Draft genomes and phenotypic characterization of Tisochrysis lutea strains. Toward the production of domesticated strains with high added value

Gregory Carrier, Caroline Baroukh, Catherine Rouxel, Laetitia Duboscq-Bidot, Nathalie Schreiber, Gaël Bougaran

Algal Research 29 (2018) 1–11

Contents lists available at ScienceDirect

Algal Research

journal homepage: www.elsevier.com/locate/algal

Draft genomes and phenotypic characterization of *Tisochrysis lutea* strains. Toward the production of domesticated strains with high added value

Gregory Carrier[a,*], Caroline Baroukh[a,b], Catherine Rouxel[a], Laëtitia Duboscq-Bidot[c], Nathalie Schreiber[a], Gaël Bougaran[a]

[a] IFREMER, Physiology and Biotechnology of Algae laboratory, rue de l'Ile d'Yeu, 44311 Nantes, France
[b] INRA UR0050, Laboratoire de Biotechnologie de l'Environnement, avenue des étangs, 11100 Narbonne, France
[c] INSEMR UMR 1087, IRS-Université de Nantes, 8 quai Moncousu, 44007 Nantes, France

ARTICLE INFO

Keywords:
Microalgae
Genome
Improvement program
Lipids
Tisochrysis lutea

ABSTRACT

Tisochrysis lutea is a microalga species currently used in aquaculture as a feed for shellfish, oysters and shrimps. It also has many other potential industrial applications, such as the production of neutral lipids for biofuels or the production of ω-3 fatty acids for nutraceuticals (human food complements). To efficiently exploit the potential of this microalga, however, higher lipid productivities are needed. To this end, improvement programs need to be developed and optimized. The diversity of strains available in microalgae has not yet been exploited in such improvement programs.

In this study, the intra-strain diversity was observed and exploited to increase neutral lipid productivity. New clonal strains with higher neutral lipid productivity were successfully selected. The best clonal strain selected accumulated 520% more triacylglycerols, with a similar growth rate to the wild-type strain in continuous light and nitrogen starvation conditions. In a photoperiod culture condition, this clonal stain also accumulated 84% more storage lipids and 30% less carbohydrates, compared to the wild-type strain. This clonal strain thus had a higher productivity which is of great interest for feed or biofuel applications.

This study also focused on identifying the genomic mechanisms responsible for the improvements in these clonal strains. With this objective, the genome of *Tisochrysis lutea* was sequenced for the first time. It is the third genome of a Haptophyte microalga sequenced so far. Different genetic polymorphisms were identified between the sequenced genomes of the wild-type strain and clonal strains. Activity of transposable elements seems to have been involved in the genome reshuffling obtained through the improvement program. The contribution of transposable elements to the adaptive capacity of microalgae remains to be demonstrated.

1. Background

Marine microalgae are unicellular photosynthetic eukaryotes. They form the first link in the food chain for aquatic organisms, and are responsible for 35% of the Earth's primary production in the ocean [1,2]. We have only just begun to explore the world of microalgae [3]. Currently, around 140,000 algae species have been inventoried (http://www.algaebase.org/), but their true number is estimated to be around 1,000,000 species [4]. With the depletion of terrestrial resources, microalgae are attracting a great deal of interest. They can be exploited in many areas such as food, feed, cosmetics, bioremediation and production of third-generation biofuels [5,6]. However, in many domains, there are obstacles that need to be solved to develop economically and environmentally-friendly processes.

The microalgae research and development community agrees that the improvement of algae strains is one of the major factors that must be addressed [7,8]. Currently, all cultivated microalgae strains are directly derived from their natural environment and are considered as wild-type strains. As with modern agriculture, microalgae exploitation requires improved strains to be obtained in order to become economically viable [9]. The improvement of microalgae is a recent idea, since the first microalgae improvement programs were set up in the last decade. However, at present, fragmented knowledge on the biology of microalgae species (biodiversity, physiology, intracellular mechanisms, phenotypes, lifecycle, *etc.*) limits the use of improvement strategies. The current challenge is thus to develop and optimize improvement methodologies to reduce the gap between the methods available for microalgae and practices already in use for higher plants and animals.

* Corresponding author.
E-mail addresses: Gregory.Carrier@ifremer.fr (G. Carrier), caroline.baroukh@inra.fr (C. Baroukh), Catherine.Rouxel@ifremer.fr (C. Rouxel), Laetitia.Duboscq-Bidot@inserm.fr (L. Duboscq-Bidot), Nathalie.Schreiber@ifremer.fr (N. Schreiber), Gael.Bougaran@ifremer.fr (G. Bougaran).

https://doi.org/10.1016/j.algal.2017.10.017
Received 22 February 2017; Received in revised form 1 September 2017; Accepted 24 October 2017

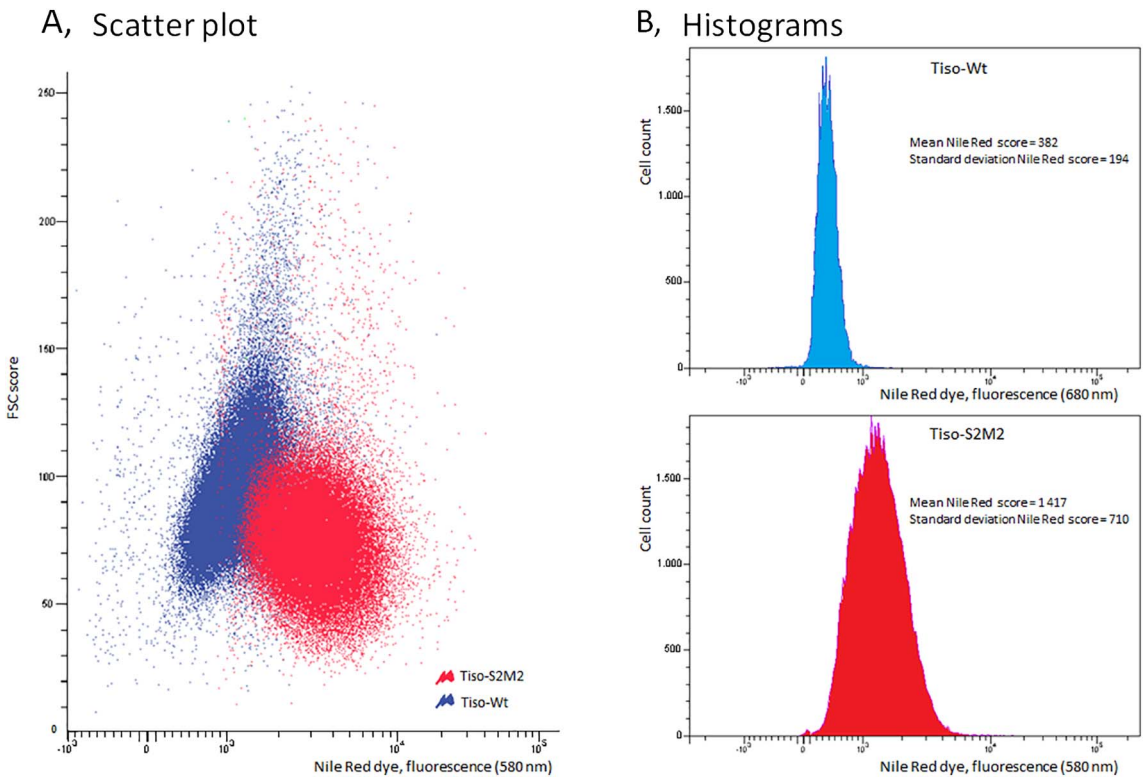

Fig. 1. Neutral lipid content of the cells of Tiso-WT and Tiso-S2M2 strains. Fig. A: scatterplot where each point represents an algae cells; blue: Tiso-WT strain; red: Tiso-S2M2 strain. All algal cells were dyed with Nile Red dye. Nile Red fluoresces at 560 nm and is linearly related to the neutral lipid content of the cell [12]. A high heterogeneity in neutral lipid content was observed in both strains. Fig. B: histograms showing the distribution of algae cell numbers as a function of Nile Red fluorescence. The mean neutral lipid content was higher for Tiso-S2M2 than for Tiso-Wt. Additionally, the standard deviation was higher for Tiso-S2M2 than Tiso-Wt, demonstrating a higher heterogeneity for neutral lipids in Tiso-S2M2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Tisochrysis lutea (Tiso) was chosen for this study because it has been used for decades in aquaculture as feedstock for shellfish, oyster and shrimps and has a promising future in the production of human food complements. Indeed, this microalga contains many antioxidant molecules of interest for health. In addition, it has been demonstrated to have a high content of long-chain polyunsaturated fatty acids such as docosahexaenoic acid (DHA) [10] and stearidonic acid (SA) [11].

An improvement program to increase lipid productivity was conducted by Bougaran et al. [12] using an empirical selection strategy based on a sequential mutation-selection procedure performed on the Tiso wild type strain. Mutagenesis was performed on Tiso-Wt with UV-c radiation treatments to increase the probability of obtaining individuals with interesting lipid profiles. To select these, a selection procedure was then applied using a flow cytometer with a cell sorter. The sorting of the best candidates according to their lipid traits was assisted by neutral lipid staining (Nile Red dye). Among the algal cells, the 10% most lipid-rich individuals were selected. Two successive rounds of mutation-selection were applied, through which the Tiso-S2M2 strain was obtained. The improved strain, demonstrated an increase in neutral lipid productivity of 80% compared to the initial Tiso-Wt strain [13]. In addition, the ability for lipid over-accumulation was proven to remain stable under our culture conditions for more than 6 years (until present).

After this first improvement program, molecular approaches were used to characterize the improved strain (Tiso-S2M2) and to guide further improvement strategies. An RNAseq approach was performed on cultures under nitrogen deficiency to compare coding sequences and their expression levels between Tiso-Wt and Tiso-S2M2 strains [14]. Through this work, the first reference transcriptome of the *Tisochrysis lutea* species was built. Variations in RNA and protein expression levels were measured to better characterize the metabolic differences between Tiso-S2M2 and Tiso-Wt [14,15]. At the different molecular levels (protein and RNA), many variations were identified associated with candidate genes (more than a hundred genes). This high candidate number suggested a large genome alteration during the improvement program leading to the Tiso-S2M2 strain. Overall, lipid synthesis in Tiso-S2M2 seemed similar to that in Tiso-Wt, whereas lipid catabolism, carbon assimilation, carbohydrate metabolism and many regulation proteins seemed to be more affected. Additionally, coding sequence analysis in the two strains revealed polymorphism (SNPs). Comparison of this genetic polymorphism between Tiso-Wt and Tiso-S2M2 revealed that, quantitatively, the amount of molecular variation was maintained in the improved strain Tiso-S2M2. These results suggested a conservation of algal cell diversity in Tiso-S2M2 despite the mutation-selection procedures. This diversity could be exploited for subsequent improvement programs. Several studies show that screening microalgal diversity is a good approach for selecting the most interesting strains [16–19].

In the present study, following the results from earlier research [12,14], the extent of intra-strain diversity in the wild-type strain and improved strain (Tiso-S2M2) were qualified on the neutral lipid content. Then, intra-strain diversity of Tiso-S2M2 was exploited to obtain new strains further improved for their neutral lipid content. Obtaining domesticated strains with high neutral lipid content is of interest for feed and biodiesel applications. To this end, cells with a high content of neutral lipids were isolated from Tiso-S2M2. The two best clonal strains (Tiso-S2M2-CL1 and Tiso-S2M2-CL2) in terms of neutral lipid content were examined to obtain details at the genomic and phenotypic level that would improve our understanding of the impact of the improvement program and the isolation procedure. To perform this step, the genome of the wild-type strain and the genome of the two new clonal strains were sequenced and their genetic polymorphism was assessed. A first draft reference genome of Tiso was reconstructed from the sequenced data. We thus also report the first sequenced genome of

Tisochrysis lutea, which is the third genome sequenced in the Haptophyta phylum.

2. Results and discussion

2.1. Highlighting intra-strain diversity of neutral lipids in Tiso-S2M2

The first improvement program resulted in a strain with a higher lipid productivity (Tiso-S2M2) than the wild-type strain (Tiso-Wt) [12]. Despite selection and mutation processes during the improvement program, molecular diversity (based on allele frequency measured in the transcriptome of each strain) persisted in the Tiso-S2M2 strain [14]. Following these results, we aimed to assess whether the molecular diversity within a strain could also be observed at the phenotypic level, particularly for the neutral lipid traits. Neutral lipid content was therefore measured in each algae cell of Tiso-Wt and Tiso-S2M2 strains, using Nile Red dye with a flow cytometer (Fig. 1). Unsurprisingly, mean neutral lipids were higher in Tiso-S2M2 than Tiso-Wt (Nile Red, fluorescence dye in mean: 382 for Tiso-Wt and 1417 for Tiso-S2M2, Fig. 1B). Neutral lipid content showed a higher heterogeneity between algae cells in Tiso-S2M2 than in Tiso-Wt (standard deviation was 710 for Tiso-S2M2 and 194 for Tiso-Wt). Contrary to the molecular variation previously observed by Carrier et al. [14], heterogeneity for neutral lipid content was not similar between strains. Why this greater variation was not observed by the transcriptomic approach remains unclear. One reason could be that Carrier et al. [14] limited their investigation to coding regions and not the complete genome. Secondly, the focus was on neutral lipids, but this single trait does not reflect the complete phenotypic heterogeneity within Tiso-S2M2.

2.2. Use of intra-strain diversity to produce strains of interest

Our previous results revealed the heterogeneity of neutral lipids between algal cells within Tiso-S2M2 strains. We aimed to exploit the intra-strain diversity to obtain clonal strains with a higher neutral lipid content. The cytometer approach is an efficient strategy to exploit intra-strain diversity in microalgae [16,20,21]. By this method, 708 algae cells among the 10% of algae with the highest Nile Red fluorescence in Tiso-S2M2 strain were isolated using a sorting flow cytometer (FAC). Among these, 106 (15%) clonal strains survived after isolation, and only nine clonal strains (1.5%) were obtained after 49 days of the recovery step (Fig. S2). After the recovery step, the neutral lipid content of the nine clonal strains was measured. All clonal strains had a higher neutral lipid content than the mean content of the Tiso-S2M2 strain (Fig. 2). Interestingly, the high neutral lipid content in clonal strains was conserved despite the 5 months of growth during the recovery stage (approximately 150 generations), meaning that the lipid trait was fixed. Additionally, clonal strains revealed variable neutral lipid contents, confirming the diversity within Tiso-S2M2. These results show the efficacy of using intra-strain diversity to obtain new algae strains of interest. Moreover, during the photoperiod experiment performed after 2 years of the isolation procedure (see results in the next section), cytometer analysis showed that the heterogeneity of the clonal strains (Tiso-S2M2-Cl1 and Cl2) was lower (standard deviation 6924 and 5818 respectively) than the heterogeneity of the wild-type strain (10,248, Fig. S3). These results confirm the reduction of diversity in clonal strains after the isolation procedure, although the heterogeneity in clonal strains is still significant after 2 years suggesting that there could be genetic drift in these strains.

Several studies on microalgal biodiversity showed that different strains belonging to the same species can show high inter-strain diversity [17–21]. A few studies also showed the interest of studying and exploiting this strain diversity of microalgal species such as *Chlorella sorokiniana* [18]. In the case of Tiso, ten different strains have been inventoried in international collections. These originated from various ecosystems, ranging from the tropical waters of Tahiti to the cold waters of the North Sea (Bendif et al. [31]). Recently, an eleventh strain was identified in the Sea of Japan [74]. Tiso inter-strain diversity could also be used to further increase the efficiency of improvement programs.

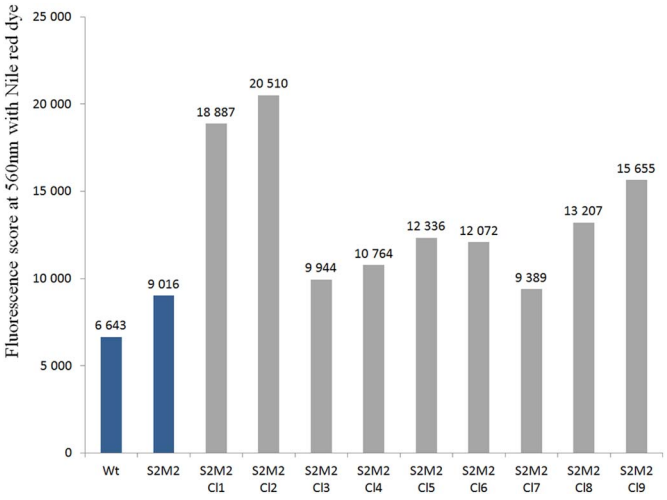

Fig. 2. Neutral lipid content among the nine clonal strains. Cells were stained with Nile Red and fluorescence was measured at 560 nm to compare neutral lipid content among the strains.

2.3. Physiological characterization of the two isolated clonal strains with the highest neutral lipid content

2.3.1. Culture characterization under continuous light and nitrogen starvation

Among the nine clonal strains obtained, the two clonal strains (Tiso-S2M2-Cl1 and Tiso-S2M2-Cl2) with the highest neutral lipid content were characterized in detail to estimate their potential for lipid production and to better understand the origin of their specificity. First, cultures were performed in batch mode under continuous light. Continuous light is the most frequently used laboratory culture mode although it does not reflect industrial outdoor production under natural light. Cultures were sampled at the stationary phase where nitrogen starvation occurred (Fig. S1). Nitrogen starvation is known to affect algal metabolism: cell division stops while neutral lipid content increases [27]. Lipid content was measured using HPTLC (Table 1 and Table S6). The two clonal strains (Tiso-S2M2-Cl1 and Tiso-S2M2-Cl2) had a higher total lipid content than Tiso-S2M2 and Tiso-Wt: + 207% and + 209% respectively than Tiso-Wt. Moreover, phospholipids and glycolipids were very similar between all strains, whereas triacylglycerols (TAG) over-accumulated in clonal strains: approximately + 460% and + 520% in Tiso-S2M2-Cl1 and Tiso-S2M2-Cl2, respectively, compared with Tiso-Wt. The clonal strain with the highest storage lipid content (mainly TAG) was Tiso-S2M2-Cl2. Additionally, the growth rate of each strain was calculated by logarithmic linearization when algae were in the exponential grow phase. The growth rate was similar for Tiso-Wt, Tiso-S2M2, Tiso-S2M2-Cl1 (2.27, 2.22 and 2.21 day^{-1}, respectively) and a little lower for Tiso-S2M2-Cl2 (1.97 day^{-1}) in these culture conditions. The lipid profile and the growth rate of these two clonal strains did not seem completely identical, meaning that the two strains have several specificities.

2.3.2. Characterization nutrient replete conditions with photoperiod

A complementary characterization of these strains was performed with monitored cultures submitted to light/dark cycles (12 h/12 h) mimicking daylight variations and without any nutrient limitation (Fig. 3). The culture conditions were closer to that used in industrial outdoor production under natural light. Several physiological traits

Table 1
Lipid content among classes in the four Tiso-strains as measured by HPTLC.

Strains	Total lipids (mg cell^{-1})		Triacylglycerols (mg cell^{-1})		Phospholipids (mg cell^{-1})		Glycolipids (mg cell^{-1})	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Tiso-Wt	2.97E − 03	9.42E − 07	4.13E − 04	2.88E − 07	9.58E − 04	3.05E − 07	1.44E − 03	7.47E − 07
Tiso-S2M2	4.33E − 03	2.95E − 05	1.39E − 03	6.24E − 06	1.03E − 03	3.57E − 06	1.49E − 03	1.12E − 05
Tiso-S2M2-Cl1	6.16E − 03	5.58E − 05	2.87E − 03	2.76E − 07	1.28E − 03	8.97E − 07	1.65E − 03	4.17E − 05
Tiso-S2M2-Cl2	6.22E − 03	2.40E − 05	3.11E − 03	7.70E − 06	1.05E − 03	1.82E − 05	1.44E − 03	1.57E − 05

were measured every 3 h for 30 h (Fig. 3): i) cell concentration, ii) size of cells iii) cell carbon and nitrogen quotas, iv) neutral lipid content v) carbohydrate content and vi) protein in algal cells. As expected from results from many algae species [28], Tiso was affected by the photoperiod.

During the light period, cell division stopped, and carbon storage metabolites (carbohydrates and neutral lipids) accumulated in the cells. Consequently, the cell carbon quota (total carbon within the cell) and cell size increased during the light period (Fig. 3-C). Conversely, during the dark period, cell division was observed (Fig. 3-A shows how cell concentration doubled within one night), carbon storage metabolites were consumed (Fig. 3-C-D-E) and, consequently, the total carbon content and cell size decreased (Fig. 3-C-B). Additionally, the total amount of protein in the cells seemed to slightly increase during the light period and to decrease during the dark period, although it did not vary substantially (Fig. 3-F). Similar effects of photoperiod were

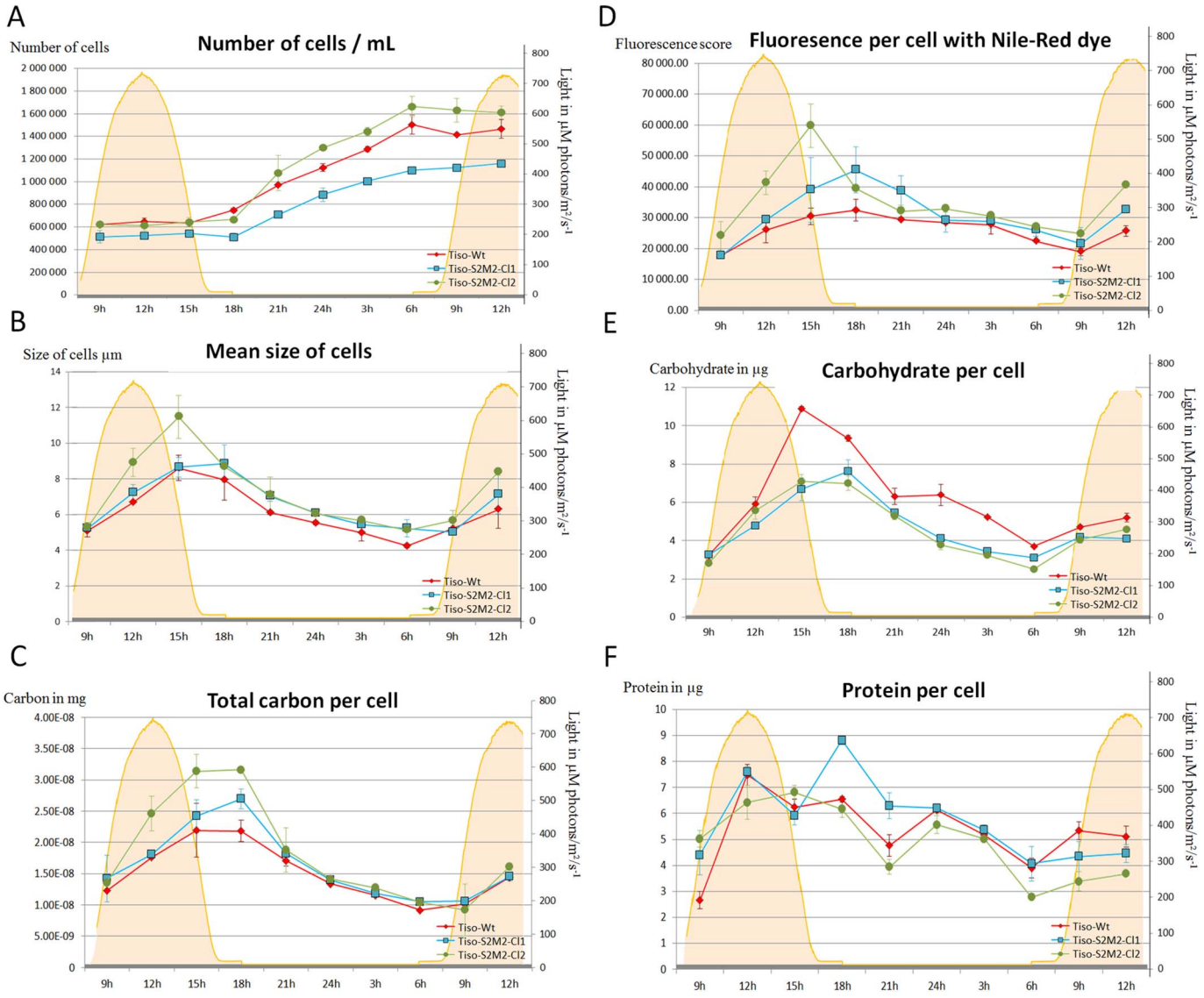

Fig. 3. Phenotypic characterization of Tiso strains (Tiso-Wt, Tiso-S2M2-Cl1 and Tiso-S2M2-Cl2) under a light-dark cycle. For all strains, cultures were performed with two biological replicates.
Over 30 h, a sample of each culture was harvested every 3 h. The cultures of microalgae were not nutrient-limited. The number of cells (A), their size (B) and Nile Red fluorescence (D) were measured by flow cytometry. Cell total carbon content (C) was measured using an elemental analyzer. Cell carbohydrate (E) and protein (F) contents were measured by spectrophotometer.

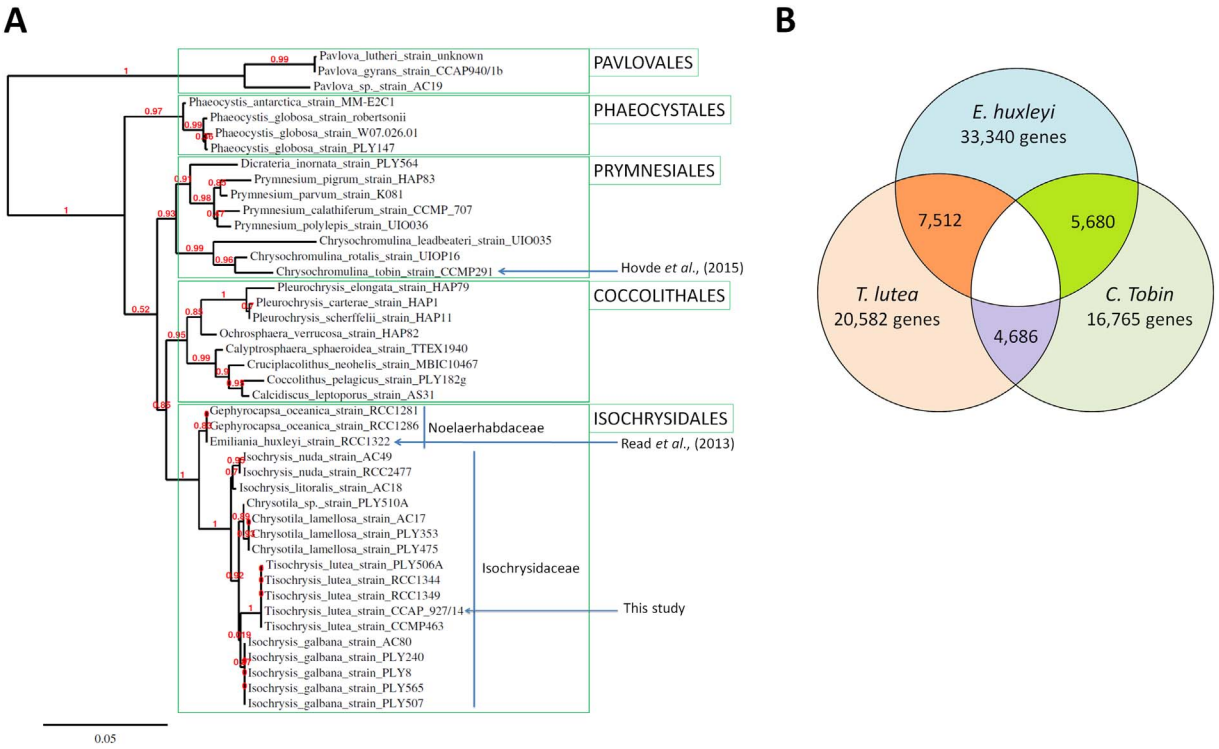

Fig. 4. Phylogeny of haptophytes and comparison of the number of genes for the three haptophytes sequenced to date.
A. A phylogeny tree of haptophytes inferred from 18S sequence using maximum likelihood. Bootstrap values are indicated in red at the nodes. B. Numbers of genes in common or not between the three haptophytes sequenced. Genes between two algae are considered homologous if the protein sequence is similar over more than 30% of its total length. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

previously reported for the same species [29].

Under these culture conditions, a high variation was observed between the three strains. Unsurprisingly, the amount of neutral lipid accumulated during the light period was higher in Tiso-S2M2-Cl1 and Tiso-S2M2-Cl2 than in Tiso-Wt (+ 40% and + 84%, respectively, Fig. 3-D). Between the two clonal strains, a variation in neutral lipid content was also observed. During the light period, Tiso-S2M2-Cl2 accumulated lipids more quickly than Tiso-S2M2-Cl1: the highest neutral lipid content was reached at approximately 15 h for Tiso-S2M2-Cl2 compared with 18 h for Tiso-S2M2-Cl1. During the dark period, lipid consumption also seemed faster for Tiso-S2M2-Cl2 than for Tiso-S2M2-Cl1. In addition, the amount of neutral lipids was similar in all strains at the end of the dark period, suggesting that the minimum neutral lipid quota is conserved between strains.

Experimental results on the amount of carbohydrates were the opposite to those on lipids (Fig. 3-E). The amount of carbohydrates was higher in the Tiso-Wt strain than in the clonal strains, particularly during the light period: − 24% and − 30% in Tiso-S2M2-Cl1 and Tiso-S2M2-Cl2, respectively, compared with Tiso-Wt (Fig. 3). Thus, we can assume that a higher quantity of carbon assimilated during the light period was allocated to lipids rather than to carbohydrates in Tiso-S2M2-Cl1 and Tiso-S2M2-Cl2. This metabolic shift observed between the wild-type and the clonal strains could explain part of the over-accumulation of lipids in the clonal strains.

This over-accumulation can also be partly explained by the growth rate of the improved strains. Indeed, during the light period, the carbon quota in Tiso-S2M2-Cl1 and Tiso-S2M2-Cl2 was higher than in Tiso-Wt: + 23% and + 44%, respectively, compared with Tiso-Wt (Fig. 3-C). Yet, the carbon quota was similar at the end of the dark period in all strains, whatever the amount of storage carbon accumulated during the light period. These results first suggest that the excess carbon stored during the day is entirely consumed during the night as the cells divide. In addition, the amount of accumulated carbon during the light period limits the growth rate during the dark period. Hence, since more carbon is stored in the improved strains during the day, the growth rate is higher for these strains during the dark period, as supported by the experimental data (Fig. 3-A): Tiso-S2M2-Cl2 grew at $2.55\,j^{-1}$ ($\pm 0.20\,j^{-1}$), Tiso-S2M2-Cl1 at $2.16\,j^{-1}$ (± 0.05) and Tiso-Wt at $2.14\,j^{-1}$ (± 0.09).

Concerning industrial culture conditions, a higher productivity of neutral lipids could be interesting. Considering one day of culture and a harvest of algae when the content in lipid is optimal (approximately 15 h), the productivity for neutral lipid is + 103% for Tiso-S2M2-Cl1 compared to Tiso-Wt. This strain could be particularly interesting in different industrial fields such as biofuels or feed production.

2.4. Draft reference genome of Tisochrysis lutea and comparison with other sequenced Haptophytes

In order to analyze the strain improvement program at the genetic level, it was necessary to sequence the Tiso genome. Nowadays, a reference genome is necessary to use many molecular approaches and also provides better knowledge of the evolutionary history of the species. A draft *Tisochrysis lutea* reference genome was obtained from the sequencing of the three Tiso strains in this study (Tiso-Wt, and the two new clonal strains). Sequencing data of these three strains were combined to obtain the Tiso reference genome, which is thus a consensus of these strains. The sequencing was performed using an Illumina Sequencer HiSeq 2000. The library preparation was performed in mate-pair mode to optimize genome assembly [30]. In all, 358 million reads were obtained and assembled into 7662 contigs (N50 = 10,571 b) with an average read depth higher than 200 × (see Material and methods). This first draft reference genome has an estimated size of 54.38 Mb. The genome was then automatically annotated and 20,582 genes were identified. On average, the genes contained a single intron and 47% of the genome was predicted to be coding sequences. A putative function of the identified genes was assigned. Among the 20,582 genes predicted, 10,383 genes (50.4%) had a putative function thanks to a

Table 2
Genome comparison of the three haptophytes sequenced to date.

	Tisochrysis lutea	*Chrysochromulina tobin*	*Emiliania huxleyi*
Genome size	54 Mb	59 Mb	141 Mb
Number of genes	20,582	16,777	38,549
Coding sequences	47%	40%	21.90%
Multi-copy genes	5321	4866	24,469

sequence or protein domain homologous to ones known in databases. In addition, 2096 genes (17.7%) had an Enzyme Code (EC) identified. The amount of genes with a putative function identified was similar to the amount for other microalgae sequenced to date (approximately 50%).

Tisochrysis lutea belongs to the Haptophyta phylum [31] (Fig. 4). This phylum is of major importance in aquatic ecosystems, being composed of algae species that are dominant either in freshwater, or in the seas and oceans [32,33]. Currently, only two species have draft sequenced genomes available: *Chrysochromulina tobin* [34] and *Emiliania huxleyi* [35] (Fig. 1). The reference genome of Tiso was compared to the genomes of these algae (Fig. 4). The Tiso genome is of a similar size to the *C. tobin* one, but is half the size of the *E. huxleyi* genome (Table 2). The number of genes and the percentage of coding regions identified in Tiso are also closer to *C. tobin* than *E. huxleyi* (Table 2). Indeed, the number of genes present in multiple copies in the Tiso and *C. tobin* genomes is smaller (25.8% and 29.0%, respectively) than in the *E. huxleyi* genome (73.3%). These data show that, in terms of genome structure, Tiso is more similar to *C. tobin* than *E. huxleyi*, despite their phylogenetic distance. The predicted proteomes of these three sequenced haptophytes were also compared in order to identify the number of genes homologous between these algae (Fig. 4). Tiso has a higher number of homolog genes with *E. huxleyi* (7512; 36.4%) than with *C. tobin* (4686; 22.7%). Contrary to the results obtained on the genome structure, the number of genes in common is in accordance with the phylogenetic distance between these algae (Fig. 4). Interestingly, the number of homolog genes in these three haptophytes is only 2733, highlighting their large phylogenetic distance.

These comparisons suggest that the evolutionary history of Tiso is complex and, on a broader level, so is the history of the different families of the Haptophyta phylum. The genome of *E. huxleyi* is twice as large and contains twice as many genes as those of Tiso and *C. tobin*. Moreover, the genome of *E. huxleyi* contains a high number of genes present in multiple copies compared to Tiso and *C. tobin* genomes and this level of multiple copies is similar to polyploid species such as soybean [36,37]. These first genomic comparisons thus suggest an ancient polyploidy event, probably generated by a particular secondary endosymbiosis, by the duplication of the genome from which *E. huxleyi* originated or, at a broader level, at the origin of the Noelaerhabdaceae family. Polyploidization events are relatively frequent in history of terrestrial plants and have an important role in species evolution [38,39]. It is therefore not surprising to see these events in algal history. In the future, a large evolutionary study on Haptophyte species at the genomic level would improve our understanding of the complex history of this phylum.

2.5. Genomic comparison of the two improved clonal strains

The draft genomes of Tiso-Wt, Tiso-S2M2-Cl1 and Tiso-S2M2-Cl2 were sequenced and compared to investigate the genomic shift induced by the improvement program and isolation procedure. The genome of each of the three strains was sequenced independently and the sequenced data for each strain was mapped on the reference genome produced in this study. Polymorphisms (SNPs and small or large insertion/deletions) from loci polyallelic within or between strains were identified using a combination of different software, as described in the Materials and methods section, to eliminate false positive polymorphism.

First, polyallelic loci were identified in each strain independently. The number of polyallelic loci was 9301 and 12,935 in Tiso-S2M2-Cl2 and Tiso-S2M2-Cl1, respectively, and 21,869 in Tiso-Wt (Fig. 5). The number of polyallelic loci is, unsurprisingly, higher in Tiso-Wt (around 2 fold) than in the clonal strains, due to the presence of several individuals in this strain. Among these data, the estimation of haplotype number was performed with HapCut and Freebayes software [40,41]. On average, 53 haplotypes were detected in Tiso-Wt. This number is a preliminary estimation because it is difficult to identify haplotypes with short reads. In the future, using long reads will improve accuracy. Concerning the clonal strains (Tiso-S2M2-Cl1 and Cl2), unsurprisingly, two haplotypes were identified confirming the diploid phase in the life cycle of Tiso.

Second, the genomic events which had given rise to the lipid-improved clonal strains (Tiso-S2M2-Cl1 and Tiso-S2M2-Cl2) were sought. Polyallelic loci and monoallelic loci of each strain were compared to identify polymorphic loci between strains (SNPs, small and large insertion/deletions). These polymorphic loci between Tiso-Wt and the

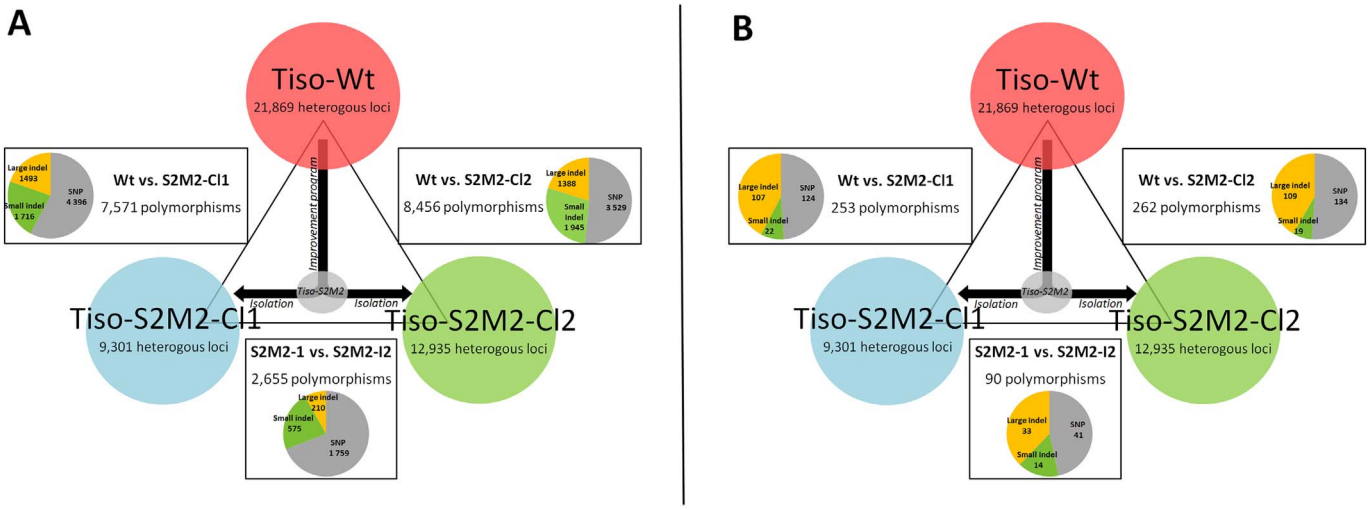

Fig. 5. Polymorphisms in and between Tiso strains.
The number of polyallelic loci identified in each strain is indicated in each circle. The number of polymorphic loci identified between two strains is given in the rectangles. The distribution of these polymorphic loci (SNP, small indel and large indel) is also given in each rectangle. A. The number of putative polymorphisms without an impact on genes is given in the rectangles. Fig. B. The number of putative polymorphisms with an impact on genes is given in the rectangles.

clonal strains would have arisen from genomic events during the selection program or isolation procedure or from genetic drift. Currently with these data, the specific moment when mutation was produced is not known. Broadly, the polymorphic loci between the Tiso-Wt strain and the clonal strains (7571 with Tiso-S2M2-Cl1 and 8456 with Tiso-S2M2-Cl1) are more numerous than the polymorphisms between the two clonal strains (2655 between Tiso-S2M2-Cl1 and Tiso-S2M2-Cl2). These results are coherent since a wild-type strain, containing diversity between individuals, was compared with clonal strains from single cells (Fig. 5). Interestingly, the genomic polymorphism between the two clonal strains (Tiso-S2M2-Cl1 and Tiso-S2M2-Cl2) is significant. The phenotypic variation previously observed between these clonal strains is probably due in part to the genomic polymorphisms identified here. From the polymorphisms observed between the clonal strains, a list of 88 candidate genes with a putative impact on strain phenotype was established (Table S3). Among these, 65 genes have no putative function and 23 genes have putative annotations that are distributed across diverse metabolic pathways, not showing any pathway that was particularly affected. However, eight genes (34% of the annotated genes) play a role in regulation of other proteins.

A list of candidate genes with shared polymorphisms in Tiso-S2M2-Cl2 and Tiso-S2M2-Cl1 but a difference with Tiso-Wt was also established (Table S4). Despite their specificity, due to their common origin, these clonal strains share a similar genetic and phenotypic background, mainly over-accumulation of lipid and under-accumulation of carbohydrates, see Section 2.3. Sixty-one candidate genes were identified and 36 had a putative function assigned (Table S4). Among the genes, only three had a putative function in lipid metabolism and two in carbohydrate metabolism. Interestingly, 14 candidate genes (38% of annotated genes) play a role in regulation of proteins including transcription factors, which can play a key role in phenotypic variation.

These primary results suggest that the regulation pathway is more affected than the core metabolism in these strains. The comparison of strains at the phenotypic level suggests that the regulation of primary pathways (carbohydrate and lipid accumulation) was affected during the light period (see results above). Additionally, a comparison study between Tiso-S2M2-Cl1 and Tiso-Wt strains in chemostat culture with different nitrogen conditions was realized [42]. Similarly, overall quantitative variation in carbon allocation was observed in Tiso-S2M2-Cl1 compared with Tiso-Wt according to change in nitrogen availability. The new phenotype seems not to be affected in its “core genes” but results from combinations of variants, which adapt the gene regulation network to make the phenotype suitable for the environment. These results also correspond to what has been observed in a few laboratory selection experiments [43], where regulation genes were more affected than structural genes such as enzymes of primary metabolism. In a future study, genetic quantitative approaches or regulation network approaches could be used to better characterize these improved strains.

2.6. The origin of polymorphisms giving rise to the improved clonal strains

The specific moment when mutations were produced cannot be identified with only theses current data. However, we can seek the major mechanisms for genome plasticity which were produced during the improvement program and which are responsible of these clonal strains. The distribution of the different classes of polymorphisms from which the clonal strains originated was established. The polymorphic loci detected between the strains do not have the same molecular origin and might not contribute similarly to the phenotypic variations observed. The polymorphisms identified could be a substitution for one Single Nucleotide Polymorphism (SNP), a gain or loss/gain of several nucleotides (small indel) or a large deletion/insertion of hundreds of nucleotides (large indel). The distribution of different polymorphisms between clonal strains and Tiso-Wt was similar for both clonal strains (55% ± 5 for SNPs, 25% ± 5 for small indels and 20% ± 1 for large indels (Fig. 5-A)). Unsurprisingly, these results showed no variation in the genomic reshuffle between the clonal strains isolated, which were derived from a single strain. The polymorphism distributions observed between strains reflect the impact of each polymorphism class on the phenotype. Comparison of the different classes reveals that the SNP polymorphisms are more frequent than the other types between clonal strains and Tiso-Wt and also between the two clonal strains (69% for SNPs, 23% for small indels and 8% for large indels (Fig. 5-A)). However, each class of polymorphism could affect the phenotype differently. It has been established that most polymorphisms maintained in genomes can be considered as of neutral effect and without impact on the phenotype [44]. Inversely, the mutations offering a phenotypic gain are rare while the deleterious mutations are eliminated quickly from the population. The SNP affects a single-nucleotide and consequently is usually neutral to the phenotype, whereas small indels and large insertions affecting several nucleotides would generate more impact on a gene. In this study, polymorphisms are considered to potentially affect the phenotype if an associated gene codes for variation that has been previously observed at the transcriptomic level [14] or impacting the amino acid(s) of the associated protein(s). It is not entirely complete and false positives could be present, but it is still a good first overview. Interestingly, SNPs remain the main polymorphism present between all strains (47% on average ± 2%), but large indels have a higher importance (38% on average ± 2) (Fig. 5-B). This is because only 3.3% of SNPs have a potential impact on the phenotype, whereas 7.5% of large indels have one. Large indels are less frequent at the scale of the whole genome (20%), but their likelihood of having an impact on the phenotype is higher than for the SNPs. Thus, in the future, large indels will be studied as carefully as SNPs. The origin of these large insertions was searched for using sequence similarity. Mainly transposable elements were identified (48% of large insertions matched with a transposable element known in the RepBase [45] database). Transposable elements are a molecular mechanism well known to respond to stress conditions and play an important role in genome evolution [43–45].

The polymorphism generated by the activity of transposable elements seems to be the major mechanism responsible for the genomic reshuffle that impacted the phenotype. Transposable elements are known to play a major role in genome evolution and the adaptation of species to their environment [49–51]. Therefore, the identification of a high activity of transposable elements induced by the improvement program is not surprising. Moreover, transposable elements are a quick solution to adapt a phenotype to the environment by modifying the gene regulation network without modifying the core genes [52,53]. Transposable elements play a major role in the explanation of the adaptation of invasive species to novel habitats despite very limited initial diversity [54]. The improvement program used here, from an initial Tiso strain of limited diversity, is a very similar case. Transposable elements could be one of the major actors of microalgal adaptation and evolution, although they have been little studied to date.

3. Conclusion

Microalgal cultivation has a great potential in many industrial domains. However, to become efficient, several challenges need to be overcome [8]. Microalgae species with potential industrial applications, such as *Tisochrysis lutea,* are not model organisms and their study has only just begun. One of the current challenges is to develop and optimize improvement methodologies as efficient as those used in the field of plants and animals [19]. This study associated a genomic and a phenotypic approach to provide a first basis for future microalgal improvement programs. Results of this study show that intra-strain diversity is present in Tiso and can be exploited to obtain improved strains of interest. An improved clonal strain with a high production of neutral lipids was obtained in this study. Moreover, at the genomic level, transposable elements seem to be one of the major actors responsible for this intra-strain diversity and deserve more extensive

study.

4. Materials and methods

4.1. Microalgae strains and culture conditions

The reference strain of this study, *Tisochrysis lutea* (Tiso-WT), previously named *Isochrysis* affinis *galbana* [31], was provided by the Culture Centre of Algae and Protozoa (CCAP 926/14). This strain was isolated in the late 1970s and has been maintained at the algae bank ever since. A previous improvement program made it possible to obtain a first lipid-improved strain (Tiso-S2M2, IFR-32B85) [12,13]. From this improved strain, two new clonal strains were obtained (Tiso-S2M2-Cl1 and Tiso-S2M2-Cl2) using an isolation program (see next section for details). Cultures of Tiso were performed in batch mode in 1-L flasks containing Conway medium [55] and aerated by bubbled 0.22 µm filtered-air. The cultures were maintained at a constant temperature of 21 °C, under a constant irradiance of 80 $\mu mol \cdot m^{-2} \cdot s^{-1}$.

4.2. Isolation procedure for the acquisition of improved clonal strains

The isolation of improved clonal strains was performed at the stationary phase of a batch culture of the strain Tiso-S2M2, when the algae were in nitrogen starvation and accumulated a large amount of neutral lipids (after 14 days of culture, similar to results shown in Fig. S1). To sort algae cells according to their neutral lipid content, Nile Red dye was used [56]. Nile Red was added to each sample (1 $\mu L\ mL^{-1}$ for 2 million cells) immediately before running samples through the cytometer, as described in Rumin et al. [56]. The linear relation between neutral lipids and Nile Red fluorescence in Tiso under these conditions was shown by Bougaran et al. [12].

Algae were isolated using a Facs Aria II Cytometer fitted with an automatic sorting device (Excitation laser 488 nm, event rate 10,000 element s^{-1} with a 85-µm nozzle and flow rate = 1, voltage settings: FSC = 220 V; SSC = 324 V; PerCP_Cy5.5(695/40) = 300 V; PE(NileRed)575/26 = 250 V). The gating protocol selected clonal strains with the most lipids (the 10% highest of the population) is 1) FSCA/FSCH et SSCA/SSCH to discriminate doubled cells; 2) histogram of PerCP Cy5.5 to conserved living cells with chlorophyll; and 3) Nile Red histogram (560 nm), the new clonal strains were isolated among the 10% of cells showing the strongest fluorescence intensity. They were collected in 200 µL of culture medium in a sterile 96-well plate (Fig. S3). After the 25 days, algae were suspended in 1 mL of culture medium. After the 10 days of culture, the algal cell concentration had reached 10.10^6 cells/mL and the algae cells were suspended in a fresh 50 mL of culture medium. To finish, after 21 days of culture, algae cells were suspended in 100 mL of culture medium. Algae cells were resuspended for conservation every 3 weeks. After 5 months, a new measure of the amount of lipids was performed by spectrometry with Nile Red dye in stationary phase culture to compare neutral lipid content in the clonal strains.

4.3. Physiological characterization of clonal strains (Tiso-Wt, Tiso-S2M2-Cl1 and Tiso-S2M2-Cl2)

A first characterization of the two clonal strains (Tiso-S2M2-Cl1 and Tiso-S2M2-Cl2) with the highest lipid content was performed in laboratory culture conditions with three biological replicates. Cultures of Tiso were performed in batch mode in 1-L flasks containing Conway medium, but with 0.59 $mMol \cdot L^{-1}$ of nitrate (N/2) [55], and aerated by bubbled 0.22-µm-filtered air. The cultures were maintained at a constant temperature of 21 °C and under a constant irradiance of 80 $\mu mol \cdot m^{-2} \cdot s^{-1}$. In stationary phase, when nitrogen was deficient, two samples were taken (Fig. S1). Absence of nitrogen in the culture medium was confirmed by spectrophotometry at 220 nm as described in [57]. The detailed lipid profiles were analyzed using High Performance Thin Layer Chromatography (HPTLC). The harvested samples (approximately 150 million algae cells in 10 mL) were suspended in 6 mL of Folch solution (methanol-chloroform 2:1). The samples prepared were deposited with an automatic TLC sampler onto a silica plate. Migration was realized in solvent (20 mL hexane; 5 mL diethanol ether; 0.5 mL acetic acid) with standard lipids (Table S5). The revelation was realized in Chromatogram Immersion Device solution (H3Po4-CuSo4) at 180 °C for 30 min. Identification and analysis of bands was performed with WinCats software. These HPTLC analyses were realized in triplicate, examples of chromatograms for each strain are shown in Table S6.

Physiological characterization of Tiso-Wt, Tiso-S2M2-Cl1 and Tiso-S2M2-Cl2 was also performed during a day/night cycle composed of 12 h of light and 12 h of dark, with a maximal light intensity of 1000 $\mu E \cdot m^{-2} \cdot s^{-1}$. Cultures of Tiso were performed with two biological replicates in 2-L flasks containing Conway medium [55] aerated by bubbled 0.22-µm-filtered air. For 30 h, a sample of culture was harvested every 3 h and the following phenotype analyses were performed: i) cell number ii) cell size, iii) nitrogen and carbon biomass, iv) neutral lipid content, v) carbohydrate content and vi) protein content. The number of cells, cell size and neutral lipids were measured by flow cytometry (BD Accury C6). For each sample, 30,000 events were analyzed with a fluidics slow (14 µL/min). A gating filter was used from the peak in FSC and the peak in FL3 (corresponding to chlorophyll pigments present in microalgae). The size of algae cells was estimated with FSC relative to standard beads added to each sample (SureCount particle standard 5 µm). The amount of lipids was measured on FL2 with microalgae dyed with Nile Red, which was added to each sample (1 $\mu L\ mL^{-1}$ for 2 million cells) as described in Rumin et al. [56]. For carbohydrate, protein, organic carbon and nitrogen analyses, the samples were collected on several precombusted GF/C filters (Whatman, 25 mm diameter). For the protein and total carbohydrates in algae, the filters were stored at − 80 °C until analysis. The filter contents were suspended in TBS buffer (Tris, 50 mM; 150 mM NaCl; pH adjusted to 7.6) and sonication (40 Khz; 2 min in ice) were performed to lyse the cells. The amount of protein was measured using a BCA protein kit (ThermoFisher 23225) and total carbohydrate was measured using the Dubois method [58]. For the amount of total carbon, the filters were deposited in limp glass, placed in a steam room and dried at 75 °C for 24 h, then deep-frozen until analysis. Analysis was performed with an elemental analyzer (Thermoelectron). Methionine, aspartic acid and nicotinamide, which have different N and C percentages, were used for calibration.

4.4. Sequencing and assembly of the reference genome of Tiso

Algal cultures were treated with antibiotics (Sigma N°A5955) to minimize bacterial contamination. The DNA extractions for each strain were performed after the 6 months of the clonal strains isolation. DNA was extracted using a phenol-chloroform method, described in detail in [59]. For each clone, 10 µg of DNA (Tiso-Wt, Tiso-S2M2-Cl1 and Tiso-S2M2-Cl2) were used for the sequencing, which was performed at BioGenOuest Nantes platform using an Illumina HiSeq2000. The libraries were built using the mate-pair method (2 × 100 b sequenced for a DNA fragment of 2000 bases on average) and the genomes of each clone (Tiso-Wt, Tiso-S2M2-Cl1 and Tiso-S2M2-Cl2) were sequenced independently in different lanes. The raw sequencing data obtained were filtered to conserve only reads with a sufficient quality: First, the raw data were treated using CutAdapt software [60] to eliminate Illumina residual adapters (Anywhere adapters in sequence). Second, the reads were filtered with a homemade script used to conserve only read pairs with a mean sequencing quality score higher than Q30, a length higher than 75 bases for both reads. After filtering was complete, 75% of read pairs were conserved (94, 71, 105 millions of read pairs from Tiso-Wt, Tiso-S2M2-Cl1 and Tiso-S2M2-Cl2, respectively, Table S1). The reference genome of the Tiso species was established by the sum of

all the sequences. Assembly was performed using CLC Assembly Cell 4.0 (launched with default parameters with an insert-size between 1500 and 3000 bases), 13,114 contigs were obtained (for an overall size of 53.73 Mb). N = 33 contigs were deleted because they were not associating with Tiso but with prokaryotic origins (bacteria or virus sequences were identified in these contigs by BLAST on the nr database). A construction of large contigs exploiting the mate-pair method was performed using SSPACE software (launched with default parameters) [61] and 7662 contigs were obtained (N50 score: 10,571 bases). Finally, this first draft reference genome of Tiso had an estimated size of 54.38 Mb.

4.5. Gene identification and annotation

Identification of genes was performed using the MAKER2 training and annotation pipeline [62]. MAKER2 was employed using several tools: a) BLASTn [63] with *Arabidopsis thaliana* coding sequences (TAIR-10 [64]); b) Tophat2 [65] for aligning the RNAseq data available for Tiso [14]; c) BLASTx for aligning all CEGMA (core eukaryotic genes) [66]; and d) Augustus [67] for models based on the gene structures of *Chlamydomonas reinhardtii*. All results of these different programs were combined by Maker 2, and a total of 20,582 genes were identified. Functional annotation was performed with BLASTx [63] on the non-redundant protein database (nr), with a similarity score higher than 30%. In addition, InterProScan software was used to identify functional domains [68].

4.6. Polymorphism identification

Sequence reads obtained independently from each clone (Tiso-Wt, Tiso-S2M2-Cl1 and Tiso-S2M2-Cl2) were mapped onto the Tiso reference genome using Mosaik software (Percentage mismatch: 0.15; Hash size: 15; Insert size: 1500–300 bases; Algorithm: very sensitive) [69]. On average, 80% of read-pairs were aligned on the reference genome, 14% were aligned in multi-loci and 6% were not aligned (Table S2). First of all, polymorphisms were identified to determine polyallelic loci from reads mapped independently on reference genome for each strain. An extrapolation at the whole genome level was realized for each strain to compare their numbers. Secondly, polyallelic loci detected in each strain and monoallelic loci in each strain were compared to identify polymorphic loci between strains. A search for SNPs and small indels was performed using Freebayes software and these alignment data [41] (minimal total coverage: 40; minimal allele coverage: 10; pooled-continuous option because we did not know the number of individuals in the Wt strain). In addition, a search for SNP and small indels was also performed using DiscoSNP software [70] directly from the reads data (Size of Kmers:21; minimal coverage per read set: 10). DiscoSNP is not biased by an alignment step but it produces more errors in polymorphism positions in the genome. An SNP or small indel in or between strains was validated if it was identified with both programs (74% of SNPs identified by Freebayes and 67% of SNPs identified by DiscoSNP). Putative impacts of SNPs and small indels on gene function were evaluated using SnpEff [71]. Only the missense variant, codon STOP variant and frameshift variant were considered to have a potential impact on genes. The large insertion polymorphism was identified by alignment of the genomes of each strain on the consensus reference using a parse script to extract each genomic sequence specific to one or two strains. In complement, MiNTheGap software [72] (size of Kmers: 21; minimal coverage 10) was used. The origin of the large insertion sequences identified between the strains was searched for on several databases (TAIR-10 [64], Uniprot [73], RepBase [45]) by similarity (BLAST software, e-value E − 20).

All data (annotation, contigs, proteins, and raw data) are available in the SEANOE databank DOI: https://doi.org/10.17882/47171. Raw data for Tiso-Wt is available in SRR3156597 in NCBI. Supplementary data associated with this article can be found in the online version, at https://doi.org/10.1016/j.algal.2017.10.017.

Acknowledgements

We acknowledge the ABiMS platform and IFREMER-Caparmor platform for bioinformatics server maintenance. We acknowledge the BioGenOuest Nantes genomic platform for the sequencing and CytoCell for use of the cytometer.

Funding

This work was funded by the Agence National de la Recherche (ANR-12-BIME-0004), Facteur 4 project and the INRIA Project Lab Algae *in silico*. C. Baroukh was supported by a “Contrat Jeune Scientifique” (CJS) INRA-INRIA fellowship. There are no non-financial competing interests. The funders had no role in study design, data collection and analysis, the decision to publish or preparation of the manuscript.

Author contributions

Conceived and designed the experiments: GC, CB, CR, GB. Performed the experiments on mircoalgae: GC, CB, CR, NS. Performed sequencing genomes: GC, NS, LDB. Analyzed the data: GC, CB, CR. Wrote and revised the publication: GC, CB, CR, LDB, NS, GB.

References

[1] C.B. Field, M.J. Behrenfeld, J.T. Randerson, P. Falkowski, Primary production of the biosphere: integrating terrestrial and oceanic components, Science 281 (1998) 237–240, http://dx.doi.org/10.1126/science.281.5374.237.

[2] L. Guidi, S. Chaffron, L. Bittner, D. Eveillard, A. Larhlimi, S. Roux, Y. Darzi, S. Audic, L. Berline, J.R. Brum, L.P. Coelho, J.C.I. Espinoza, S. Malviya, S. Sunagawa, C. Dimier, S. Kandels-Lewis, M. Picheral, J. Poulain, S. Searson, Tara Oceans Consortium Coordinators, L. Stemmann, F. Not, P. Hingamp, S. Speich, M. Follows, L. Karp-Boss, E. Boss, H. Ogata, S. Pesant, J. Weissenbach, P. Wincker, S.G. Acinas, P. Bork, C. de Vargas, D. Iudicone, M.B. Sullivan, J. Raes, E. Karsenti, C. Bowler, G. Gorsky, Plankton networks driving carbon export in the oligotrophic ocean, Nature 532 (7600) (2016) 16942 (advance online publication), https://doi.org/10.1038/nature16942.

[3] P. Bork, C. Bowler, C. de Vargas, G. Gorsky, E. Karsenti, P. Wincker, Tara Oceans studies plankton at planetary scale, Science 348 (2015) 873, http://dx.doi.org/10.1126/science.aac5605.

[4] M.D. Guiry, How many species of algae are there? J. Phycol. 48 (2012) 1057–1063, http://dx.doi.org/10.1111/j.1529-8817.2012.01222.x.

[5] M. Koller, A. Muhr, G. Braunegg, Microalgae as versatile cellular factories for valued products, Algal Res. 6 (Part A) (2014) 52–63, http://dx.doi.org/10.1016/j.algal.2014.09.002.

[6] D. Stengel, S. Connan, Marine algae: a source of biomass for biotechnological applications, in: D.B. Stengel, S. Connan (Eds.), Nat. Prod. Mar. Algae, Springer, New York, 2015, pp. 1–37, , http://dx.doi.org/10.1007/978-1-4939-2684-8_1.

[7] C. Enzing, M. Ploeg, M. Barbosa, L. Sijtsma, Microalgae-based Products for the Food and Feed Sector: An Outlook for Europe, Publications Office of the European Union, 2014, http://publications.jrc.ec.europa.eu/repository/handle/111111111/31366 , Accessed date: 4 March 2016.

[8] M. Vigani, C. Parisi, E. Rodríguez-Cerezo, M.J. Barbosa, L. Sijtsma, M. Ploeg, C. Enzing, Food and feed products from micro-algae: market opportunities and challenges for the EU, Trends Food Sci. Technol. 42 (2015) 81–92, http://dx.doi.org/10.1016/j.tifs.2014.12.004.

[9] A.W.D. Larkum, I.L. Ross, O. Kruse, B. Hankamer, Selection, breeding and engineering of microalgae for bioenergy and biofuel production, Trends Biotechnol. 30 (2012) 198–205, http://dx.doi.org/10.1016/j.tibtech.2011.11.003.

[10] L.A. Horrocks, Y.K. Yeo, Health benefits of docosahexaenoic acid (DHA), Pharmacol. Res. 40 (1999) 211–225, http://dx.doi.org/10.1006/phrs.1999.0495.

[11] J.L. Guil-Guerrero, Stearidonic acid (18:4n-3): metabolism, nutritional importance, medical uses and natural sources, Eur. J. Lipid Sci. Technol. 109 (2007) 1226–1236, http://dx.doi.org/10.1002/ejlt.200700207.

[12] G. Bougaran, C. Rouxel, N. Dubois, R. Kaas, S. Grouas, E. Lukomska, J.-R. Le Coz, J.-P. Cadoret, Enhancement of neutral lipid productivity in the microalga *Isochrysis* affinis *Galbana* (T-Iso) by a mutation-selection procedure, Biotechnol. Bioeng. 109 (11) (2012) 2737–2745 https://doi.org/10.1002/bit.24560.

[13] C. Rouxel, G. Bougaran, S. Doulin Grouas, N. Dubois, J.-P. Cadoret, Novel *Isochrysis* sp. Tahitan Clone and Uses Therefore, EP 11006712.1, (2011).

[14] G. Carrier, M. Garnier, L. Le Cunff, G. Bougaran, I. Probert, C. De Vargas, E. Corre, J.-P. Cadoret, B. Saint-Jean, 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 (2014) e86889, , http://dx.doi.org/

10.1371/journal.pone.0086889.
[15] M. Garnier, G. Carrier, H. Rogniaux, E. Nicolau, G. Bougaran, B. Saint-Jean, J.P. Cadoret, Comparative proteomics reveals proteins impacted by nitrogen deprivation in wild-type and high lipid-accumulating mutant strains of *Tisochrysis lutea*, J. Proteome 105 (2014) 107–120, http://dx.doi.org/10.1016/j.jprot.2014.02.022.
[16] T.T.Y. Doan, B. Sivaloganathan, J.P. Obbard, Screening of marine microalgae for biodiesel feedstock, Biomass Bioenergy 35 (2011) 2534–2544, http://dx.doi.org/10.1016/j.biombioe.2011.02.021.
[17] A.N. Barry, S.R. Starkenburg, R.T. Sayre, Strategies for optimizing algal biology for enhanced biomass production, Front. Energy Res. 3 (2015), http://dx.doi.org/10.3389/fenrg.2015.00001.
[18] J.N. Rosenberg, N. Kobayashi, A. Barnes, E.A. Noel, M.J. Betenbaugh, G.A. Oyler, Comparative analyses of three *Chlorella* species in response to light and sugar reveal distinctive lipid accumulation patterns in the microalga *C. sorokiniana*, PLoS ONE 9 (2014) e92460, , http://dx.doi.org/10.1371/journal.pone.0092460.
[19] C.J. Unkefer, R.T. Sayre, J.K. Magnuson, D.B. Anderson, I. Baxter, I.K. Blaby, J.K. Brown, M. Carleton, R.A. Cattolico, T. Dale, T.P. Devarenne, C.M. Downes, S.K. Dutcher, D.T. Fox, U. Goodenough, J. Jaworski, J.E. Holladay, D.M. Kramer, A.T. Koppisch, M.S. Lipton, B.L. Marrone, M. McCormick, I. Molnár, J.B. Mott, K.L. Ogden, E.A. Panisko, M. Pellegrini, J. Polle, J.W. Richardson, M. Sabarsky, S.R. Starkenburg, G.D. Stormo, M. Teshima, S.N. Twary, P.J. Unkefer, J.S. Yuan, J.A. Olivares, Review of the algal biology program within the National Alliance for Advanced Biofuels and Bioproducts, Algal Res. 22 (2017) 187–215, http://dx.doi.org/10.1016/j.algal.2016.06.002.
[20] K. Manandhar-Shrestha, M. Hildebrand, Development of flow cytometric procedures for the efficient isolation of improved lipid accumulation mutants in a *Chlorella* sp. microalga, J. Appl. Phycol. 25 (2013) 1643–1651, http://dx.doi.org/10.1007/s10811-013-0021-8.
[21] M.S. Cooper, W.R. Hardin, T.W. Petersen, R.A. Cattolico, Visualizing “green oil” in live algal cells, J. Biosci. Bioeng. 109 (2010) 198–201, http://dx.doi.org/10.1016/j.jbiosc.2009.08.004.
[27] N.S. Shifrin, S.W. Chisholm, Phytoplankton lipids: interspecific differences and effects of nitrate, silicate and light-dark cycles, J. Phycol. 17 (1981) 374–384, http://dx.doi.org/10.1111/j.1529-8817.1981.tb00865.x.
[28] M.J. Dring, Photocontrol of development in algae, Annu. Rev. Plant Physiol. Plant Mol. Biol. 39 (1988) 157–174, http://dx.doi.org/10.1146/annurev.pp.39.060188.001105.
[29] T. Lacour, A. Sciandra, A. Talec, P. Mayzaud, O. Bernard, Diel variations of carbohydrates and neutral lipids in nitrogen-sufficient and nitrogen-starved cyclostat cultures of *Isochrysis* sp.1, J. Phycol. 48 (2012) 966–975, http://dx.doi.org/10.1111/j.1529-8817.2012.01177.x.
[30] M.J. Chaisson, D. Brinza, P.A. Pevzner, De novo fragment assembly with short mate-paired reads: does the read length matter? Genome Res. 19 (2009) 336–346, http://dx.doi.org/10.1101/gr.079053.108.
[31] E.M. Bendif, I. Probert, D.C. Schroeder, C. de Vargas, 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. 25 (2013) 1763–1776, http://dx.doi.org/10.1007/s10811-013-0037-0.
[32] L. Bittner, A. Gobet, S. Audic, S. Romac, E.S. Egge, S. Santini, H. Ogata, I. Probert, B. Edvardsen, C. de Vargas, Diversity patterns of uncultured Haptophytes unravelled by pyrosequencing in Naples Bay, Mol. Ecol. 22 (2013) 87–101, http://dx.doi.org/10.1111/mec.12108.
[33] M. Simon, P. López-García, D. Moreira, L. Jardillier, New haptophyte lineages and multiple independent colonizations of freshwater ecosystems, Environ. Microbiol. Rep. 5 (2013) 322–332, http://dx.doi.org/10.1111/1758-2229.12023.
[34] B.T. Hovde, C.R. Deodato, H.M. Hunsperger, S.A. Ryken, W. Yost, R.K. Jha, J. Patterson, R.J. Monnat Jr., S.B. Barlow, S.R. Starkenburg, R.A. Cattolico, Genome sequence and transcriptome analyses of *Chrysochromulina tobin*: metabolic tools for enhanced algal fitness in the prominent order Prymnesiales (Haptophyceae), PLoS Genet. 11 (2015) e1005469, , http://dx.doi.org/10.1371/journal.pgen.1005469.
[35] B.A. Read, J. Kegel, M.J. Klute, A. Kuo, S.C. Lefebvre, F. Maumus, C. Mayer, J. Miller, A. Monier, A. Salamov, J. Young, M. Aguilar, J.-M. Claverie, S. Frickenhaus, K. Gonzalez, E.K. Herman, Y.-C. Lin, J. Napier, H. Ogata, A.F. Sarno, J. Shmutz, D. Schroeder, C. de Vargas, F. Verret, P. von Dassow, K. Valentin, Y. Van de Peer, G. Wheeler, Emiliania huxleyi Annotation Consortium, J.B. Dacks, C.F. Delwiche, S.T. Dyhrman, G. Glöckner, U. John, T. Richards, A.Z. Worden, X. Zhang, I.V. Grigoriev, Pan genome of the phytoplankton Emiliania underpins its global distribution, Nature 499 (2013) 209–213, http://dx.doi.org/10.1038/nature12221.
[36] A. Roulin, P.L. Auer, M. Libault, J. Schlueter, A. Farmer, G. May, G. Stacey, R.W. Doerge, S.A. Jackson, The fate of duplicated genes in a polyploid plant genome, Plant J. 73 (2013) 143–153, http://dx.doi.org/10.1111/tpj.12026.
[37] D.E. Soltis, P.S. Soltis, Polyploidy: recurrent formation and genome evolution, Trends Ecol. Evol. 14 (1999) 348–352, http://dx.doi.org/10.1016/S0169-5347(99)01638-9.
[38] K.L. Adams, J.F. Wendel, Polyploidy and genome evolution in plants, Curr. Opin. Plant Biol. 8 (2005) 135–141, http://dx.doi.org/10.1016/j.pbi.2005.01.001.
[39] J.D. Thompson, R. Lumaret, The evolutionary dynamics of polyploid plants: origins, establishment and persistence, Trends Ecol. Evol. 7 (1992) 302–307, http://dx.doi.org/10.1016/0169-5347(92)90228-4.
[40] V. Bansal, V. Bafna, HapCUT: an efficient and accurate algorithm for the haplotype assembly problem, Bioinformatics 24 (2008) i153–i159, http://dx.doi.org/10.1093/bioinformatics/btn298.
[41] E. Garrison, G. Marth, Haplotype-based Variant Detection From Short-Read Sequencing, ArXiv12073907 Q-Bio, http://arxiv.org/abs/1207.3907, (2012) , Accessed date: 20 July 2016.
[42] M. Garnier, G. Bougaran, M. Pavlovic, J.-B. Berard, G. Carrier, A. Charrier, F. Le Grand, E. Lukomska, C. Rouxel, N. Schreiber, J.-P. Cadoret, H. Rogniaux, B. Saint-Jean, Use of a lipid rich strain reveals mechanisms of nitrogen limitation and carbon partitioning in the haptophyte *Tisochrysis lutea*, Algal Res. 20 (2016) 229–248, http://dx.doi.org/10.1016/j.algal.2016.10.017.
[43] R.J. Greenspan, Selection, gene interaction, and flexible gene networks, Cold Spring Harb. Symp. Quant. Biol, 2009 (sqb.2009.74.029), https://doi.org/10.1101/sqb.2009.74.029.
[44] M. Kimura, The Neutral Theory of Molecular Evolution, Cambridge University Press, Cambridge, 1983http://ebooks.cambridge.org/ref/id/CBO9780511623486 (accessed July 28, 2016).
[45] W. Bao, K.K. Kojima, O. Kohany, Repbase update, a database of repetitive elements in eukaryotic genomes, Mob. DNA 6 (2015) 11, , http://dx.doi.org/10.1186/s13100-015-0041-9.
[49] E. Casacuberta, J. González, The impact of transposable elements in environmental adaptation, Mol. Ecol. 22 (2013) 1503–1517, http://dx.doi.org/10.1111/mec.12170.
[50] D. Lisch, How important are transposons for plant evolution? Nat. Rev. Genet. 14 (2013) 49–61, http://dx.doi.org/10.1038/nrg3374.
[51] O. Rey, E. Danchin, M. Mirouze, C. Loot, S. Blanchet, Adaptation to global change: a transposable element–epigenetics perspective, Trends Ecol. Evol. 31 (2016) 514–526, http://dx.doi.org/10.1016/j.tree.2016.03.013.
[52] C. Feschotte, Transposable elements and the evolution of regulatory networks, Nat. Rev. Genet. 9 (2008) 397–405, http://dx.doi.org/10.1038/nrg2337.
[53] C.D. Hirsch, N.M. Springer, Transposable element influences on gene expression in plants, Biochim. Biophys. Acta Gene Regul. Mech. (n.d.). doi:https://doi.org/10.1016/j.bbagrm.2016.05.010.
[54] J. Stapley, A.W. Santure, S.R. Dennis, Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species, Mol. Ecol. 24 (2015) 2241–2252, http://dx.doi.org/10.1111/mec.13089.
[55] P. Walne, Experiments in the Large-scale Culture of the Larvae of *Ostrea edulis* L. Minist. Agric. Fish. Invest., London, 1966, p. 53970 http://openlibrary.org/b/OL5568472M/Experiments_in_the_large-scale_culture_of_the_larvae_of_Ostrea_edulis_L.
[56] J. Rumin, H. Bonnefond, B. Saint-Jean, C. Rouxel, A. Sciandra, O. Bernard, J.-P. Cadoret, G. Bougaran, The use of fluorescent Nile red and BODIPY for lipid measurement in microalgae, Biotechnol. Biofuels 8 (2015) 42, http://dx.doi.org/10.1186/s13068-015-0220-4.
[57] Y. Collos, F. Mornet, A. Sciandra, N. Waser, A. Larson, P.J. Harrison, An optical method for the rapid measurement of micromolar concentrations of nitrate in marine phytoplankton cultures, J. Appl. Phycol. 11 (1999) 179–184, http://dx.doi.org/10.1023/A:1008046023487.
[58] M. DuBois, K.A. Gilles, J.K. Hamilton, P.A. Rebers, F. Smith, Colorimetric method for determination of sugars and related substances, Anal. Chem. 28 (1956) 350–356, http://dx.doi.org/10.1021/ac60111a017.
[59] Z. Hu, X. Zeng, A. Wang, C. Shi, D. Duan, An efficient method for DNA isolation from red algae, J. Appl. Phycol. 16 (2004) 161–166, http://dx.doi.org/10.1023/B:JAPH.0000048456.26639.1a.
[60] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, EMBnetjournal 17 (1) (2011) (Gener. Seq. Data Anal.), http://journal.embnet.org/index.php/embnetjournal/article/view/200 (accessed January 1, 2011).
[61] M. Boetzer, C.V. Henkel, H.J. Jansen, D. Butler, W. Pirovano, Scaffolding pre-assembled contigs using SSPACE, Bioinformatics 27 (2011) 578–579, http://dx.doi.org/10.1093/bioinformatics/btq683.
[62] C. Holt, M. Yandell, MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects, BMC Bioinf. 12 (2011) 491, http://dx.doi.org/10.1186/1471-2105-12-491.
[63] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410, http://dx.doi.org/10.1016/S0022-2836(05)80360-2.
[64] T.Z. Berardini, L. Reiser, D. Li, Y. Mezheritsky, R. Muller, E. Strait, E. Huala, The arabidopsis information resource: making and mining the “gold standard” annotated reference plant genome, Genesis 53 (2015) 474–485, http://dx.doi.org/10.1002/dvg.22877.
[65] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S.L. Salzberg, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions, Genome Biol. 14 (2013) R36, http://dx.doi.org/10.1186/gb-2013-14-4-r36.
[66] G. Parra, K. Bradnam, I. Korf, CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes, Bioinformatics 23 (2007) 1061–1067, http://dx.doi.org/10.1093/bioinformatics/btm071.
[67] M. Stanke, M. Diekhans, R. Baertsch, D. Haussler, Using native and syntenically mapped cDNA alignments to improve de novo gene finding, Bioinformatics 24 (2008) 637–644, http://dx.doi.org/10.1093/bioinformatics/btn013.
[68] E.M. Zdobnov, R. Apweiler, InterProScan – an integration platform for the signature-recognition methods in InterPro, Bioinformatics 17 (2001) 847–848, http://dx.doi.org/10.1093/bioinformatics/17.9.847.
[69] W.-P. Lee, M. Stromberg, A. Ward, C. Stewart, E. Garrison, G.T. Marth, MOSAIK: A Hash-based Algorithm for Accurate Next-Generation Sequencing Read Mapping, ArXiv13091149 Q-Bio, http://arxiv.org/abs/1309.1149, (2013) , Accessed date: 29 January 2014.
[70] R. Uricaru, G. Rizk, V. Lacroix, E. Quillery, O. Plantard, R. Chikhi, C. Lemaitre, P. Peterlongo, Reference-free detection of isolated SNPs, Nucleic Acids Res. (2014) gku1187, , http://dx.doi.org/10.1093/nar/gku1187.

[71] P. Cingolani, A. Platts, L.L. Wang, M. Coon, T. Nguyen, L. Wang, S.J. Land, X. Lu, D.M. Ruden, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff, Fly (Austin) 6 (2012) 80–92, http://dx.doi.org/10.4161/fly.19695.

[72] G. Rizk, A. Gouin, R. Chikhi, C. Lemaitre, MindTheGap: integrated detection and assembly of short and long insertions, Bioinformatics 30 (2014) 3451–3457, http://dx.doi.org/10.1093/bioinformatics/btu545.

[73] T.U. Consortium, UniProt: a hub for protein information, Nucleic Acids Res. 43 (2015) D204–D212, http://dx.doi.org/10.1093/nar/gku989.

[74] K.V. Efimova, T.Y. Orlova, V.A. Brykov, Molecular genetic identification of a new *Tisochrysis lutea* (Bendif et Probert, 2013) strain isolated from the Russian coastal waters of the Sea of Japan, Microbiology 85 (2016) 325–332, http://dx.doi.org/10.1134/S0026261716030036.