbana*

Résumé

Chez tout organisme, l'évolution et l'acclimatation aux changements du milieu de vie sont orchestrés par de nombreux acteurs moléculaires. Parmi eux, les facteurs de transcription (FTs) jouent un rôle clé en régulant l'expression des gènes. Identifier les FTs impliqués dans la production de composés d'intérêt est donc une étape importante dans un contexte biotechnologique. Le laboratoire dispose d'une souche mutante de la microalgue haptophyte *Tisochrysis lutea* produisant deux fois plus de lipides de réserve que la souche sauvage en condition de privation azotée. Compte tenu du rôle clé des FTs dans l'établissement du phénotype, cette thèse vise à identifier les FTs impliqués dans la mise en place de ce phénotype mutant.
Un pipeline bio-informatique d'identification et classification des FTs présents dans le génome de *T. lutea* a été élaboré. Le manque de donnée chez les haptophytes constituant un vide dans l'étude de l'histoire évolutive des microalgues, une étude comparative des FTs présents dans le génome d'algues de différentes lignées a été réalisée. Celle-ci révèle que l'étude des FTs aide à comprendre et illustrer l'histoire évolutive des microalgues par la mise en évidence de présences/absences de familles de FTs spécifiques de lignée.
Afin de comprendre l'établissement du phénotype de la souche mutante de *T. lutea*, des données transcriptomiques ont permis la construction de réseaux de co-expression et de régulation des gènes chez les deux souches. Leur analyse croisée a identifié sept FTs candidats potentiellement liés au phénotype mutant. Une approche de p-RT-PCR a confirmé l'implication de deux FTs dans la remobilisation de l'azote en condition de privation azotée.

Mots clés
Bioinformatique – biologie moléculaire – évolution - facteur de transcription - micro-algue - réseau de gènes - RNA-seq

Abstract

In every organism, evolution and acclimation to environmental changes are orchestrated by numerous molecular players. Among them, transcription factors (TFs) play a crucial role by regulating gene expression. Therefore, identify TFs involved in the production of high value products is a significant step in a biotechnological context. The laboratory has at its disposal a mutant strain of the haptophyte microalga *Tisochrysis lutea* producing twice more storage lipids than the wild type strain when exposed to nitrogen deprivation. Given the key role of TFs in phenotype establishment, this PhD aim at identify the TFs involved in that of the mutant phenotype of *T. lutea*.
A TFs identification and classification pipeline was elaborated and applied to *T. lutea*'s genome. Since the lack of data in haptophytes constitutes a limit in studies on microalgae evolutionary history, a comparative study of TFs identified in the genome of microalgae belonging to different lineages was carried out. This study reveals that TFs could be used to understand and illustrate microalgae evolutionary history through the highlight of lineage specific presence/absence of TF families.
Aiming at understanding *T. lutea*'s mutant strain phenotype establishment, transcriptomic data were used to build gene co-expression networks and gene regulatory networks for both strains. Their comparative analysis identified seven TFs potentially liked to the mutant phenotype. A q-RT-PCR approach confirmed the involvement of two TFs in nitrogen recycling under nitrogen deprivation.

Key Words
Bioinformatic – evolution – gene network - molecular biology – microalgae – RNAseq - transcription factor