

RESEARCH ARTICLE

Phenotype stability and dynamics of transposable elements in a strain of the microalga *Tisochrysis lutea* with improved lipid traits

Jérémy Berthelier^{1,2,3*}, Bruno Saint-Jean^{1,4}, Nathalie Casse², Gaël Bougaran^{1,5}, Grégory Carrier^{1,4}

1 Laboratory PBA, Ifremer, Nantes, France, **2** Laboratory of Biology of Organisms Stress Health Environment (BiOSSE), Le Mans University, Le Mans, France, **3** Plant Epigenetics Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa, Japan, **4** Laboratory PHYTOX GENALG, Ifremer, Nantes, France, **5** Laboratory PHYTOX PHYSALG, Ifremer, Nantes, France

* berthelier.j@laposte.net



OPEN ACCESS

Citation: Berthelier J, Saint-Jean B, Casse N, Bougaran G, Carrier G (2023) Phenotype stability and dynamics of transposable elements in a strain of the microalga *Tisochrysis lutea* with improved lipid traits. PLoS ONE 18(4): e0284656. <https://doi.org/10.1371/journal.pone.0284656>

Editor: Balamurugan Srinivasan, Bharathidasan University, INDIA

Received: November 2, 2022

Accepted: April 5, 2023

Published: April 27, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0284656>

Copyright: © 2023 Berthelier et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are publicly available from the SEANO database (<https://doi.org/10.17882/83584>).

Abstract

Microalgal domestication is an expanding research field that aims to multiply and accelerate the potential of microalgae for various biotechnological purposes. We investigated the stability of improved lipid traits and genetic changes of a domesticated strain of the haptophyte *Tisochrysis lutea*, TisoS2M2, previously obtained by a mutation-selection improvement program. After 7 years of maintenance, TisoS2M2 still displayed improved lipid traits compared with the native strain, demonstrating that a mutation-selection improvement program is suitable for obtaining a domesticated strain with stable, improved phenotype over time. We identified specific genetic variations between the native and domesticated strains and focused on the dynamics of transposable elements (TEs). DNA transposons mainly caused specific TE indels of the domesticated strain TisoS2M2, and some specific TE indels may have impacted genes associated to the neutral lipid pathway. We revealed transposition events for TEs in *T. lutea* and discussed on the potential role of the improvement program on their activity.

1. Background

Microalgae are regarded as a promising resource to tackle the increase in global population and depletion of natural resources [1] and have been studied for multiple applications, such as health, feedstock, cosmetics, bioremediation, and biofuels [2, 3]. Indeed, they constitute a sink of valuable molecules, and their short generation times make them interesting candidates for industrial purposes [2, 3]. The haptophyte *Tisochrysis lutea* was originally isolated from tropical seawater (Tahiti, French Polynesia). This microalga is currently used in aquaculture and is widely studied in the ecophysiology and biotechnology fields [4–9]. Its high content of neutral lipids makes it a potential source of biofuels [10–13]. *T. lutea* is also considered a promising feedstock for health compounds because this microalga accumulates high contents of xanthophylls (such as

Funding: This work was funded by the French Research Institute for Exploitation of the Sea (IFREMER), the French Region of Pays de la Loire with the Atlantic Microalgae program, the DynAlgue ANR JCJC, and LANCOM Paris Scientifique 2020 Pays de La Loire. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

fucoxanthin) and polyunsaturated fatty acid (PUFA) (such as docosahexaenoic acid, DHA) that play roles in the health and development of aquatic organisms. Fucoxanthin is associated with numerous health benefits, including antioxidant and anti-inflammatory effects. DHA improves the cardiovascular condition and fetal development, making this microalga potentially attractive to be used in various medicinal applications [14, 15]. Despite the promising use of microalgae such as *T. lutea* in biotechnological applications, several factors must be overcome before microalgal resources become economically viable on an industrial scale [16]. One current issue pointed out by the scientific community is the lack of knowledge on the stability of the improved traits of domesticated strains over time [16–18]. While the domestication of crops started thousands of years ago through selection and breeding for thousands of generations [19], the domestication of microalgae has a history of only two decades. Unlike crops, the sexual reproduction of microalgae is mostly uncontrolled, and their domestication is instead performed by improvement programs, the most common of which are genetic engineering [20, 21], controlled evolution [22, 23], selection [24, 25], and mutation-selection [10, 26]. While domesticated strains produced right after these improvement programs demonstrate improved traits, the stability of the improved traits over time has, to our knowledge, never been investigated after years [17, 18]. Indeed, most studies assumed that phenotypic traits measured in the laboratory are stable over time. They underestimate the possibility that the domesticated strains could turn back to the wildtype phenotype or have a drift of the improved traits [17, 18]. Such an evaluation is of major importance in the context of industrial applications [16]. In addition, while they encode for the genetic novelties giving rise to the improved domesticated traits, genetic changes in the genome of domesticated strains have been few investigated [22, 27–32]. For instance, previous studies on the genome of domesticated strains of *Chlamydomonas reinhardtii* identified new single-nucleotide polymorphisms (SNPs) and DNA deletions/insertions (indels) compared with the native strain [22, 29]. For the haptophyte *T. lutea*, we have previously investigated the genetic changes in two clonal domesticated strains compared with the native strain. Among the indels, DNA sequence similarity analysis suggested the involvement of transposable elements (TEs) [11]. TEs are DNA sequences found in eukaryotic and prokaryotic genomes that can move and spread [33, 34]. In eukaryotes, TEs are commonly grouped into two classes according to their mechanism of mobility: class I retrotransposons move by a copy-paste mechanism via RNA intermediaries; class II DNA transposons move by cut-paste mechanisms [34]. Although TE indels are mostly harmful or neutral to the host, they can also participate in the rise of evolved phenotypic traits [35, 36]. For plant crops, TEs have been found to promote various domesticated traits [35, 37, 38]. Because of the major role of TEs in the domestication of organisms, we previously performed an accurate annotation of TEs in *T. lutea*. We identified two DNA transposon families, hAT/*Ace* and Tc1-Mariner/*Luffy*, which can be expressed and get their transposase processed by the cell [39]. We suggested that these DNA transposon families may participate in the genome dynamics of the *T. lutea*.

In the present study, we evaluated the phenotypic stability over time of the domesticated *T. lutea* strain TisoS2M2, previously obtained by a mutation-selection program, which has a higher total fatty acid (TFA) content compared to the native strain TisoArg [10]. We next investigated specific genetic changes in the domesticated strain compared with the native strain by focusing on the dynamics of TEs and their potential role in the improved phenotype of TisoS2M2.

2. Results

2.1 TisoS2M2 has stable and improved lipids trait over time

The domesticated strain TisoS2M2 had previously been obtained by submitting the native TisoArg strain to a mutation-selection improvement program [10] (Fig 1A). This experiment

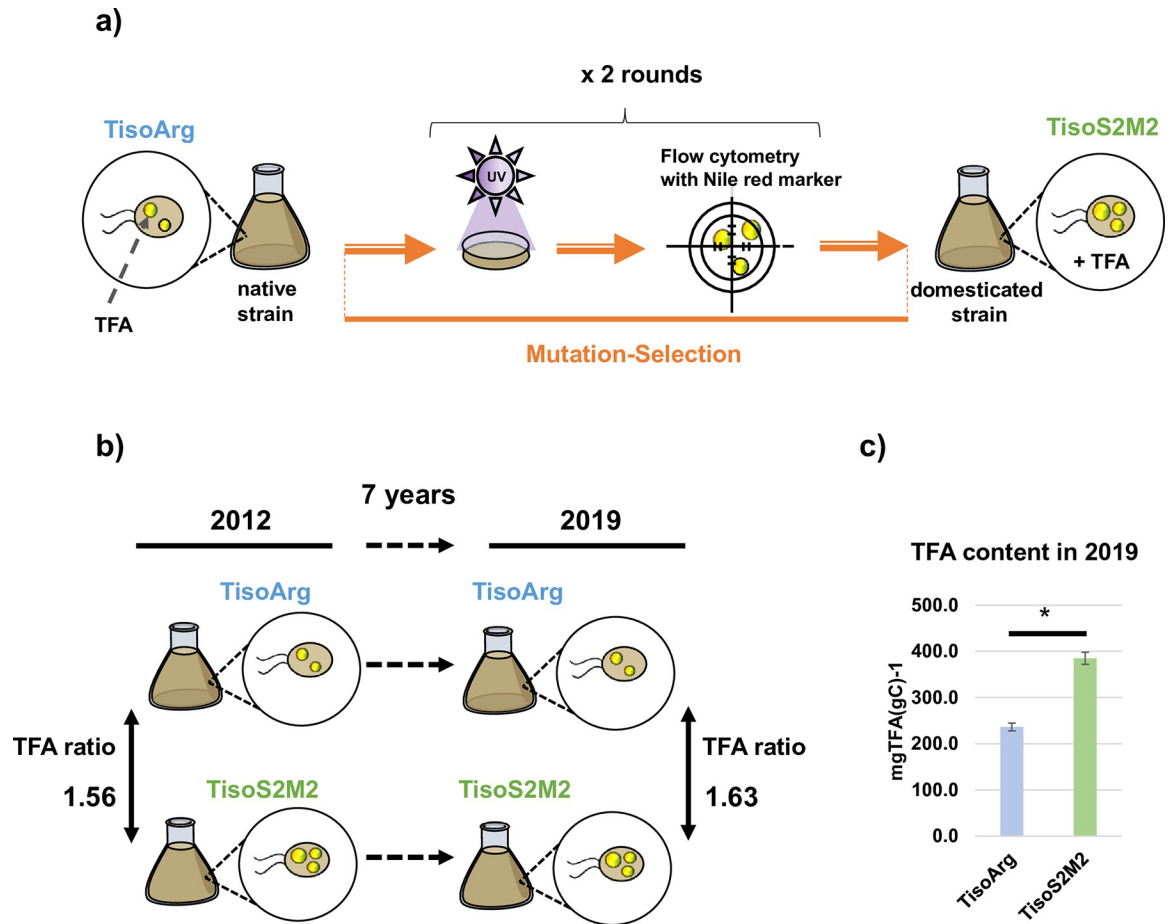


Fig 1. Strategy of the improvement program and comparison of the total fatty acid (TFA) content between the domesticated strain TisoS2M2 strain and the native strain TisoArg. a) Illustration of the strategy of the improvement program used in 2012 [10] to obtain TisoS2M2. b) Comparison of TFA ratio obtained in 2012 [10] and 2019 (this study) for TisoArg and TisoS2M2. c) Comparison of the TFA content of TisoArg and TisoS2M2 measured in 2019. Bars represent the means of two biological replicates \pm SD. *, $p < 0.05$ by t-test.

<https://doi.org/10.1371/journal.pone.0284656.g001>

was conducted in 2012 and the TFA content of TisoS2M2 was 1.56 times higher than that of TisoArg [10] (Fig 1B). Here, the TFA content of both strains was again measured in 2019. A mean of $236.4 \pm 8.4 \text{ mgTFA(gC)}^{-1}$ for the wildtype TisoArg and $385.1 \pm 13.1 \text{ mgTFA(gC)}^{-1}$ for the domesticated TisoS2M2 were found, which were significantly different (Fig 1C; S1 Data). In 2019, the TFA content of TisoS2M2 was 1.63 times higher than TisoArg (Fig 1B; S1 Data). Interestingly, the proportions of monounsaturated fatty acid (MUFA), saturated fatty acids (SFA) and PUFA among TFA were similar between both strains (S1 Data; S1 Fig). Among PUFA, we found that TisoS2M2 had a significantly higher content of $\Sigma\Omega 3$ and $\Sigma\Omega 6$ compared with TisoArg. We measured a $\Sigma\Omega 3$ and $\Sigma\Omega 6$ content of $13.6 \pm 0.6 \text{ mg}\Sigma\Omega 3(\text{gC})^{-1}$ and $11.0 \pm 0.3 \text{ mg}\Sigma\Omega 6(\text{gC})^{-1}$ in TisoArg, but $24.9 \pm 0.2 \text{ mg}\Sigma\Omega 3(\text{gC})^{-1}$ and $19.3 \pm 0.6 \text{ mg}\Sigma\Omega 6(\text{gC})^{-1}$ in TisoS2M2 (S1 Data; S1B Fig). The contents in $\Sigma\Omega 3$ and $\Sigma\Omega 6$ were 1.83 and 1.76 times higher in TisoS2M2 compared with the native TisoArg (S1 Data). Similar growth rates were found for both strains in 2019 with $0.37 \pm 0.03 \mu_{\text{max}} (\text{day}^{-1})$ and $0.35 \pm 0.08 \mu_{\text{max}} (\text{day}^{-1})$ for TisoArg and TisoS2M2, respectively, under the chosen growing condition (S1 Data; S1C Fig). In conclusion, the TFA content of TisoS2M2 continued to be superior to TisoArg.

2.2 Prediction of specific polymorphisms in the native and domesticated strains

In addition to the phenotypic analyses, we investigated specific genetic variants in the native and domesticated strains. The genomes of TisoArg and TisoS2M2 were sequenced with Illumina technology in 2018. Overall, the paired-end reads of each strain covered 99% of the reference genome assembly of *T. lutea* CCAP927/14. Read depths of 58.3× and 59.0× were obtained for TisoArg and TisoS2M2, respectively. First, the detection of SNPs in TisoArg and TisoS2M2 was performed. We found 52,122 SNPs in TisoArg and 47,300 SNPs in TisoS2M2. Among them, 33,671 SNPs were common to both strains, while 18,451 and 13,629 SNPs were specific to TisoArg and TisoS2M2, respectively (Fig 2A, S2 Data).

Next, we investigated specific indels in TisoArg and TisoS2M2 and focused on the ones caused by TEs, because we revealed in a previous study that the *T. lutea* genome contains autonomous TE families encoding for transposase [39]. We designed a bioinformatic pipeline to predict specific TE indels in TisoArg and TisoS2M2 genomes (S3 Data). Combining the indel predictions of the tools Pindel and Breakdancer, 57 and 53 deletions were overlapping and considered to be shared [Fig 2B; S4 Data]. However, we detected 20 TisoArg-specific deletions and 34 TisoS2M2-specific deletions. Focusing on TisoS2M2-specific predictions, we identified two high confidence deletion events detected by both tools. The first was a deletion of 862 bp located at contig 93 (S4 and S5 Data). The edges of the predicted deletion sequence display terminal inverted repeats (TIRs) identical to those of the DNA transposon family *Ace*

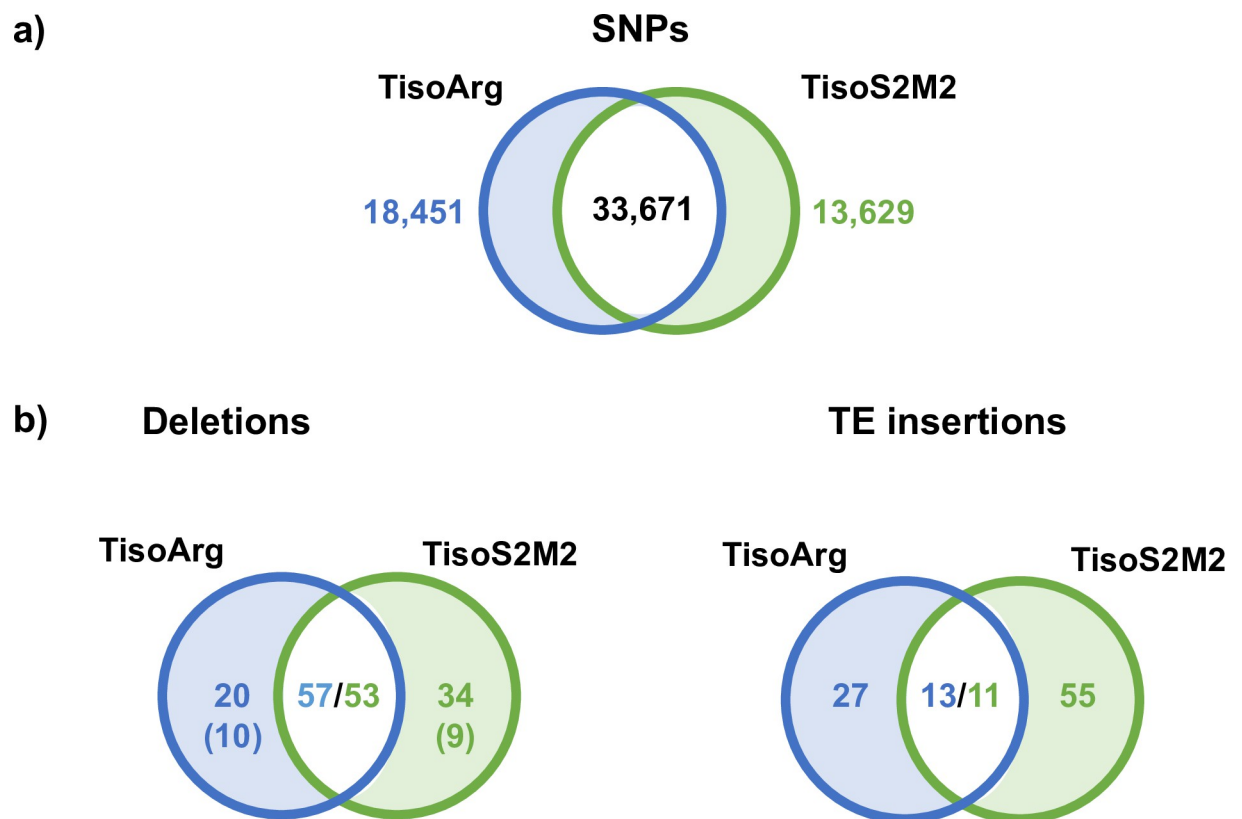


Fig 2. a) Shared and specific SNPs predicted for the wildtype TisoArg and the domesticated strain TisoS2M2 using Illumina reads. b) Shared and specific deletions (left) and insertions (right) predicted for the wildtype TisoArg and the domesticated strain TisoS2M2 using Illumina reads. The number of TEs associated with the deletion events is given in parentheses.

<https://doi.org/10.1371/journal.pone.0284656.g002>

[39] (S5 Data). However, the short length of the sequence indicated that it was caused by a potential non-autonomous element. The second deletion event of 1,100 bp was predicted at contig 142. The deleted sequence had no similarity to any previously detected TEs of *T. lutea*. However, we detected TIRs belonging to a potential 531-bp TE sequence [39] (S4 and S5 Data). From the TIR sequences, we identified several copies of a new TE family in the reference genome of *T. lutea* (S6 Data), which we named *Shanks* (see 2.4) and produced an annotation of TE copies belonging to this family (S6 Data). Using this improved TE library, we found that 10 TisoArg-specific deletions and 9 TisoS2M2-specific deletions corresponded to TEs described in *T. lutea* (S4 Data). We next investigated specific TE insertion events in TisoArg and TisoS2M2. Considering the new TE family *Shanks*, we detected 26 TisoArg-specific and 55 TisoS2M2-specific TE insertions (Fig 2B; S4 Data). We found that 13 and 11 predictions of specific TE insertions in TisoArg and TisoS2M2 were overlapping. These predictions were considered to be shared (Fig 2B; S4 Data).

2.3 Long read sequencing analysis confirmed specific TE indels predicted by Illumina sequencing

Because numerous specific TE indels were predicted in TisoArg and TisoS2M2, we also sequenced in 2018 the genomes of TisoArg and TisoS2M2 with the Oxford Nanopore Technology (ONT). This technology produces long read sequencing, making possible to span the entire length of TEs and flanking genomic regions [40]. Our objectives were to validate TE indel events predicted by the Illumina sequencing strategy, next to confirm the TE families predicted to be associated and to discover the structure of the involved sequences (autonomous or non-autonomous TEs). Read depths of 4.1× and 7.9× were obtained for TisoArg (37,516 reads) and TisoS2M2 (91,449 reads), respectively. The tool Sniffles predicted 166 and 240 deletions in TisoArg and TisoS2M2. Among them, 90 deletions were shared between TisoArg and TisoS2M2, and none contradicted the predictions obtained with Illumina reads (S4 Data). Sniffles did not retrieve any of the specific TE deletions predicted with Illumina reads in TisoArg but confirmed the existence of 2 out of the 9 TisoS2M2-specific TE deletions detected by Illumina reads (Fig 3A; S4 Data). Sniffles predictions supported the TE deletions caused by a Tc1-Mariner/*Luffy* at the contig 97 and a hAT/*Ace* at the contig 93 (S2A Fig; S4 Data). Focusing on the second one, we retrieved 5 reads with the TE sequence, 1 read without the TE sequence and 2 reads without the TE sequence but bearing a footprint of a target site duplication (TSD) (S2 Fig; S7 Data).

Concerning insertion events, Sniffles detected 311 and 439 insertions in TisoArg and TisoS2M2, respectively (S4 Data). Among those, 19 were shared between strains, and none contradicted Mobster predictions (S4 Data). Sniffles confirmed 16 out of the 27 TisoArg-specific TE insertions and 36 out of the 55 TisoS2M2-specific TE insertions predicted by Mobster with Illumina reads (Fig 3; S4 Data). Long reads of TisoArg confirmed the presence of 12 hAT/*Ace*, 3 Tc1-Mariner/*Luffy*, and 1 Harbinger7 (Fig 3A; S4 Data). The lengths of the insertions predicted for these elements were superior to 2,500 bp, indicating the involvement of autonomous TEs (S3 Fig and S4 Data). For TisoS2M2, long reads confirmed the presence of 31 hAT/*Ace*, two hAT/*Shanks*, two Tc1-Mariner/*Luffy*, and one Harbinger7 (Fig 3B; S4 Data). The lengths of the insertions suggested the involvement of autonomous TEs for the hAT/*Ace*, Tc1-Mariner/*Luffy* and Harbinger7. Among the hAT/*Ace* insertion events, one was two times longer than a regular autonomous *Ace* and corresponded to a chimeric TE sequence composed of two distinct autonomous *Ace* elements (S3 Fig). On the other hand, insertions associated with the hAT/*Shanks* family were two times shorter than the described length of autonomous elements, suggesting the involvement of non-autonomous elements (See 2.4, S3 Fig; S4 Data).

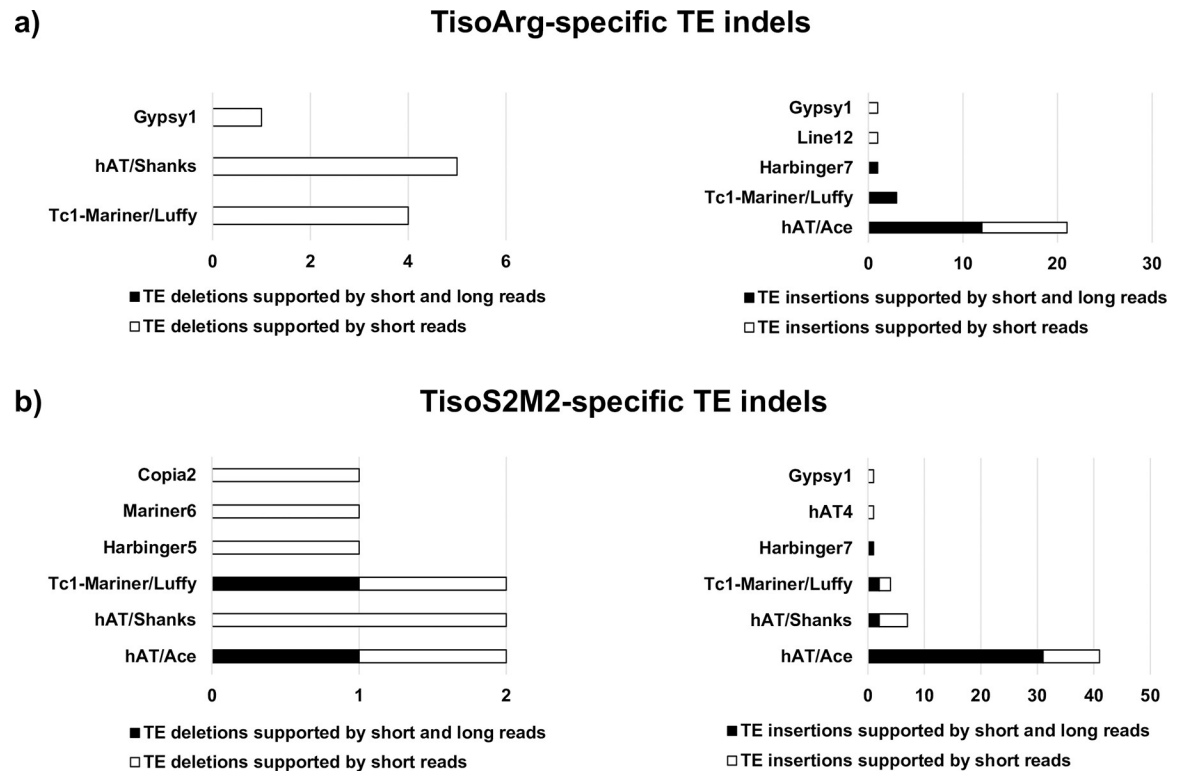


Fig 3. a) Proportion of TE families involved in a) specific TE indels of TisoArg supported by Illumina short reads and ONT long reads b) specific TE indels of TisoS2M2 supported by Illumina short reads and ONT long reads.

<https://doi.org/10.1371/journal.pone.0284656.g003>

2.4 Characterization of TE families to improve the TE annotation of *T. lutea*

Most of the specific TE indels identified in TisoS2M2 were related to three DNA transposon families: *Ace*, *Shanks*, and *Luffy*. Because of the high involvement of these TE families, we characterized their autonomous elements and annotated potential non-autonomous TEs to improve the current TE annotation of *T. lutea* [39].

The *Luffy* family had been previously identified in the reference genome of *T. lutea* with 20 potential autonomous copies [39]. Here, we noted that the autonomous elements of this family have a DNA sequence length of approximately 2,567 bp, terminal sequences of 14 bp, and are bordered by “TA” TSD motifs, which are characteristic of the Tc1-Mariner superfamily [34]. This family has a single open reading frame (ORF) encoding for a peptide of 463 amino acids (aa; S4 Fig). No domain predictions were obtained for the ORF with Pfam, but HHpred detected a transposase domain belonging to the Tc1-Mariner superfamily. We searched for potential non-autonomous elements of this TE family (bearing intact TIRs) using an autonomous referent sequence, but none were identified.

The *Ace* family is composed of three autonomous elements with a length of approximately 2,590 bp, a terminal sequence of 12 bp, and are bordered by TSD motifs of 8 bp (S4 Fig; S6 Data). They have a single ORF encoding for a peptide of 713 aa. Pfam detected two protein families (PF04937 and PF05699) related to the TIR/*hAT* superfamily. We generated an accurate annotation of potential non-autonomous TEs for this TE family and found 92 potential non-autonomous TEs, with lengths ranging from 244 to 1,632 bp. Among them, 55 elements

displayed potential intact TIRs with TSD motifs, and 37 had different TIRs, TSD motifs, or both (S5 Data).

Autonomous *Shanks* elements have a DNA sequence length of approximate 6,768 bp, non-identical TIRs of 26 bp, and are bordered by TSD motifs of 7 bp (S6 Data). They bear two ORFs encoding for enzymes of approximate 793 aa (ORF1) and 580 aa (ORF2; S4 Fig). However, no predictions were obtained for either ORF with Pfam or HHpred. During the course of this study, the TE databank Repbase classified this TE in a small subgroup of transposons TIR/hATv, composed of 16 elements, belonging to three members of the Haptophyta (*T. lutea*, *Chrysochromulina* sp. and *Emiliania huxleyi*). We annotated 1,429 copies in the reference *T. lutea* genome for the *Shanks* family. Among them, we identified 30 potential autonomous TE copies and detected 80 potential non-autonomous TEs belonging to this family. The 80 non-autonomous elements have lengths ranging from 486 to 3,050 bp, and 66 have intact TSD motifs (S6 Data). The new annotations obtained for *Ace* (92 non-autonomous TE) and *Shanks* (1,429 copies) were added to the previous TE annotation of *T. lutea* [39], constituting the second version of the TE annotation of *T. lutea* (see Material and Methods).

2.5 Impact of SNPs and TE indels on genes

Assuming that the improvement program induced specific genetic variants between TisoArg and TisoS2M2, we searched for SNPs or TE indels that could have impacted genes and be causal in the improved phenotype of the domesticated strain. For TisoArg, 804 of the specific SNPs were intragenic, and 385 were predicted to impact the coding sequence by promoting missense variant or change in stop codon (S2 Data). Concerning TisoS2M2, 745 specific SNPs were intragenic, and 330 were predicted to impact the coding sequence (S2 Data). We next focused on the impacted genes that are associated with lipid or catabolic metabolism using Gene Ontology (GO) terms. We retrieved 11 and 9 candidate genes in TisoArg and TisoS2M2, respectively (S2 Data). We searched for homologous genes in the literature, but none have been found to regulate the TFA content of any organisms (S2 Data).

Next, we investigated the potential impact of TE indels located within or close to genes. For instance, de novo TE insertions can drive change in the host phenotype by affecting genes [35, 36, 41]. In TisoArg, 37 TE indels were located within or close ($\pm 1,000$ bp) to 57 genes. For TisoS2M2, 35 TE indels were predicted to impact 41 genes (S8 Data). Using GO terms, we found 3 candidate genes in TisoArg that were associated with catabolic or lipid metabolisms. However, none had a predicted function that could be related to the regulation of the TFA content. In TisoS2M2, we found nine candidate genes potentially impacted by specific TE indels. Interestingly, two gene candidates had homologous genes known to be able to regulate the TFA content. First, an insertion of a hAT/*Shanks* is located downstream (-511 bp) to the gene Tiso_v2_10131 predicted to encode a Cyclin-dependent kinase (S5 Fig; Table 1; S8 Data). Second, a hAT/*Shanks* deletion event is located inside the gene Tiso_v2_15006 predicted to encode for a Long-chain-enoil-CoA hydratase and/or downstream (-50 bp) to the gene Tiso_v2_15005 predicted to encode for a class 3 lipase (S5 Fig; Table 1; S8 Data). The gene predicted to encode for the long-chain-enoil-CoA hydratase has a length of 37,896 bp and seems to correspond to a false positive gene prediction. This gene prediction was not discussed further. Long read alignments made possible to visualize the deletion and insertion events in the TisoS2M2 dataset (S5 Fig).

3. Discussion

This study evaluated the phenotypic stability and genetic changes in the lipid-enhanced domesticated strain TisoS2M2 of *T. lutea*. This domesticated strain was obtained during a

Table 1. Specific TE indels of TisoS2M2 located close to or within genes involved in lipid or catabolic metabolism. The frequencies of the insertion and deletion events were measured by Mobster and Pindel, respectively.

Indel ID	Contig	Start	End	TE family	Gene ID	Predicted enzyme	GO term	TE location	Frequency (%)
Pindel-TisoS2M2-19	142	46892	48002	hAT/Shanks	Tiso_v2_15006 (potential false positive)	Long-chain-enoyl-CoA hydratase	lipid	Inside gene	17
Pindel-TisoS2M2-19	142	46892	48002	hAT/Shanks	Tiso_v2_15005	Lipase Class 3	Lipid and catabolic	Downstream gene (+50 bp)	17
Mobster-TisoS2M2-47	016	260041	260041	hAT/Shanks	Tiso_v2_13322	Arylsulfatase	Lipid and catabolic	Inside gene	23
Mobster-TisoS2M2-272	086	172943	172943	hAT/Shanks	Tiso_v2_10131	Cyclin-dependent kinase A-1	Lipid and catabolic	Upstream gene (-511 bp)	28
Mobster-TisoS2M2-747	189	830328	830328	hAT/Ace	Tiso_v2_06146	60S ribosomal L10-2	catabolic	Upstream gene (-324 pb)	38
Mobster-TisoS2M2-476	128	377408	377408	hAT/Ace	Tiso_v2_11435	Mitogen-activated kinase 5 Short	Lipid and catabolic	Upstream gene (-709)	38
Mobster-TisoS2M2-138 & 139	043	312781	312781	hAT/Shanks	Tiso_v2_09258	RAC family serine threonine- kinase homolog	Lipid and catabolic	Downstream gene (+392)	10
Mobster-TisoS2M2-138	043	312776	312776	hAT/Shanks	Tiso_v2_09258	RAC family serine threonine- kinase homolog	Lipid and catabolic	Downstream gene (+397)	30
Mobster-TisoS2M2-699	179	341961	341961	Tc1-Mariner/ <i>Luffy</i>	Tiso_v2_09731	Cystathionine gamma-synthase	Catabolic	Upstream gene (-27)	37
Mobster-TisoS2M2-351	097	1811743	1811743	hAT/Ace	Tiso_v2_02964	Unknown	Catabolic	Downstream gene (+853 pb)	27

<https://doi.org/10.1371/journal.pone.0284656.t001>

previous study from a native Wt strain TisoArg through a UV-based mutation-selection program [10]. UV radiation is a robust and inexpensive method to generate random mutations and obtain a pool of mutants without previous knowledge of the genetics and metabolisms of the studied microalgae [18]. However, random mutations can be deleterious for cell fitness, and the obtained domesticated strain with a phenotype of interest could occasionally revert to wildtype over time [18, 42]. While the stability of improved phenotypic traits is essential in an industrial context, studies investigating the phenotypic stability of domesticated strains are lacking [18]. Here, we confirmed that seven years after a UV-based improvement program, the domesticated strain TisoS2M2 had a stable, lipid-enhanced trait over time. By comparing the TFA content of the domesticated and native strains in 2019, we revealed that TisoS2M2 had 1.63 times higher TFA content than the native. This ratio is very similar to the ratio previously measured in 2012 (ratio of 1.56) [10]. In addition, a similar proportion of PUFA, SFA and MUFA was found among the TFA content of TisoArg and TisoS2M2 (S1 Data and S1A Fig). The increase in TFA content in TisoS2M2 is therefore caused by a global increase of the fatty acid groups, which is stable over time. The growth rates of TisoS2M2 and TisoArg were very similar and not statistically different in 2019, as also reported in 2012 [10]. Our results confirmed that the improvement program of Bougaran et al. is suitable for the creation of domesticated strains with a stable and improved TFA trait over time [10]. For over two decades, forward genetics strategies comprising random mutagenesis have been used with microalgal strain. Improvement of microalgae was conducted in response to the growing demand for biomolecules of interest such as PUFAs and antioxidant molecules, biomass productivity [20, 43–46], photosynthetic efficiency [47], pesticide tolerance [48], global demand for sustainable biofuels [10, 47, 49–62]. It is, therefore, crucial to control that the improved phenotypic traits of domesticated strains are stable over time [63].

The genetic characterization of domesticated strains is also important to identify the genetic mutation(s) that caused the improved phenotype and to control its persistence over time. Few studies investigated genetic differences between native and domesticated strains, and some identified the causal genetic mutation that gave rise to the improved phenotype [27, 28, 30–32]. A previous study used a reference genome assembly and removed shared background variants between the wildtype strain and mutants to identify the causal mutations [27]. We used a similar method to characterize the genetic changes (SNPs and TE indels) between TisoArg and TisoS2M2. While a part of genetic variants could have spontaneously appeared during the maintenance period, we assumed that the UV-based improvement program caused most of the specific SNPs and TE indels in TisoS2M2.

UV is known to generate random mutations such as SNPs in algal genome [27, 64, 65], and this type of mutation can be causal in the improved phenotype of a domesticated microalgae [27]. We identified thousands of specific SNPs in TisoArg (18,451 SNPs) and TisoS2M2 (13,629 SNPs), and hundreds were predicted to have a potential impact on candidate genes associated with “lipid” or “catabolic” GO terms. However, we found no homologous genes that have been previously described to regulate the TFA content in the literature. This analysis suggested that the improvement program induced changes at the single nucleotide level in TisoS2M2, leading to specific SNPs predictions in both strains. According to our analysis, SNPs did not seem to be the causal type of mutation associated with the lipid-enhanced phenotype of TisoS2M2. However, the fact that predictions of SNPs and their impacts on genes are challenging, and that the knowledge about gene functions of *T. lutea* is limited, may have prevented the detection of causal mutation(s).

TEs are known to be major actors in the genome evolution of organisms and can give rise to new phenotypes by impacting gene expression or function [35, 36, 41]. Because potential autonomous TEs were characterized in *T. lutea* [39], we analyzed the dynamics of TEs in the genome of TisoS2M2. Previous works combining Illumina and ONT sequencing technologies were performed to identify indels [31, 66].

Using this strategy, we found numerous specific TE indels in TisoArg and TisoS2M2. About half of the predictions obtained by the tools using Illumina sequencing were supported by ONT long reads (Fig 3). Most of the specific TE indels were associated with three DNA transposons families indicating their ability to transpose in the genome of *T. lutea*: the TE families hAT/*Ace* and Tc1-Mariner/*Luffy*, which had previously been reported as expressed in the reference genome of *T. lutea* [11], as well as the hAT/*Shanks* family, newly identified in the present study. Most of the TE indels spanned by ONT long reads were found to be caused by autonomous TEs. However, we also revealed that several TE indels, associated with hAT/*Shanks* families, were caused by potential non-autonomous TEs. We annotated many potential non-autonomous elements of the hAT/*Ace* and *Shanks* families in the reference genome of *T. lutea*, suggesting that the *T. lutea* genome hosts numerous non-autonomous TEs belonging to *Ace* and *Shanks* families that could be mobilized. Long read alignments made also possible to pinpoint a locus with heterozygous TE indel events in TisoS2M2 (S2 Fig and S7 Data). The life cycle of *T. lutea* is still unknown, but is potentially haploid-diploid as described for other haptophytes [67]. The life cycle of *T. lutea* is uncontrolled at laboratory conditions and could lead to the presence of haploids and diploid cells in the algal population.

The finding of specific TEs in both the native TisoArg and the domesticated TisoS2M2 was unexpected. We assumed that the UV-based steps of the improvement program previously conducted in 2012 could have promoted TE mobility leading to specific TE indels in TisoS2M2 [10] (Fig 1A). Conversely, the selection effects of the improvement program could have discarded haplotypes of the native TisoArg with specific TE loci, which were not selected in the domesticated TisoS2M2 [10, 68] (Fig 1A). Another hypothesis would be that specific TE indels in

TisoArg and TisoS2M2, or a part, could have appeared during the years of maintenance. TE mobility events were recently reported in a strain of *C. reinhardtii* maintained at laboratory condition [69]. However, almost two times more specific TE indels were found in TisoS2M2 compared to TisoArg, reinforcing the hypothesis that the improvement program had an impact on the mobility of TEs in TisoS2M2.

In microalgal field, previous studies found that TE indels can affect genes, causing phenotypic changes [70–73]. For instance, a previous study identified that an insertion of a TE upstream a desaturase gene caused decrease in the fatty acid contents of a *C. reinhardtii* mutant [74]. Assuming that specific TE indels in TisoS2M2 appeared during the improvement program in 2012 [10], we searched for TE indels that could have impacted genes known to regulate the TFA content of organisms. While no candidates were found for TisoArg, two candidate genes were identified in TisoS2M2. First, a TE insertion was found upstream of a gene encoding for a Cyclin-dependent kinase (CDK). This gene family was mainly reported to be involved in cellular processes, but a study also revealed that a CDK regulated the lipid content of *Arabidopsis thaliana* [75]. Indeed, a knockout mutant of the gene CDK8 in *A. thaliana* has a significant reduction in fatty acid content. Conversely, a higher fatty acid content was measured in a mutant of *A. thaliana* overexpressing this gene [75]. In *T. lutea*, the TE insertion located upstream of a predicted CDK might affect its expression and could lead to an overaccumulation of the TFA content. In addition, a TE deletion was identified downstream a gene encoding for a class 3 lipase. This lipase class is known to be directly involved in the TFA degradation of plants [76], and homologous genes have also been described in microalgae [76–78]. The downregulation of class 3 lipases was notably reported to increase the TFA content of microalgal cells [77]. Nevertheless, these two TE indels are predicted with low allele frequencies (28 and 17%), and are probably not the main genetic(s) cause(s) of the lipid-enhanced phenotype of TisoS2M2. The selection of clonal strains having these TE indels would be an interesting proposition for further investigations. The causal mutation(s) of the TisoS2M2 phenotype remain to be explored.

Although this cannot be confirmed with our current data, the detection of specific TE indels in TisoS2M2 suggested that the UV induced the activity and mobility of several TE families during the improvement program. UV has been reported to trigger TE expression in crops [79, 80]. Indeed, UV caused chromatin remodeling in maize, and induced the expression of epigenetically silenced TEs [81]. A similar scenario could have happened during the improvement program of *T. lutea*. Epigenetic mechanisms are thought to play an essential role in the adaptation processes of microalgae because of fluctuating environmental conditions [82]. While epigenetics mechanisms involving DNA methylation, histone modification, or RNAi have been reported in Chlorophyte, Diatoms, and Dinoflagellate [82–86], they are still unknown for *T. lutea* and, more generally, for haptophytes. Nevertheless, knowledge about the epigenetic regulation of TE activity is still limited and appear to differs depending on the microalgal taxa [83, 87–90]. To our knowledge, no previous studies investigated the effect of UV on the activity of TEs in microalgae. However, the expression of TEs was pinpointed to be seasonally responsive to UV radiations in natural conditions for the Rhodophyta *Cyanidioschyzon* sp. [91]. An effect of UV on the TE activity of *T. lutea* during the improvement program is conceivable and would be a key question to investigate further.

4. Conclusion

We investigated the stability of improved lipid traits and specific genetic changes of the domesticated strain TisoS2M2, previously obtained by a mutation-selection program. After 7 years of maintenance, TisoS2M2 still conserves a higher TFA content compared to the native TisoArg,

revealing a stable, improved lipid phenotype over time. In an industrial context, evaluations of this kind are crucial to ensure the development and use of suitable domesticated strains. In addition, we characterized the genetic changes between the native and the domesticated strain by focusing on the dynamics of TEs. We predicted and confirmed numerous specific TE indels in TisoS2M2, revealing transposition events for several TE families of *T. lutea*. While candidate genes could be impacted by TE indels, the causal mutation(s) explaining the improved phenotype of TisoS2M2 is still undetermined. Our results suggest that UV may have induced the TE activity during the improvement program, which would be an important question to address in the future.

5. Material and methods

5.1 Microalgae strains and culture conditions

The wildtype *T. lutea* strain, TisoArg, was provided by the French Research Institute for Exploitation of the Sea (France) in 1998. TisoS2M2, which accumulated more TFA content compared with TisoArg, was previously obtained from an improvement mutation-selection program [10]. TisoArg and TisoS2M2 are not clonal but composed of populations of cells. For culture maintenance, these strains were grown in 200 mL flasks with natural seawater enriched with sterile Conway medium [92]. These cultures were maintained for 3 weeks, during which time they were supplied with bubbles from 0.22-mm filtered air and kept at a constant temperature of 21 °C and under a continuous light irradiance of 50 $\mu\text{mol m}^{-2} \text{s}^{-1}$.

5.2 Physiological characterization of the studied strains

Comparisons between the native and domesticated strains were conducted in duplicated cultures. Cultures were grown under similar conditions than ones used in the experiment of 2012 [10]. First, 300,000 cells from the mother culture were used to inoculate new culture flasks that were designed for phenotype characterization [93]. Each flask contained 300 mL natural seawater enriched with sterile Conway medium [92] with modified nitrogen content (150 μM instead of 1,500 μM). Each flask was placed in a culture control system with the following parameters: regulated temperature at 23 °C; continuous light irradiance of 350 $\mu\text{E m}^{-2} \text{s}^{-1}$; regulated pH of 7.5 with a CO_2 pulse. The growth of microalgae was automatically monitored by optical density measurements. The formula used to calculate the growth rate was $\mu = (\text{Ln}(X_2) - \text{Ln}(X_1)) / (t_2 - t_1)$. The uptake of NO_3 was measured twice a day as follows: (1) a 2-mL sample was collected; (2) microalgae were eliminated by filtration (0.2 μm); (3) the medium culture was analyzed by spectroscopy at 220 nm to establish the level of remaining NO_3 [94]. After exhaustion of the nitrogen, microalgae growth was stopped after approximately 2 days. After 1 day in stationary phase (corresponding to 6 culture days), harvesting was performed to characterize the microalgae. Each sample was filtered using two pre-combusted GF/C filters for CN and for lipid respectively (diameter, 47 mm; Whatman, Maidstone, UK). To measure the carbon content in each microalgal cell, we used a carbon elemental analyzer to test the filters, which had been placed in an oven and dried at 75 °C for 24 h. The filters were deposited in bottles filled with 6 mL of Folch reagent and deep frozen (-80°C) for lipid measurement. Lipid class separation was performed by column chromatography according to the method by Soudant et al. [95], with a BPX-70 capillary column (60 m long, 0.25 mm internal diameter, 0.25 μm film thickness; SGE, Austin, TX) containing a polar stationary phase (cyanopropyl-siloxane). The upper organic phases containing fatty acid methyl esters (FAMES) were collected and assayed by gas chromatography coupled with flame-ionization detection (GC-FID). FAME quantification was compared with the C17 internal standard (Sigma-Aldrich, St. Louis, MO) by GC-FID using a gas chromatograph

(Autosystem Gas Chromatograph; Perkin-Elmer, Waltham, MA). TFA content was calculated as the sum of SFA, PUFA and MUFA.

5.3 Whole genome sequencing of the *T. lutea* strains

Microalgal cultures were treated with antibiotics (A5955; Sigma-Aldrich) to avoid bacterial contamination. DNA was extracted using a phenol-chloroform method detailed in Hu et al. [96]. Sequencing was conducted with an HiSeq4000 sequencer (150 bp, paired-end; Illumina, San Diego, CA). The raw sequencing data were trimmed with TrimGalore (v0.6.2) (<https://github.com/FelixKrueger/TrimGalore>) to remove adapters and get a quality score higher than Q30. In addition, TisoArg and TisoS2M2 were sequenced with a MinION device (Oxford Nanopore Technologies, Oxford, UK) using the Ligation Sequencing Kit SQK-LSK109 and one flow cell R9.4 for each strain. Basecalling was performed for reads with a minimum quality score of 8 using Guppy v3.3.

5.4 Prediction of polymorphisms

Illumina reads were obtained independently from each strain (TisoArg and TisoS2M2). To compare the polymorphisms of each strain, reads were mapped independently on the reference genome assembly v2 of *T. lutea* (strain CCAP 927/14 [39]) using Mosaik [97] or bwa-mem [98]. SNPs were predicted with DiscoSnp++ [99, 100]. Because the reference genome assembly belongs to a different strain than the studied ones, only the predicted 0/1 and 1/1 SNPs were considered. Predicted SNPs of TisoArg and TisoS2M2 were compared with BCFtools isec [101].

5.5 Prediction of TE indels

Deletions were predicted with the tools Breakdancer [102] and Pindel [103], which were applied to the bwa-mem alignments of each strain. Only deletions with a minimum of 150 bp and a maximum of 15,000 bp were selected as positive candidates (see details in S3 Data). The tool PASTEC [104] was used to detect the deletions caused by TEs and predict their TE family. TE insertions were detected with the tool Mobster [105] using the Mosaik alignments obtained for each of the two strains. Mobster was used with a library of manually curated TE sequences of *T. lutea* (also containing sequences of *Shanks*). To prevent false positive events, we compared the location of the TE insertions detected by Mobster against the TE annotation v2 of *T. lutea* (see 5.6). We removed TE insertions detected by Mobster that were overlapping annotated TEs. Next, we compared the locations of TE indels detected in each strain with bedtools' intersect function [106] and selected only those specific to the domesticated strain (see details in S3 Data). Finally, The tool Sniffles was used with long reads to support TE indels predicted with Illumina sequencing dataset [107]. Sniffles predictions were not used to detect new specific TE indels because of the low read depth of the ONT long reads obtained for TisoArg or TisoS2M2. The tool PASTEC [104] was used to detect the indels caused by TEs and predict their TE family. Read alignments were visualized using the genome browser IGV [108].

5.6 Characterization of the TE families and new TE annotation

To annotate *Shanks* elements and non-autonomous elements of *Ace*, curated reference TE sequences were submitted to TEannot [109] implemented in PiRATE [39]. The new annotated copies were added to the previous TE annotation of *T. lutea* [39], constituting TE annotation v2. We used ORFfinder (www.ncbi.nlm.nih.gov/orffinder/), Pfam [110], HHpred [111], and

the online tool Censor [112] to characterize the autonomous TEs in the *Luffy*, *Ace*, and *Shanks* families.

5.7 Analysis of the genes potentially impacted by SNPs or TE indels

The gene annotation of the genome assembly v2 of *T. lutea* was used [39] to identify genes that could potentially be impacted by SNPs or TE indels. SnpEFF was used to predict the potential effect of SNPs on gene amino acid and we selected the SNPs predicted to have a “moderate” (missense variant) or “high” (changing stop codon) impact [113]. The genes located in a window $\pm 1,000$ bp from the predicted TE indel positions were retrieved using bedtools’ intersect function [106]. The functional annotation of genes was used to retrieve genes with GO term associated to keywords “lipid” and “catabolic”. Function of gene was also manually checked by finding the largest ORF with ORFfinder, then submitting the protein sequence to NCBI BLASTp and HMMER from The MPI Bioinformatics Toolkit [111].

Supporting information

S1 Fig. Additional phenotypic comparison between the native TisoArg and the domesticated strain TisoS2M2 in 2019. a) Comparison of the proportion of SFA, MUFA and PUFA. b) Comparison of the $\Sigma\Omega 3$ and $\Sigma\Omega 6$ content. c) Growth curves of TisoArg and TisoS2M2. For a) and b), bars represent the means of two biological replicates \pm SD. *, $p < 0.05$ by t-test. N.S. means Not significant. For c) Each dot represents the mean of two biological replicates \pm SD. (TIF)

S2 Fig. Alignment of ONT long reads supporting a TE deletion event in TisoS2M2 with heterozygosity. a) Track of the locus at Contig 93 with a predicted hAT/*Ace* deletion in TisoS2M2. The TE deletion event is visible for 3 reads. b) Illustration of the heterozygosity at the loci of the hAT/*Ace* deletion events at Contig 93 and the supporting nucleic motifs found from the long reads of TisoS2M2. The highlight of the motifs are in [S7 Data](#). (TIF)

S3 Fig. Violin plot of the length distribution of specific TE insertions in TisoArg and TisoS2M2 supported by short reads and long reads. Lengths were retrieved from Sniffles predictions in [S4 Data](#)). TE families with one predicted insertion are not shown. (TIF)

S4 Fig. Sequence structures of the TE families Tc1-Mariner/*Luffy*, hAT/*Ace*, and hAT/*Shanks*. (TIF)

S5 Fig. Alignment of ONT long reads supporting TE indels in TisoS2M2 that may impact two candidate genes associated to lipid GO term. a) A TE deletion is predicted close to a gene encoding for a class 3 lipase in TisoS2M2. b) A TE insertion is predicted to be closely located to a gene encoding for a CDK. The indels are highlighted in violet and the lengths of the events are indicated. (TIF)

S1 Data. Comparison of the lipid contents and growth rates between TisoArg and TisoS2M2 in 2019. (XLSX)

S2 Data. List of shared and specific SNPs between TisoArg and TisoS2M2, and predictions of their impacts on genes.

(XLSX)

S3 Data. Pipeline used to predict specific TE indels in TisoArg and TisoS2M2 using Illumina sequencing data.

(PDF)

S4 Data. List of predictions of shared and specific indels detected by Pindel, Breakdancer, Mobster, and Sniffles in TisoArg and TisoS2M2.

(XLSX)

S5 Data. Predicted sequences by Pindel of the deletion events located at the contigs 93 and 142 of *T. lutea*.

(PDF)

S6 Data. List of annotated TEs of *Shanks* and non-autonomous TEs of *Ace*.

(XLSX)

S7 Data. Detailed alignment of the ONT long reads supporting the prediction of a TE deletion with heterozygosity in TisoS2M2.

(PDF)

S8 Data. List of genes predicted to be impacted by TE indels.

(XLSX)

Acknowledgments

We acknowledge the platform Genotoul GeT-PlaGe for the Illumina sequencing of *T. lutea* strains and OIST English editing service for proofreading of the manuscript.

Author Contributions

Conceptualization: Jérémy Berthelie, Bruno Saint-Jean, Grégory Carrier.

Data curation: Jérémy Berthelie, Bruno Saint-Jean, Gaël Bougaran.

Formal analysis: Jérémy Berthelie, Bruno Saint-Jean, Grégory Carrier.

Funding acquisition: Bruno Saint-Jean, Nathalie Casse, Gaël Bougaran, Grégory Carrier.

Investigation: Grégory Carrier.

Methodology: Jérémy Berthelie, Grégory Carrier.

Software: Grégory Carrier.

Supervision: Bruno Saint-Jean, Nathalie Casse, Grégory Carrier.

Validation: Gaël Bougaran.

Visualization: Jérémy Berthelie.

Writing – original draft: Jérémy Berthelie.

Writing – review & editing: Bruno Saint-Jean, Nathalie Casse, Gaël Bougaran, Grégory Carrier.

References

1. Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, et al. Food security: the challenge of feeding 9 billion people. *Science*. 2010; 327(5967):812–8. <https://doi.org/10.1126/science.1185383> PMID: 20110467
2. Milledge JJ. Commercial application of microalgae other than as biofuels: a brief review. *Rev Environ Sci Biotechnol*. 2011 Mar 1; 10(1):31–41.
3. Cadoret JP, Garnier M, Saint-Jean B. Microalgae, functional genomics and biotechnology. In: *Advances in botanical research*. Elsevier; 2012. p. 285–341.
4. Garnier M, Carrier G, Rogniaux H, Nicolau E, Bougaran G, Saint-Jean B, et al. Comparative proteomics reveals proteins impacted by nitrogen deprivation in wild-type and high lipid-accumulating mutant strains of *Tisochrysis lutea*. *J Proteomics*. 2014 Jun; 105:107–20. <https://doi.org/10.1016/j.jprot.2014.02.022> PMID: 24583506
5. Thiriet-Rupert S, Carrier G, Chénais B, Trottier C, Bougaran G, Cadoret JP, et al. Transcription factors in microalgae: genome-wide prediction and comparative analysis. *BMC Genomics*. 2016; 17:1–16.
6. Nef C, Jung S, Mairet F, Kaas R, Grizeau D, Garnier M. How haptophytes microalgae mitigate vitamin B 12 limitation. *Sci Rep*. 2019; 9(1):1–11.
7. Nef C, Henry C, Nicolau É, Bérard JB, Hervé F, Caruana A, et al. Cobalamin scarcity modifies carbon allocation and impairs DMSP production through methionine metabolism in the haptophyte microalgae *Tisochrysis lutea*. *Front Mar Sci*. 2020;7.
8. Nef C, Dittami S, Kaas R, Briand E, Noël C, Mairet F, et al. Sharing Vitamin B12 between Bacteria and Microalgae Does Not Systematically Occur: Case Study of the Haptophyte *Tisochrysis lutea*. *Microorganisms*. 2022; 10(7). <https://doi.org/10.3390/microorganisms10071337> PMID: 35889056
9. Pajot A, Lavaud J, Carrier G, Garnier M, Saint-Jean B, Rabilloud N, et al. The fucoxanthin chlorophyll a/c-binding protein in *tisochrysis lutea*: influence of nitrogen and light on fucoxanthin and chlorophyll a/c-binding protein gene expression and fucoxanthin synthesis. *Front Plant Sci*. 2022; 13. <https://doi.org/10.3389/fpls.2022.830069> PMID: 35251102
10. Bougaran G, Rouxel C, Dubois N, Kaas R, Grouas S, Lukomska E, et al. Enhancement of neutral lipid productivity in the microalga *Isochrysis affinis Galbana* (T-Iso) by a mutation-selection procedure. *Biotechnol Bioeng*. 2012; 109(11):2737–45. <https://doi.org/10.1002/bit.24560> PMID: 22614222
11. Carrier G, Baroukh C, Rouxel C, Duboscq-Bidol L, Schreiber N, Bougaran G. Draft genomes and phenotypic characterization of *Tisochrysis lutea* strains. Toward the production of domesticated strains with high added value. *Algal Res*. 2018; 29:1–11.
12. Thiriet-Rupert S, Carrier G, Trottier C, Eveillard D, Schoefs B, Bougaran G, et al. Identification of transcription factors involved in the phenotype of a domesticated oleaginous microalgae strain of *Tisochrysis lutea*. *Algal Res*. 2018 Mar; 30:59–72.
13. Almutairi AW. Improvement of chemical composition of *Tisochrysis lutea* grown mixotrophically under nitrogen depletion towards biodiesel production. *Molecules*. 2020; 25(20):4609. <https://doi.org/10.3390/molecules25204609> PMID: 33050388
14. Mohamadnia S, Tavakoli O, Faramarzi MA, Shamsollahi Z. Production of fucoxanthin by the microalga *Tisochrysis lutea*: A review of recent developments. *Aquaculture*. 2020; 516:734637.
15. Bigagli E D'Ambrosio M, Cinci L, Niccolai A Biondi N, Rodolfi L, et al. A comparative in vitro evaluation of the anti-inflammatory effects of a *tisochrysis lutea* extract and fucoxanthin. *Mar Drugs*. 2021; 19(6):334. <https://doi.org/10.3390/md19060334> PMID: 34207952
16. Vigani M, Parisi C, Rodríguez-Cerezo E, Barbosa MJ, Sijtsma L, Ploeg M, et al. Food and feed products from micro-algae: Market opportunities and challenges for the EU. *Trends Food Sci Technol*. 2015 Mar; 42(1):81–92.
17. Jebali A, Sanchez MR, Hanschen ER, Starckenburg SR, Corcoran AA. Trait drift in microalgae and applications for strain improvement. *Biotechnol Adv*. 2022 Nov 1; 60:108034. <https://doi.org/10.1016/j.biotechadv.2022.108034> PMID: 36089253
18. Trovão M, Schüller LM, Machado A, Bombo G, Navalho S, Barros A, et al. Random mutagenesis as a promising tool for microalgal strain improvement towards industrial production. *Mar Drugs*. 2022; 20(7):440. <https://doi.org/10.3390/md20070440> PMID: 35877733
19. Meyer RS, Purugganan MD. Evolution of crop species: genetics of domestication and diversification. *Nat Rev Genet*. 2013 Dec; 14(12):840–52. <https://doi.org/10.1038/nrg3605> PMID: 24240513
20. Xue J, Niu YF, Huang T, Yang WD, Liu JS, Li HY. Genetic improvement of the microalga *Phaeodactylum tricornutum* for boosting neutral lipid accumulation. *Metab Eng*. 2015 Jan; 27:1–9. <https://doi.org/10.1016/j.ymben.2014.10.002> PMID: 25447640

21. Yang J, Pan Y, Bowler C, Zhang L, Hu H. Knockdown of phosphoenolpyruvate carboxykinase increases carbon flux to lipid synthesis in *Phaeodactylum tricornutum*. *Algal Res*. 2016 Apr; 15:50–8.
22. Perrineau MM, Gross J, Zelzion E, Price DC, Levitan O, Boyd J, et al. Using natural selection to explore the adaptive potential of *Chlamydomonas reinhardtii*. *PLoS One*. 2014; 9(3):e92533. <https://doi.org/10.1371/journal.pone.0092533> PMID: 24658261
23. Bonnefond H, Grimaud G, Rumin J, Bougaran G, Talec A, Gachelin M, et al. Continuous selection pressure to improve temperature acclimation of *Tisochrysis lutea*. *PLoS One*. 2017; 12(9):e0183547. <https://doi.org/10.1371/journal.pone.0183547> PMID: 28902878
24. Grima EM, Pérez JS, Camacho FG, Medina AR, Giménez AG, Alonso DL. The production of polyunsaturated fatty acids by microalgae: from strain selection to product purification. *Process Biochem*. 1995; 30(8):711–9.
25. Chen Y, Li D, Lu W, Xing J, Hui B, Han Y. Screening and characterization of astaxanthin-hyperproducing mutants of *Haematococcus pluvialis*. *Biotechnol Lett*. 2003 Apr 1; 25(7):527–9. <https://doi.org/10.1023/a:1022877703008> PMID: 12882139
26. Shaish A, Ben-Amotz A, Avron M. Production and selection of high beta carotene mutants of *dunaliella bardawil* chlorophyta. *J Phycol*. 1991 Oct; 27(5):652–6.
27. Schierenbeck L, Ries D, Rogge K, Grewe S, Weisshaar B, Kruse O. Fast forward genetics to identify mutations causing a high light tolerant phenotype in *Chlamydomonas reinhardtii* by whole-genome-sequencing. *BMC Genomics*. 2015; 16:1–15.
28. Sarayloo E, Tardu M, Unlu YS, Simsek S, Cevahir G, Erkey C, et al. Understanding lipid metabolism in high-lipid-producing *Chlorella vulgaris* mutants at the genome-wide level. *Algal Res*. 2017; 28:244–52.
29. Shin SE, Koh HG, Kang NK, Suh WI, Jeong B ryool, Lee B, et al. Isolation, phenotypic characterization and genome wide analysis of a *Chlamydomonas reinhardtii* strain naturally modified under laboratory conditions: towards enhanced microalgal biomass and lipid production for biofuels. *Biotechnol Biofuels*. 2017 Dec; 10(1):308. <https://doi.org/10.1186/s13068-017-1000-0> PMID: 29296121
30. Takeshita T, Ivanov IN, Oshima K, Ishii K, Kawamoto H, Ota S, et al. Comparison of lipid productivity of *Parachlorella kessleri* heavy-ion beam irradiation mutant PK4 in laboratory and 150-L mass bioreactor, identification and characterization of its genetic variation. *Algal Res*. 2018; 35:416–26.
31. Thomy J, Sanchez F, Gut M, Cruz F, Alioto T, Piganeau G, et al. Combining Nanopore and Illumina sequencing permits detailed analysis of insertion mutations and structural variations produced by PEG-mediated transformation in *Ostreococcus tauri*. *Cells*. 2021; 10(3):664. <https://doi.org/10.3390/cells10030664> PMID: 33802698
32. Barten R, Van Workum DJM, de Bakker E, Risse J, Kleisman M, Navalho S, et al. Genetic mechanisms underlying increased microalgal thermotolerance, maximal growth rate, and yield on light following adaptive laboratory evolution. *BMC Biol*. 2022; 20(1):242.
33. Siguier P, Filée J, Chandler M. Insertion sequences in prokaryotic genomes. *Curr Opin Microbiol*. 2006 Oct; 9(5):526–31. <https://doi.org/10.1016/j.mib.2006.08.005> PMID: 16935554
34. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet*. 2007; 8(12):973–82. <https://doi.org/10.1038/nrg2165> PMID: 17984973
35. Lisch D. How important are transposons for plant evolution? *Nat Rev Genet*. 2013 Jan; 14(1):49–61. <https://doi.org/10.1038/nrg3374> PMID: 23247435
36. Furci L, Berthelie J, Juez O, Miryeganeh M, Saze H. Plant Epigenomics. In: *Handbook of Epigenetics*. Elsevier; 2023. p. 263–86.
37. Carrier G, Le Cunff L, Dereeper A, Legrand D, Sabot F, Bouchez O, et al. Transposable elements are a major cause of somatic polymorphism in *Vitis vinifera* L. *PLoS One*. 2012; 7(3):e32973. <https://doi.org/10.1371/journal.pone.0032973> PMID: 22427919
38. Akakpo R, Carpentier M, le Hsing Y, Panaud O. The impact of transposable elements on the structure, evolution and function of the rice genome. *New Phytol*. 2020 Apr; 226(1):44–9. <https://doi.org/10.1111/nph.16356> PMID: 31797393
39. Berthelie J, Casse N, Daccord N, Jamilloux V, Saint-Jean B, Carrier G. A transposable element annotation pipeline and expression analysis reveal potentially active elements in the microalga *Tisochrysis lutea*. *BMC Genomics*. 2018 May 22; 19(1):378. <https://doi.org/10.1186/s12864-018-4763-1> PMID: 29783941
40. Ewing AD, Smits N, Sanchez-Luque FJ, Faivre J, Brennan PM, Richardson SR, et al. Nanopore sequencing enables comprehensive transposable element epigenomic profiling. *Mol Cell*. 2020; 80(5):915–28. <https://doi.org/10.1016/j.molcel.2020.10.024> PMID: 33186547
41. Kazazian HH. Mobile elements: drivers of genome evolution. *Science*. 2004 Mar 12; 303(5664):1626–32. <https://doi.org/10.1126/science.1089670> PMID: 15016989

42. Manandhar-Shrestha K, Hildebrand M. Development of flow cytometric procedures for the efficient isolation of improved lipid accumulation mutants in a *Chlorella* sp. microalga. *J Appl Phycol*. 2013; 25:1643–51. <https://doi.org/10.1007/s10811-013-0021-8> PMID: 24273383
43. Choi GG, Bae MS, Ahn CY, Oh HM. Enhanced biomass and gamma-linolenic acid production of mutant strain *Arthrospira platensis*. *J Microbiol Biotechnol*. 2008; 18(3):539–44. PMID: 18388474
44. Huesemann MH, Hausmann TS, Bartha R, Aksoy M, Weissman JC, Benemann JR. Biomass productivities in wild type and pigment mutant of *Cyclotella* sp. (Diatom). *Appl Biochem Biotechnol*. 2009; 157:507–26. <https://doi.org/10.1007/s12010-008-8298-9> PMID: 18597048
45. Smalley T, Fields FJ, Berndt AJ, Ostrand JT, Heredia V, Mayfield SP. Improving biomass and lipid yields of *Desmodesmus armatus* and *Chlorella vulgaris* through mutagenesis and high-throughput screening. *Biomass Bioenergy*. 2020; 142:105755.
46. López-Sánchez A, Silva-Gálvez AL, Aguilar-Juárez Ó, Senés-Guerrero C, Orozco-Nunnally DA, Carrillo-Nieves D, et al. Microalgae-based livestock wastewater treatment (MbWT) as a circular bioeconomy approach: Enhancement of biomass productivity, pollutant removal and high-value compound production. *J Environ Manage*. 2022; 308:114612. <https://doi.org/10.1016/j.jenvman.2022.114612> PMID: 35149401
47. Kumar V, Sharma N, Jaiswal KK, Vlaskin MS, Nanda M, Tripathi MK, et al. Microalgae with a truncated light-harvesting antenna to maximize photosynthetic efficiency and biomass productivity: Recent advances and current challenges. *Process Biochem*. 2021; 104:83–91.
48. Corcoran AA, Saunders MA, Hanley AP, Lee PA, Lopez S, Ryan R, et al. Iterative screening of an evolutionary engineered *Desmodesmus* generates robust field strains with pesticide tolerance. *Algal Res*. 2018; 31:443–53.
49. Tillich UM, Lehmann S, Schulze K, Dühring U, Frohme M. The optimal mutagen dosage to induce point-mutations in *Synechocystis* sp. PCC6803 and its application to promote temperature tolerance. *PLoS One*. 2012; 7(11):e49467. <https://doi.org/10.1371/journal.pone.0049467> PMID: 23185339
50. Vigeolas H, Duby F, Kaymak E, Niessen G, Motte P, Franck F, et al. Isolation and partial characterization of mutants with elevated lipid content in *Chlorella sorokiniana* and *Scenedesmus obliquus*. *J Biotechnol*. 2012; 162(1):3–12. <https://doi.org/10.1016/j.jbiotec.2012.03.017> PMID: 22480533
51. Zayadan BK, Purton S, Sadvakasova AK, Userbaeva AA, Bolatkhan K. Isolation, mutagenesis, and optimization of cultivation conditions of microalgal strains for biodiesel production. *Russ J Plant Physiol*. 2014 Jan 1; 61(1):124–30.
52. de Jaeger L, Verbeek RE, Draaisma RB, Martens DE, Springer J, Eggink G, et al. Superior triacylglycerol (TAG) accumulation in starchless mutants of *Scenedesmus obliquus*:(l) mutant generation and characterization. *Biotechnol Biofuels*. 2014; 7:1–11.
53. Liu S, Zhao Y, Liu L, Ao X, Ma L, Wu M, et al. Improving cell growth and lipid accumulation in green microalgae *Chlorella* sp. via UV irradiation. *Appl Biochem Biotechnol*. 2015; 175:3507–18. <https://doi.org/10.1007/s12010-015-1521-6> PMID: 25724975
54. Lim DK, Schuhmann H, Sharma K, Schenk PM. Isolation of high-lipid *Tetraselmis suecica* strains following repeated UV-C mutagenesis, facs, and high-throughput growth selection. *BioEnergy Res*. 2015; 8:750–9.
55. Sharma SK, Nelson DR, Abdrabu R, Khraiweh B, Jijakli K, Arnoux M, et al. An integrative Raman microscopy-based workflow for rapid in situ analysis of microalgal lipid bodies. *Biotechnol Biofuels*. 2015; 8(1):1–14.
56. Sivaramakrishnan R, Incharoensakdi A. Enhancement of lipid production in *Scenedesmus* sp. by UV mutagenesis and hydrogen peroxide treatment. *Bioresour Technol*. 2017; 235:366–70. <https://doi.org/10.1016/j.biortech.2017.03.102> PMID: 28384589
57. Moha-León JD, Pérez-Legaspi IA, Ortega-Clemente LA, Rubio-Franchini I, Ríos-Leal E. Improving the lipid content of *Nannochloropsis oculata* by a mutation-selection program using UV radiation and quizalofop. *J Appl Phycol*. 2019; 31:191–9.
58. Mishra S, Liu YJ, Chen CS, Yao DJ. An easily accessible microfluidic chip for high-throughput microalgae screening for biofuel production. *Energies*. 2021; 14(7):1817.
59. Carino JD, Vital PG. Characterization of isolated UV-C-irradiated mutants of microalga *Chlorella vulgaris* for future biofuel application. *Environ Dev Sustain*. 2022; 1–18. <https://doi.org/10.1007/s10668-021-02091-8> PMID: 35002483
60. Lo E, Arora N, Philippidis GP. Physiological insights into enhanced lipid accumulation and temperature tolerance by *Tetraselmis suecica* ultraviolet mutants. *Sci Total Environ*. 2022; 839:156361. <https://doi.org/10.1016/j.scitotenv.2022.156361> PMID: 35640758
61. Tensingh JAS, Shankar V. Sustainable production of biodiesel using UV mutagenesis as a strategy to enhance the lipid productivity in *R. mucilaginosa*. *Sustainability*. 2022; 14(15):9079.

62. Anthony J, Rangamaran VR, Sivasankarasubbiah KT, Nagaraj N, Gopal D, Ramalingam K. An insight into the influence of random mutagenesis on growth and lipid accumulation in *Chlorella vulgaris*—a transcriptome study. *Biomass Convers Biorefinery*. 2022;1–16.
63. Szyjka SJ, Mandal S, Schoepp NG, Tyler BM, Yohn CB, Poon YS, et al. Evaluation of phenotype stability and ecological risk of a genetically engineered alga in open pond production. *Algal Res*. 2017 Jun 1; 24:378–86.
64. Godfroy O, Peters AF, Coelho SM, Cock JM. Genome-wide comparison of ultraviolet and ethyl methanesulphonate mutagenesis methods for the brown alga *Ectocarpus*. *Mar Genomics*. 2015; 24:109–13. <https://doi.org/10.1016/j.margen.2015.03.007> PMID: 25861732
65. Zhang J, Müller BS, Tyre KN, Hersh HL, Bai F, Hu Y, et al. Competitive growth assay of mutagenized *Chlamydomonas reinhardtii* compatible with the international space station veggie plant growth chamber. *Front Plant Sci*. 2020; 11:631. <https://doi.org/10.3389/fpls.2020.00631> PMID: 32523594
66. Lemay MA, Sibbesen JA, Torkamaneh D, Hamel J, Levesque RC, Belzile F. Combined use of Oxford Nanopore and Illumina sequencing yields insights into soybean structural variation biology. *BMC Biol*. 2022; 20(1):1–24.
67. Houdan A, Billard C, Marie D, Not F, Sáez AG, Young JR, et al. Holococcolithophore-heterococcolithophore (Haptophyta) life cycles: Flow cytometric analysis of relative ploidy levels. *Syst Biodivers*. 2004 Jun; 1(4):453–65.
68. Krasovec M, Sanchez-Brosseau S, Grimsley N, Piganeau G. Spontaneous mutation rate as a source of diversity for improving desirable traits in cultured microalgae. *Algal Res*. 2018 Nov; 35:85–90.
69. Craig RJ, Gallaher SD, Shu S, Salomé PA, Jenkins JW, Blaby-Haas CE, et al. The *Chlamydomonas* Genome Project, version 6: Reference assemblies for mating-type plus and minus strains reveal extensive structural mutation in the laboratory. *Plant Cell*. 2023 Feb 1; 35(2):644–72. <https://doi.org/10.1093/plcell/koac347> PMID: 36562730
70. Miller SM, Kirk DL. *glsA*, a *Volvox* gene required for asymmetric division and germ cell specification, encodes a chaperone-like protein. *Development*. 1999; 126(4):649–58. <https://doi.org/10.1242/dev.126.4.649> PMID: 9895313
71. Kirk MM, Stark K, Miller SM, Muller W, Taillon BE, Gruber H, et al. *regA*, a *Volvox* gene that plays a central role in germ-soma differentiation, encodes a novel regulatory protein. *Development*. 1999; 126(4):639–47. <https://doi.org/10.1242/dev.126.4.639> PMID: 9895312
72. Nishii I, Ogihara S, Kirk DL. A kinesin, *InvA*, plays an essential role in *Volvox* morphogenesis. *Cell*. 2003; 113(6):743–53. [https://doi.org/10.1016/s0092-8674\(03\)00431-8](https://doi.org/10.1016/s0092-8674(03)00431-8) PMID: 12809605
73. Helliwell KE, Collins S, Kazamia E, Purton S, Wheeler GL, Smith AG. Fundamental shift in vitamin B12 eco-physiology of a model alga demonstrated by experimental evolution. *ISME J*. 2015; 9(6):1446–55. <https://doi.org/10.1038/ismej.2014.230> PMID: 25526368
74. Nguyen HM, Cuine S, Beyly-Adriano A, Legeret B, Billon E, Auroy P, et al. The green microalga *Chlamydomonas reinhardtii* has a single ω -3 fatty acid desaturase that localizes to the chloroplast and impacts both plastidic and extraplastidic membrane lipids. *Plant Physiol*. 2013 Oct 1; 163(2):914–28.
75. Zhai Z, Blanford JK, Cai Y, Sun J, Liu H, Shi H, et al. CYCLIN-DEPENDENT KINASE 8 positively regulates oil synthesis by activating WRINKLED1 transcription. *New Phytol*. 2023; <https://doi.org/10.1111/nph.18764> PMID: 36683527
76. Nurniwalis A, Zubaidah R, Akmar ASN, Zulkifli H, Arif MM, Massawe F, et al. Genomic structure and characterization of a lipase class 3 gene and promoter from oil palm. *Biol Plant*. 2015; 59(2):227–36.
77. Barka F, Angstenberger M, Ahrendt T, Lorenzen W, Bode HB, Büchel C. Identification of a triacylglycerol lipase in the diatom *Phaeodactylum tricornutum*. *Biochim Biophys Acta BBA—Mol Cell Biol Lipids*. 2016 Mar; 1861(3):239–48. <https://doi.org/10.1016/j.bbalip.2015.12.023> PMID: 26747649
78. Nomaguchi T, Maeda Y, Liang Y, Yoshino T, Asahi T, Tanaka T. Comprehensive analysis of triacylglycerol lipases in the oleaginous diatom *Fistulifera solaris* JPCC DA0580 with transcriptomics under lipid degradation. *J Biosci Bioeng*. 2018 Aug; 126(2):258–65. <https://doi.org/10.1016/j.jbiosc.2018.03.003> PMID: 29628268
79. Ramallo E, Kalendar R, Schulman AH, Martínez-Izquierdo JA. *Reme1*, a *Copia* retrotransposon in melon, is transcriptionally induced by UV light. *Plant Mol Biol*. 2008 Jan; 66(1–2):137–50. <https://doi.org/10.1007/s11103-007-9258-4> PMID: 18034313
80. Walbot V. Reactivation of Mutator transposable elements of maize by ultraviolet light. *MGG Mol Gen Genet*. 1992 Sep; 234(3):353–60. <https://doi.org/10.1007/BF00538694> PMID: 1328840
81. Qüesta JI, Walbot V, Casati P. Mutator transposon activation after UV-B involves chromatin remodeling. *Epigenetics*. 2010 May 16; 5(4):352–63. <https://doi.org/10.4161/epi.5.4.11751> PMID: 20421734
82. Bacova R, Kolackova M, Klejdus B, Adam V, Huska D. Epigenetic mechanisms leading to genetic flexibility during abiotic stress responses in microalgae: A review. *Algal Res*. 2020; 50:101999.

83. Veluchamy A, Lin X, Maumus F, Rivarola M, Bhavsar J, Creasy T, et al. Insights into the role of DNA methylation in diatoms by genome-wide profiling in *Phaeodactylum tricornutum*. *Nat Commun*. 2013; 4(1):2091. <https://doi.org/10.1038/ncomms3091> PMID: 23820484
84. Tirichine L, Rastogi A, Bowler C. Recent progress in diatom genomics and epigenomics. *Curr Opin Plant Biol*. 2017; 36:46–55. <https://doi.org/10.1016/j.pbi.2017.02.001> PMID: 28226268
85. Yang F, Li L, Lin S. Methylation pattern and expression dynamics of methylase and photosystem genes under varying light intensities in *Fugacium kawagutii* (Symbiodiniaceae). *J Phycol*. 2020; 56(6):1738–47. <https://doi.org/10.1111/jpy.13070> PMID: 32920818
86. Li Y. Algal epigenetics: insights from DNA methylation in a symbiotic dinoflagellate. *J Phycol*. 2022;
87. Wu-Scharf D, Jeong B ryool, Zhang C, Cerutti H. Transgene and transposon silencing in *Chlamydomonas reinhardtii* by a DEAH-box RNA helicase. *Science*. 2000; 290(5494):1159–62. <https://doi.org/10.1126/science.290.5494.1159> PMID: 11073454
88. Casas-Mollano JA, Rohr J, Kim EJ, Balassa E, Van Dijk K, Cerutti H. Diversification of the core RNA interference machinery in *Chlamydomonas reinhardtii* and the role of DCL1 in transposon silencing. *Genetics*. 2008; 179(1):69–81. <https://doi.org/10.1534/genetics.107.086546> PMID: 18493041
89. Khan A, Eikani CK, Khan H, Iavarone AT, Pesavento JJ. Characterization of *Chlamydomonas reinhardtii* core histones by top-down mass spectrometry reveals unique algae-specific variants and post-translational modifications. *J Proteome Res*. 2018; 17(1):23–32. <https://doi.org/10.1021/acs.jproteome.7b00780> PMID: 29198113
90. Vigneau J, Borg M. The epigenetic origin of life history transitions in plants and algae. *Plant Reprod*. 2021; 34(4):267–85. <https://doi.org/10.1007/s00497-021-00422-3> PMID: 34236522
91. Skorupa DJ, Castenholz RW, Mazurie A, Carey C, Rosenzweig F, McDermott TR. In situ gene expression profiling of the thermoacidophilic alga *Cyanidioschyzon* in relation to visible and ultraviolet irradiance: *Cyanidioschyzon* gene expression profiling? *Environ Microbiol*. 2014 Jun; 16(6):1627–41.
92. Walne PR. Culture of bivalve molluscs: 50 years' experience at Conwy. Fishing News Books Ltd.; 1979.
93. Dubois N, Cadoret JP, Bougaran G, Rouxel C, Doulin S. *Isochrysis* sp tahitian clone and uses therefore. 2016 Aug 9;
94. Collos Y, Mornet F, Sciandra A, Waser N, Larson A, Harrison P. An optical method for the rapid measurement of micromolar concentrations of nitrate in marine phytoplankton cultures. *J Appl Phycol*. 1999; 11(2):179–84.
95. Soudant P, Marty Y, Moal J, Samain J. Separation of major polar lipids in *Pecten maximus* by high-performance liquid chromatography and subsequent determination of their fatty acids using gas chromatography. *J Chromatogr B Biomed Sci App*. 1995; 673(1):15–26. [https://doi.org/10.1016/0378-4347\(95\)00241-a](https://doi.org/10.1016/0378-4347(95)00241-a) PMID: 8925070
96. Hu Z, Zeng X, Wang A, Shi C, Duan D. An efficient method for DNA isolation from red algae. *J Appl Phycol*. 2004; 16(3):161–6.
97. Lee WP, Stromberg MP, Ward A, Stewart C, Garrison EP, Marth GT. MOSAIK: a hash-based algorithm for accurate next-generation sequencing short-read mapping. *PLoS One*. 2014; 9(3):e90581. <https://doi.org/10.1371/journal.pone.0090581> PMID: 24599324
98. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv Prepr ArXiv13033997*. 2013;
99. Uricaru R, Rizk G, Lacroix V, Quillery E, Plantard O, Chikhi R, et al. Reference-free detection of isolated SNPs. *Nucleic Acids Res*. 2015; 43(2):e11–e11. <https://doi.org/10.1093/nar/gku1187> PMID: 25404127
100. Peterlongo P, Riou C, Drezen E, Lemaitre C. DiscoSnp++: de novo detection of small variants from raw unassembled read set (s). *BioRxiv*. 2017; 209965.
101. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011; 27(21):2987–93. <https://doi.org/10.1093/bioinformatics/btr509> PMID: 21903627
102. Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods*. 2009 Sep; 6(9):677–81. <https://doi.org/10.1038/nmeth.1363> PMID: 19668202
103. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*. 2009 Nov 1; 25(21):2865–71. <https://doi.org/10.1093/bioinformatics/btp394> PMID: 19561018
104. Hoede C, Arnoux S, Moisset M, Chaumier T, Inizan O, Jamilloux V, et al. PASTEC: an automatic transposable element classification tool. Cordaux R, editor. *PLoS ONE*. 2014 May 2; 9(5):e91929.

105. Thung DT, Slagboom EP, Ye K, Veltman JA, Hehir-Kwa JY. Mobster: accurate detection of mobile element insertions in next generation sequencing data. 2014; 11.
106. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26(6):841–2. <https://doi.org/10.1093/bioinformatics/btq033> PMID: 20110278
107. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods*. 2018 juin; 15(6):461–8. <https://doi.org/10.1038/s41592-018-0001-7> PMID: 29713083
108. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013; 14(2):178–92. <https://doi.org/10.1093/bib/bbs017> PMID: 22517427
109. Flutre T, Duprat E, Feuillet C, Quesneville H. Considering transposable element diversification in de novo annotation approaches. *PLoS One*. 2011; 6(1):e16526. <https://doi.org/10.1371/journal.pone.0016526> PMID: 21304975
110. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016 Jan 4; 44(D1):D279–85. <https://doi.org/10.1093/nar/gkv1344> PMID: 26673716
111. Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol*. 2018; 430(15):2237–43. <https://doi.org/10.1016/j.jmb.2017.12.007> PMID: 29258817
112. Kohany O, Gentles AJ, Hankus L, Jurka J. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics*. 2006; 7. <https://doi.org/10.1186/1471-2105-7-474> PMID: 17064419
113. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012; 6(2):80–92. <https://doi.org/10.4161/fly.19695> PMID: 22728672