Algal Research
July 2023, Volume 74, Pages 103148 (8p.)
https://doi.org/10.1016/j.algal.2023.103148
https://archimer.ifremer.fr/doc/00839/95111/

Towards the optimization of genetic polymorphism with EMS-induced mutagenesis in *Phaeodactylum tricornutum*

Rumin Judith [1, 2], Carrier Gregory [1, 3], Rouxel Catherine [1], Charrier Aurelie [1, 4], Raimbault Virginie [1, 4], Cadoret Jean-Paul [1, 5], Bougaran Gael [1, 4], Saint-Jean Bruno [1, 3, *]

[1] IFREMER, Laboratoire PBA, 44311 Nantes, France
[2] LISAQUA, 21 Rue Marcel Schowb, 44100 Nantes, France
[3] Ifremer, PHYTOX, laboratoire GENALG, F-44000 Nantes, France
[4] Ifremer, PHYTOX, laboratoire PHYSALG, F-44000 Nantes, France
[5] ALGAMA, 81 rue Réaumur, 75002 Paris, France

* Corresponding author : Bruno Saint-Jean, email address : Bruno.saintjean@ifremer.fr

Abstract :

Microalgae remain an exciting target for biotechnology as they offer a largely unexploited reservoir of novel and valuable bioactive compounds. Strain improvement programs are an expanding research field aiming to multiply microalgal potential. This study evaluates the genetic diversity created in populations of the diatom Phaeodactylum tricornutum subjected to random mutagenesis. We explored the genetic diversity using genotyping-by-sequencing (GBS) to estimate and compare the impact of the most common chemical mutagen (ethyl methanesulfonate, EMS). Five microalga populations obtained following EMS treatment had survival rates between 1 and 98 %. High genetic diversity was obtained for only one of these P. tricornutum populations, with a survival rate close to 30 %.

Highlights

► A literature data-based modeling approach was used to assess the best mutagenesis conditions aiming to maximize the genetic diversity. ► Survival rate of EMS mutagenized-population of *P. tricornutum* is an helpfull criterion for mutagenesis optimization ► Mutagenesis experiments on *P. tricornutum* showed that the survival rate (83 to 12%) has a negative linear trend as a function of the EMS dose applied (3.65 to 4.00 M). ► Allelic frequency showed a threshold effect with EMS dose, with the maximum frequency of mutations obtained for 26% survival rate (3.92 M EMS dose).

Keywords : Microalgae, Mutation, Ethyl methanesulfonate, Genetic polymorphism, Survival rate, Genotyping-by-sequencing

1. Introduction

Microalgae, a natural and renewable resource, have been a promising source of bio-based feedstocks and bioactive compounds since the beginning of this century. The huge potential of microalgae in the biotechnology market, including applications in nutrition, cosmetics and health, is constantly growing. These real-cell factories can provide a number of valuable natural chemicals, including pigments and food supplements, such as eicosapentaenoic acid (EPA), docosahexaenoic acid (DHA), and vitamins, as well as other compounds of interest. However, microalgal production costs remain high [1]. Apart from

efforts to design more efficient cultivation processes and photobioreactors, species improvement appears a promising alternative to reduce production costs.

Research work on boosting the accumulation of high added-value molecules is expanding. Initially driven by biodiesel production studies, investigations are now focusing on other bio-molecules, including those with antioxidant, antiviral, antibacterial, antifungal, anti-inflammatory, antitumor, and antimalarial activity [2,3,4]. The main experimental approaches for strain improvement include genetic engineering and synthetic biology, adaptive laboratory evolution (ALE), or *in silico* design of algal strains [5], but the most widespread method is random mutagenesis. Over the last decade, an increasing number of studies have used UV mutagenesis, ethyl methanesulfonate (EMS) mutagenesis or gamma mutagenesis to induce higher genetic polymorphism in microalgal genomes. Among recent advances in strain improvement programs, the most advanced studies have worked on the development of high-throughput procedures to increase screening efficiency of selecting strains from mutagenesis [6]. Yi et al., (2018) efficiently combined a diphenylamine (DPA) inhibition of the carotenogenic pathway, an EMS chemical mutagenesis, selection of strains with fluorescence intensity, and the selection of strains with high total carotenoid content after pigment extraction [6]. The greater genetic diversity induced provides higher phenotypic diversity. As a first step, mutagenesis is generally used to increase the frequency of mutation in microalgal populations, then a second step screens the resulting phenotypes to select the best individuals [7].

The main bottleneck for the whole mutagenesis procedure remains the random step of the mutations themselves, which makes it difficult to establish an ideal protocol. Since lethal and beneficial mutations occur randomly, the effectiveness of a mutagenesis process relies on the balance between the gain in genetic diversity and the rate of mortality within the population [6]. Therefore, efficiency mainly depends on the type and concentration of the mutagen. In the literature, the efficiency of random mutagenesis is arbitrarily estimated by survival rate, usually set at 10% [9–11], 40% [12], 50% [13], or 65% [14], depending on the study. However, no studies dealing with microalgae have yet established a link between the estimated survival rate following mutagenesis and the gain in genetic diversity, meaning this issue has not been approached systematically.

At the same time, the advent of cost-effective next generation sequencing (NGS) platforms has made it possible to produce sequence information and to have access to the

genetic polymorphism of an entire single organism. Simplify the genotyping methods is a particularly appealing approach to analyzing genetic diversity within microalgae populations. Different methods based on restriction enzyme digestion have been developed for this task, including reduced-representation libraries (RRLs), complexity reduction of polymorphic sequences (CRoPS), restriction-site-associated DNA sequencing (RADseq), or genotyping-by-sequencing (GBS) [15]. The GBS approach, developed by Elshire (2011)[16], is a powerful method, increasingly used to study genetic diversity or to develop molecular markers for agronomic plants of interest. Its low cost, rapid application, and flexibility make it a method of choice for genotyping plants in agronomy. Genotyping-by-sequencing relies on the use of restriction enzymes. By choosing the right enzymes, shared regions of the genome can be compared between different samples. Specifically, by significantly reducing the genome size, GBS provides an affordable means to search for genetic polymorphisms (single nucleotide polymorphism (SNP), Insertion (IN), or Deletion (DEL)) among a large number of individuals.

In order to improve gains in genetic diversity by mutagenesis, and thus the efficiency of microalgal domestication programs, we propose a dual approach based on: (1) the experimental optimization of genetic polymorphism among *Phaeodactylum tricornutum* populations by identifying the optimal survival rate after EMS-induced mutagenesis, and (2) a modeling approach to mutagenesis. The marine diatom *P. tricornutum* is a good potential producer of valuable lipids due to its fast growth, lipid accumulation capability, and established genetic tools [14,17,18], making it a suitable candidate species to examine [17,19,20]. This study is part of the first step of a domestication process (mutation/selection) aiming to obtain improved strains. Here, GBS was used to assess the genetic polymorphism following EMS-induced mutations. Although this method has been frequently used for terrestrial plants, this is the first report of its application in the field of microalgal improvement programs.

2. Material and methods

2.1. Algal material

2.1.1. Algal strains and their maintenance

Phaeodactylum tricornutum CCAP 1055/1 was obtained from the culture collection of algae of the CCAP (Culture Center of Algae and Protozoa) and maintained at 20°C under continuous light (50 µmol m^{-2} s^{-1}). Cultures were grown in fresh Erlenmeyer flasks containing seawater filtered at 0.22 µm and sterilized, enriched with a Walne's medium [21] supplemented with Na_2SiO_3*$5H_2O$ 40 g L^{-1}. Xenic cultures were treated with a specific mix of antibiotics based on Cho et al. (2002)[22] and regularly centrifuged twice at 500 g for 10 min at 20°C to eliminate bacteria before mutagenesis.

2.1.2. Cultivation conditions

Cultures were grown in 2-L flasks in batch mode and bubbled with 0.22-µm filtered air. Culture medium flasks enriched with 1 mL L^{-1} Conway (Walne, 1966 [21]) supplemented with Na_2SiO_3*$5H_2O$ 40 g L^{-1} were previously sterilized for 20 min at 120°C before the experiment [23]. Synchronization of *P. tricornutum* cells in 12:12 (light:dark) cycle was applied for cell division. The cultures were maintained at a constant temperature of 21°C, under an irradiance of 160 µmol m^{-2} s^{-1} for 12 h. Initial cell concentration was 4 x 10^5 cell mL^{-1}.

2.1.3. Random mutagenesis by ethyl methanesulfonate

Ten mL samples of synchronized cell culture in exponential growth phase at 2 x 10^6 cells mL^{-1} were exposed to seven EMS concentrations (3.65; 3.775; 3.8; 3.875; 3.9; 3.925; 4 M) for 30 min. These seven concentrations allowed us to obtain survival rates ranging from 10 to 90%. Exposures beyond 30 min resulted in 100% mortality of the populations. Exposures were carried out in darkness at room temperature with agitation by turning. A control population (*c*) without EMS exposure was prepared following the same procedure (Figure 1). Following incubation, the microalgal cells were centrifuged at 600 g for 10 min at 20°C and cell pellets were washed three times with 10 mL seawater that had been filtered at 0.22 µm and autoclaved (hereafter referred to as "sterilized seawater"), containing 10% (w/v) sodium thiosulfate. Cell pellets from each treatment were re-suspended in 2 mL of sterilized seawater enriched with Walne's medium supplemented with Na_2SiO_3*$5H_2O$ 40 g.L^{-1} and distributed in

24-well plates (Thermo Scientific™ Nunc™). After an initial measurement of Chl *a* fluorescence (450/685nm), OD_{680nm} and OD_{800nm} performed on a Tecan spectrofluorometer (by night), cells were kept at 20°C in darkness overnight to prevent photoactivation. Microplate cultures were then maintained under an irradiance of 80 µmol m^{-2} s^{-1} and monitored for cell growth (Chl *a* fluorescence, OD_{680nm}, OD_{800nm}) with the Tecan spectrofluorometer. The survival rate (*Ys*) was calculated as the ratio between the lowest value (*Xmin*) measured by Chl *a* fluorescence, OD_{680nm}, OD_{800nm} during the growth monitoring and the initial value (*Xi*):

$$Ys = \frac{X_{min}}{X_i} 100 \quad \text{(Eq. 1)}$$

After recovery of cultures, the microalgal populations were re-suspended in 100 mL of sterilized seawater enriched with Walne's medium supplemented with Na2SiO3*5H2O 40 g L^{-1}. Five populations exposed to EMS concentrations of 3.65; 3.775; 3.8; 3.925; 4 M were selected for DNA extraction.

Before DNA extraction, 50 mL samples of 1 x 10^{6} cells ml^{-1} cultures were centrifuged. One *P. tricornutum* population was used as a reference population (*R*) before the mutagenesis procedure. This allowed a subsequent comparison with the five populations exposed to the mutagen and with the control population (*c*).

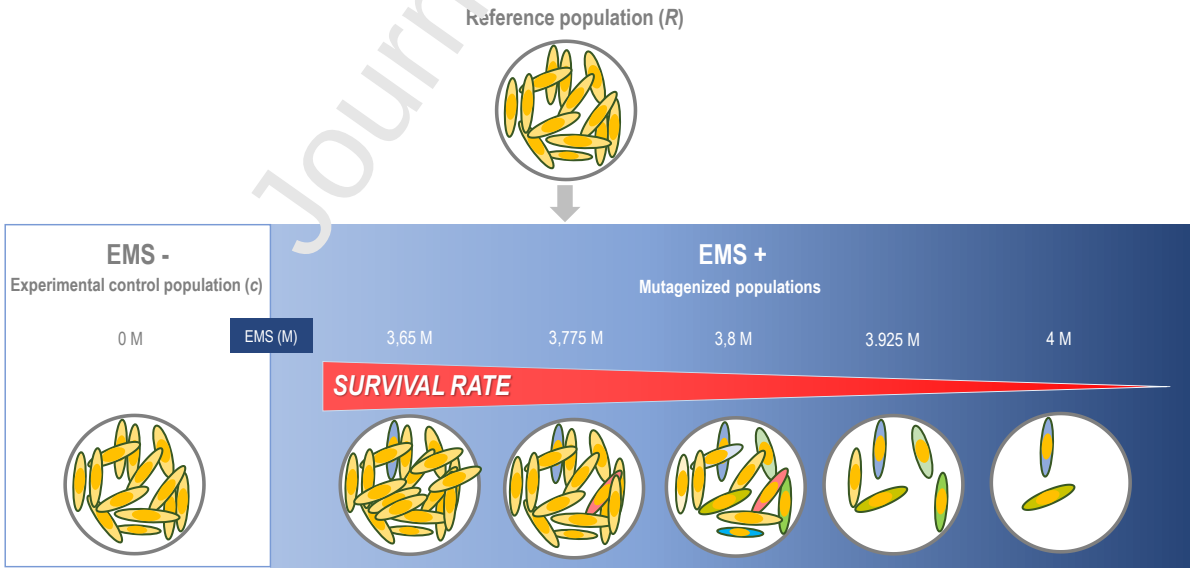

Figure 1. Schematic representation of the EMS mutagenesis steps. Five EMS mutagenesis treatments with five different EMS concentrations were performed on the reference population strain (*R*) of *P. tricornutum* to obtain five microalgal population strains. An experimental control was also included in the experiment (strain *c*).

2.2. *In silico* assessment of restriction enzymes, DNA extraction, and restriction digest

In silico digestion of the *P. tricornutum* reference genome was performed with different restriction enzymes: ApeKI 5‘-GCWG C-3’, where W is A or T; MspI 5‘-C CGG-3’; and PstI 5‘-CTGCA G-3’ (New England Biolabs), with a computing command developed in our laboratory. The frequency distribution of fragment size was evaluated for each enzyme to determine the proportion of fragments in the preferred size range (around 400 bp) and to obtain 10% of the total genome. The frequency distribution of fragment size was evaluated for each enzyme ApeKI, MspI, and PstI. In this study, the digestion protocol was carried out with ApeKI, which is also the most frequently used enzyme for GBS analyses [24].

Then, DNA from the seven populations was extracted from culture pellets and quantified using an Infinite® 200 PRO NanoQuant microplate reader (Tecan). One µg of genomic DNA was digested for 2 h at 72°C in a 5 µL reaction volume of NEB buffer with 5 U of ApeKI enzymes (New England Biolabs). Genomic DNA was then purified with the Invitrogen PureLink^{TM} kit (Quick PCR Purification Kit) following the manufacturer’s protocol before preparation of the GBS library.

2.3. Preparation of libraries

2.3.1. Adapter design and ligation

A set of seven barcode adapters and one common adapter with an ApeKI overhang were designed using barcode design considerations (http://www.deenabio.com). The barcode adapter and the common adapter were hybridized and normalized at 0.75 ng mL^{-1} in TE with Qubit® dsDNA HS (High Sensitivity) Assay Kit quantification. To minimize the possibility of producing adapter dimers and DNA chimeras during ligation, eight adapter concentrations were tested with 200 ng DNA samples of *P. tricornutum* with T4 DNA Ligase (New England Biolabs, M0202S). Samples were then purified with the Invitrogen PureLink^{TM} kit (Quick PCR Purification Kit) before the amplification with the NEBNext High-Fidelity 2X PCR Master Mix kit (New England Biolabs, M0541S) following the manufacturer’s protocol (with a temperature cycling program of 72°C for 5 min; 98°C for 30 sec followed by 15 cycles of 98°C for 10 sec; 65°C for 30 sec; 72°C for 30 sec, and a final step of 72°C for 5 min). After purification with Invitrogen PureLink^{TM} (Quick PCR Purification Kit, K310001), the appropriate quantity of adapter was determined with DNA analyses performed with a

TapeStation DNA 1000 Bioanalyzer 2200 TapeStation Instrument (Agilent Technologies) specific to DNA fragments between 35 and 1000 bp.

2.3.2. Multiplexing and sizing

The seven DNA samples were multiplexed in equimolar quantity and 1 µg of the mix was taken for sizing by microfluidic electrophoresis with a LabChip XT DNA (Caliper Life Technologies) instrument specific for fragments between 50 and 750 bp. Ten nM of selected DNA fragments from 250 to 500 bp were finally quantified on a 2200 TapeStation before NGS sequencing.

2.4. Sequencing

2.4.1. Illumina processing

Paired-end genome sequencing was performed on an Illumina HiSeq 2000 sequencer (Illumina Corporation Inc.). Raw 100 bp sequence data were cleaned with FastQC software developed by S. Andrews (2010) at the Babraham Institute (www.bioinformatics.bbsrc.ac.uk) to validate run quality (read number and quality score of nucleotides sequenced [25], composition of reads). Read pairs obtained for each sample were filtered to select only those of sufficient quality to be collected. Reads containing the Illumina adapter sequences and Illumina control sequences were eliminated with Cutadapt software [26], reads with a quality score below Q30 on a mean quality score of all nucleotides and a length below 75 bp were eliminated. Finally, nucleotides with a quality score below Q30 were also eliminated.

2.4.2. DNA sequence alignments and genetic mapping

Genome datasets of the six experimental populations were aligned using MosaikAssembler [27] and the *P. tricornutum* reference genome. To improve read alignment, the percentage of maximum polymorphism between the sequence and the reference genome was set at 15% and the size separating the two sequences of a pair was set at between 100 and 400 bp. On average 82.3% of read pairs were aligned on a single locus, 15.1% aligned on multiple loci and 2.6% were not aligned. The maximum sequencing depth per base was then estimated depending on relevant sequencing coverage for all seven populations.

Mutation discovery and genotyping were performed by polymorphism identification: SNP (Single Nucleotide Polymorphism), IN (INsertion polymorphism), and DEL (DELetion polymorphism). Freebayes (https://arxiv.org/abs/1207.3907) was used on alignments obtained from each population with the options: freebayes -L Aligmenent-Files.bam -f Reference-Genome-v2.fasta -v Output.vcf --min-mapping-quality 5 --min-base-quality 30 --min-alternate-fraction 0.05 --min-alternate-count 5 --min-coverage 20. A higher frequency of 5% polymorphism was set (min-alternate-fraction option) with five minimal reads (min-alternate-count option) to consider true polymorphism, limit false positives, and consider only polymorphisms sufficiently anchored in the population. The detected polymorphisms were compared with the reference strain control and the origin of the polymorphism, mutation or selection, was also determined (S2). The polymorphism ratio was computed using the number of polymorphisms in a treatment population compared with the number of polymorphisms in the reference population (*R*). The mutation frequency was calculated over 6 generations so as to obtain the required genetic material after culture recovery.

3. Results and discussion

In order to assess the genetic diversity among the mutagenized and control populations, we aimed to examine a representative portion of the genomes, highlighting the common and diverging regions between the populations. On the one hand, while the cost of sequencing is constantly falling [28], it is still too high to fully sequence several individuals at a time. On the other hand, GBS is a powerful method to reduce genome representations by using a combination of restriction enzymes, together with the selection of an efficient DNA fragment size. By reducing the sequencing cost, GBS makes it possible to consider the sequencing of several samples at a time.

However, choosing the appropriate restriction enzyme is a critical step in developing a GBS protocol for a novel organism. *In silico* digestions of the *P. tricornutum* reference genome were performed for different restriction enzymes to obtain around 10% of the genome with a DNA fragment size of around 400 bases to remain under Illumina sequencing protocol conditions. The frequency distribution of fragment size was evaluated for each enzyme ApeKI, MspI, and PstI. *In silico* libraries after the ApeKI and MspI digestions showed a preponderance of smaller fragments, in particular 38 000 and 65 000 fragments in the size range of interest (around 400 bp) for ApeKI and MspI digestion, respectively (**Error!**

Reference source not found.). The *in silico* digestion with PstI generated size fragments that were too large (< 1500 bp). The percentages of sequenced genome including size fragments between 250 and 550 bp were 11.1% and 14.1% for the ApeKI and MspI digestions, respectively. Finally, with the enzyme ApeKI, we covered 5.6% of the *P. tricornutum* genome (1,359,215 bp) at a depth coverage of 50X for all populations. It is worth noting that previous studies aiming at screening for induced mutations analyzed only a limited number of genes (often a dozen or so genes) [29–31]. In our study, the percentage of the genome studied without bias (coding and non-coding regions) appears to be more efficient at estimating the mutagenic effects on the genome.

Table 1. Percentage of the sequenced genome in DNA fragments of different sizes after *in silico* digestions of the *P. tricornutum* reference genome with ApeKI and MspI.

Size of DNA fragments (bp)	Percentage of the sequenced genome (%)	
	ApeKI	MspI
200–600	14.6	18.5
250–550	11.1	14.1
300–500	7.4	9.5
350–500	5.6	7.0

3.1. Modeling the mutagenesis procedure

The challenge is to apply the appropriate dose of mutagen to increase and optimize genetic diversity. To identify the optimal survival rate and increase the mutability, i.e., the capacity of an organism to be mutated [8], the determination of the number of viable mutants in the resulting population is a critical issue. The hypothesis made in this study is based on the results of [32] for EMS and the observation of [33] for X-rays. It defines the number of viable mutants N_m as the product of two functions of the mutagen dose (Eq. 2): the survival rate, expressed as the percentage of survivors Y_s, which is an affine function (Eq. 3); and the number of mutated cells among surviving cells Y_m, which is a linear function (Eq. 4).

$$N_m = N_0 . Y_s . Y_m \quad \text{(Eq. 2)}$$

$$Y_S = -b . D + 1 \quad \text{(Eq. 3)}$$

$$Y_m = a . D \quad \text{(Eq. 4)}$$

where D is the mutagen dose, N_0 the initial cell number subjected to mutagenesis and a and b are constants. From Eq. 3, it can be demonstrated that N_m is a parabolic function of EMS dose with a maximum obtained when Y_s is 50%.

The linear trend for Y_s is in agreement with other data from the literature [9,13,32]. However, some studies did not find an obvious linear relationship between EMS dose and Y_s on microalgae and reported a threshold effect [10], while others used a single survival rate without addressing the correlation between the two parameters [8,11,12]. Another source of uncertainty is that many previous publications used different survival rates [9,11–14], suggesting disagreements in the scientific community. In the following sections, we show how we used this modeling approach to guide our investigations to find the best mutagenesis conditions.

3.2. Obtention of mutagenized populations of *P. tricornutum* with EMS-induced mutagenesis

Mutagenesis experiments were carried out on *P. tricornutum* strains in this study, producing five mutagenized microalga populations with different EMS concentrations (3.650; 3.775; 3.800; 3.925 and 4.000 M) (Figure 2). Survival rate showed a negative linear trend as a function of the EMS dose applied, ranging from 83 to 12% (R^2 = 0.87; Figure 2). This result confirms the first hypothesis of our model.

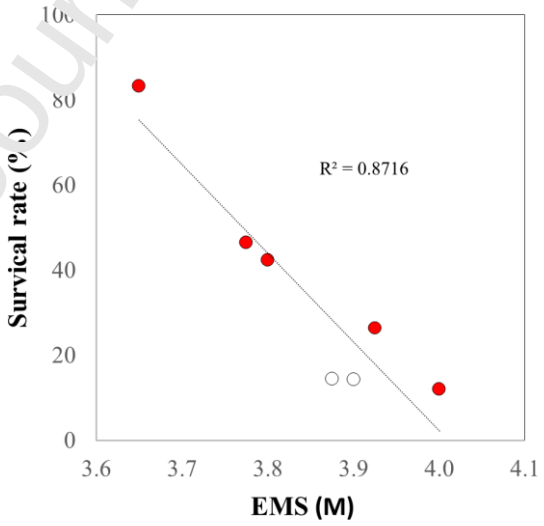

Figure 2. Correlation between survival rate (%) of *P. tricornutum* populations and EMS dose applied (M). The five populations selected for DNA analysis are in red.

3.3. GBS method for comparing the genetic polymorphism of populations

Genomes of each population of *P. tricornutum* were sequenced using Illumina sequencing technology. For each population, reads were independently mapped on the genome of the reference strain (R) after quality filtering. Finally, 70% of the total read pairs were kept for the polymorphism analysis for the six populations, confirming experimental reproducibility of the genome digestion with ApeKI, adapter ligation, and 250–550 bp fragment sizing steps (Table 2).

Table 2. EMS mutagenesis treatments, survival rates after mutagenesis, number of sequenced bases, barcodes and percentage discarded after quality filtering for each *P. tricornutum* strain.

Survival rate	EMS concentration (M)	Duration treatment (min)	Number of sequence bases	Barcodes	Deleted percentage after filtering (%)
83	3.65	30	4,047,897	GGTGT	30
47	3.775	30	5,128,048	AACAA	35
42	3.8	30	6,974,093	CCACT	33
26	3.925	30	3,787,158	TTGTA	33
12	4	30	7,745,369	AAGGT	34
c	0	30	4,347,142	CGTAA	33
R	-	-	7,104,794	CTCGA	34

The polymorphism of the reference strain (*R*) was estimated at 17,566 SNPs and 778 IN/DELs in 6.9% of the genome. For further comparison, the number of polymorphisms was extrapolated to the complete genome of *P. tricornutum* corresponding to 254,580 SNPs and 11,275 IN/DELs.

The genetic polymorphism of the control population (*c*) was investigated. Since this population was put though all the experimental steps except EMS, it displays the impact of the procedure itself on the genetic diversity of a *P. tricornutum* population. Slight increases of 3 449 SNPs and 699 IN/DELs, when extrapolated to the complete genome of *P. tricornutum*, were identified as being induced by the procedure. This polymorphism represents a diversity variation of 1.56% compared with the reference strain (*R*) and indicates that the experimental procedure has a low impact.

Regarding the type of polymorphisms generated by EMS, SNPs were more frequent than IN/DELs, as observed in the literature (Figure 3). The number of polymorphisms was not equally affected for all survival rates. As shown in Figure 3, the number of SNP and IN/DELs remained unchanged in the different populations, except for the population with a survival rate of 26%. EMS at 3.925 M result had a 26% survival rate and generated 17,182 new polymorphic bases when extrapolated to the complete genome of *P. tricornutum*. Only this population showed an increase in the polymorphism ratio from 1.55 to 7.55 (Figure 4). The other EMS treatments resulted in different survival rates, while the polymorphism they induced was not significantly different from that of control conditions. The mutations identified in this study are those not yet removed by selection and drift, fixed in enough cells to be detectable by sequencing, and generated in a population undergoing selection. This points to an inherent impact of the experimental protocol rather than a real impact of EMS concentration on DNA, and the importance of the survival rate of mutagenized populations.

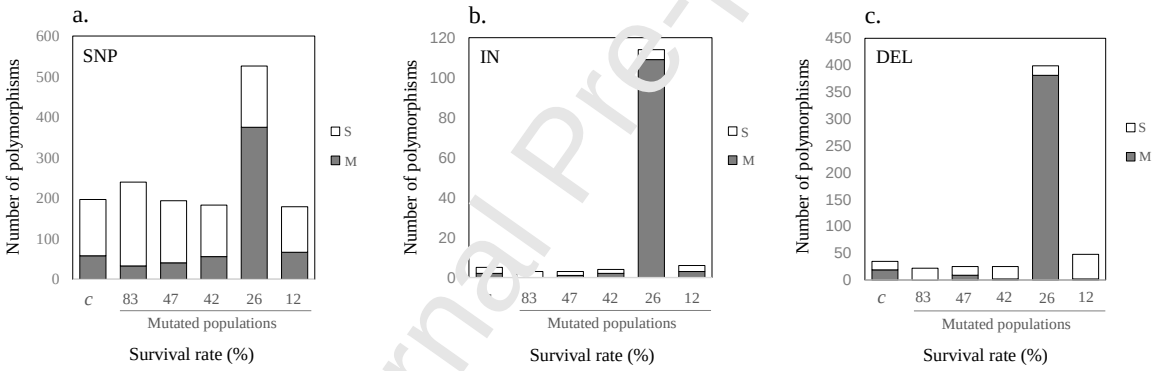

Figure 3. Number of polymorphisms in the control and five EMS-treated populations of *P. tricornutum*. (a) SNPs, (b) INs, (c) DELs resulting from allelic selection (in white) or allelic mutation (in grey). The abscissa axis indicates control (*c*) and mutagenized populations labeled with their respective survival rates.

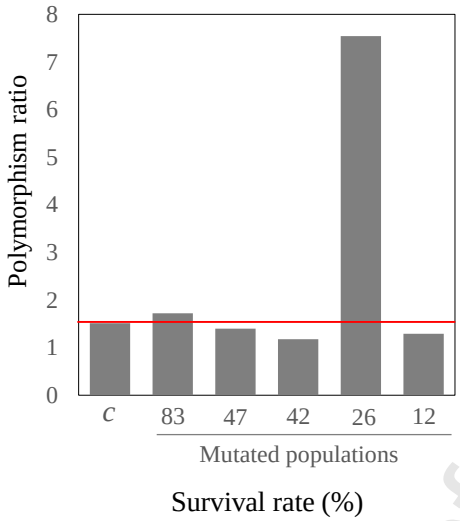

Figure 4. Polymorphism ratio in the control and five EMS-treated populations of *P. tricornutum*. The abscissa axis indicates the control (*c*) and the mutagenized populations labelled with their respective survival rates. The red line indicates the polymorphism rate of the control population (*c*).

3.4. Importance of survival rate as a criterion for mutagenesis optimization

Assessing the impact of mutagenesis on cell populations is a key issue for improvement programs. The use of indicators reflecting mutagenesis efficiency would greatly help in developing efficient procedures. Here, based on the predictions of the model, we used survival rate as an indicator for optimizing EMS dose, with a target survival rate of 50%. This study attempts to reveal differences in induced genetic polymorphism with increasing concentration of EMS mutagen. Of the five concentrations of EMS tested, only the population with the 26% survival rate has a higher polymorphism ratio (by 4.97) than the reference strain (Figure 4). We were not able to find any similar studies on microalgae in the literature to specifically compare our results, but Godfroy et al., (2015) [42] conducted an EMS mutagenesis study on *Ectocarpus* sp. These authors reported a polymorphism frequency of 7.72^{e-7}, computed on the basis of six generations, which is similar to the 1.45^{e-5} frequency observed here for the 26% survival rate in *P. tricornutum*. Spontaneous mutation rate has recently been estimated for some microalgae species: *Emiliania huxleyi* per nucleotide per cell division mutation rate ($\mu = 5.55^{e-10}$) [34] is close to estimates for other eukaryotic plankton such as the diatom *P. tricornutum* ($\mu = 4.77^{e-10}$) or unicellular green alga species *Ostreococcus tauri* ($\mu = 4.79^{e-10}$) and *Chlamydomonas reinhardtii* ($\mu = 3.23^{e-10}$) [33–35]. The mutation rate of *P. tricornutum* in our results is higher than the mutation rates of the order $\mu = 5^{e-10}$ determined across Haptophytes, Stramenopiles, and Chlorophyta.

Under these conditions (26% survival rate), we observed 1 mutation every 261 kbp, which is in line with the results observed with EMS mutagenesis in the model plant *Arabidopsis thaliana* (1 mutation every 400 kbp) [31], *Pisum sativum* (1 mutation every 200 kbp) [30], and *Cucumis melo* (1 mutation every 146 kbp at 848 kbp) [29]. However, these comparisons have important limitations, including the differences in the mutagenesis protocols used, the variable doses of EMS tested and the specificities of the tested organisms.

Our modeling approach to mutagenesis optimization is based on two hypotheses. The first one is that survival rate is a linear function of EMS dose, which is supported by the R^2 of 0.87 we found, as well as findings of previous studies [9,32,38]. The second hypothesis postulates that the percentage of mutant cells among surviving cells is also a linear function of EMS dose. To test this hypothesis, we used the GBS approach with the allelic frequency as a proxy to estimate the mutant frequency in surviving cells Y_m. Our results found a discrepancy between the qualitative modeling approach and the experimental mutation occurrence as measured by GBS. Conversely to the second hypothesis of the model, the GBS experiment resulted in a clear threshold effect, with the maximum frequency of mutations corresponding to a 26% survival rate (3.92 M EMS dose), while no effect was observed at lower doses. Several studies have addressed the question of the dose-response relationship with chemical mutagens. The question of a threshold response has been particular debated. Tano and Yamaguchi reported a threshold effect for N-nitrosodimethylamine (NDMA) in *Tradescantia*, while no such effect was observed for EMS [39]. To add to the complexity, the authors highlighted the specificity of the response, with different responses observed in two different clones of the same species. Conversely, Jenkins et al. (2005) [40] reported a threshold effect for MMS, and Dobo et al. (2011) [39] observed a threshold effect for both EMS and ethylnitrosoureain rats. The threshold effect can be accounted for by DNA repair and other protection mechanisms [39,41]. Further investigation, focusing on the range 42-12% survival rate, is therefore needed to better describe and understand the effect of EMS on polymorphism.

For the highest dose tested here (4M EMS), the allelic polymorphism decreased back to the level below the above-mentioned threshold. The reason for this effect is probably a very high mutation load that results in too many deleterious mutations, causing cell death, thus removing the new diversity generated. It can be hypothesized that this drop in diversity at a

very low survival rate could result from a cytotoxic effect through indirect, non-canonical effects [42] or deleterious mutations at low concentrations, reducing the survival rate of populations. Indeed, several studies highlighted the toxicity of EMS on animal [43] and human cells [44]. To uncover the underlying reasons, it would be necessary to test more EMS doses in the 3.92–4.00 M range and make the associated GBS analyzes.

According to GBS, allelic frequency increased over a very narrow range of EMS dose. When applied to the modeling approach, the threshold effect would result in a shift of the maximum mutant number N_m towards lower survival rates. This trend is supported by results of another recent experimental study on *P. tricornutum* populations exposed to EMS, where the highest phenotypic diversity for carotenoid was observed at a survival rate of 29% (compared with 58%) [6]. Accordingly, the observed threshold effect decreases the predictive capacity of our modeling approach since the threshold must be assessed before predicting the optimal survival rate.

3.5. The GBS approach for microalga improvement programs

Over the past decade, next generation sequencing (NGS) technologies have revolutionized sequencing capabilities by massive analysis of low-cost data [45]. Applied to microalgae, NGS technologies have, in particular, provided new insights for studies characterizing the diversity, dynamics and metabolic pathways of phytoplankton communities [46,47]. Nevertheless, these new techniques face challenges related to genome size and the level of genome complexity of the species under investigation. The GBS approach is a recent NGS-based technique for the discovery of SNPs and allows the reduction of genome representations by the use of restriction enzymes. In contrast to whole genome sequencing, this method is to have a subset of a genome sampled and sequenced. [45]. Applied to plants [24,48,49], mammals [50], insects [51], and fish [52], the GBS approach for the identification of SNPs for population genomic investigations has also recently been applied to the harmful benthic dinoflagellate *Ostreopsis* cf. ovata [45]. This dinoflagellate is known to produce palytoxin-like compounds, which are listed among the most dangerous marine toxins [53]. However, using the GBS approach for the analyses of induced polymorphism in microalgae improvement programs has not yet been assessed. It provides crucial information in the analysis of SNPs of mutant populations and makes it possible to optimize the efficiency of the mutagenesis procedure. In this study, GBS coupled with a modeling approach allowed us to gain a better understanding of the optimal conditions for EMS mutagenesis and to propose

survival rate as a proxy to optimize the experimental conditions of mutagenesis. In strain improvement programs, EMS-mutagenized populations with a survival rate around 30% would show the highest genetic diversity and provide the maximum number of mutant cells, thereby increasing the probability of obtaining individuals with a phenotype of interest. We underline that a protocol for optimizing mutagenesis is an important step toward strain improvement that would make it possible to increase the speed of algal domestication and increase the chances of obtaining traits of interest. A more comprehensive analysis could provide more ways to optimize the GBS procedure for microalgal improvement programs, including: (1) the percentage of the genome analyzed, as we probably underestimated the number of mutations in the genome of *P. tricornutum* populations to some extent because only 5.6% of the genome was sequenced at 50X coverage; (2) the number of mutagenized populations examined, to refine the effect on and the trend of survival rate (Y_s) and mutants among the surviving cells (Y_m), within the efficient range of EMS. Thus, in order to develop an adequate tool for population genomic analyses based on GBS, precise optimization steps are required, which will vary according to the taxa being investigated [45].

4. Conclusion

The application of the GBS approach to microalgae is innovative in a research program on strain improvement. The GBS approach evaluated the created genetic diversity (SNPs and IN/DELs) of *P. tricornutum* populations with different survival rates following EMS-induced mutagenesis. Studies on mutagenesis in the literature have examined survival rates of populations or organisms in the range 10–90%, but without a clear rationale or real gain in genetic polymorphism. This study is the first to assess the effect of mutagenesis on induced-polymorphism in a microalga, and reveals higher generated polymorphism at a survival rate of approximately 30% in the diatom *P. tricornutum* following EMS-induced mutagenesis. It opens new perspectives for the optimized management of microalga improvement programs, which contribute to the expansion of their large-scale production.

Abbreviations

DEL: Deletion polymorphism; EMS: Ethyl methanesulfonate; GBS: Genotyping-by-sequencing; IN: Insertion polymorphism; SNP: Single nucleotide polymorphism.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author contributions

Judith Rumin: **Conceptualization, Methodology, Formal analysis, Investigation, Writing - Review & Editing.** Grégory Carrier: **Supervision, Conceptualization, Methodology, Formal analysis, Investigation, Writing - Review & Editing.** Catherine Rouxel: **Methodology, Formal analysis.** Aurélie Charrier: **Methodology, Formal analysis.** Virginie Raimbault: **Methodology, Formal analysis.** Jean-Paul Cadoret**: Funding acquisition.** Gaël Bougaran**: Project administration, Supervision, Conceptualization, Methodology, Investigation, Writing - Review & Editing.** Bruno Saint-Jean: **Supervision, Conceptualization, Methodology, Investigation, Writing - Review & Editing.**

Acknowledgements

This work was supported by the Facteur 4 ANR project (ANR-12-BIME-0004). We thank Audrey Donnard, UFR Médecine et Techniques Médicales, University of Nantes, for assistance with the 2200 TapeStation and the Illumina HiSeq 2000 sequencer (Illumina Corporation Inc.), Sophie Michon-Coudouel for assistance with the LabChip XT DNA (Caliper Life Technologies) and Amandine Launay and Muriel Latreille (UMR AGAP) for advice on the GBS protocol. We thank Helen McCombie-Boudry for the proofreading of the manuscript.

Statement of Informed Consent

No conflicts, informed consent, or human or animal rights are applicable to this study.

References

[1] A.J. Klok, P.P. Lamers, D.E. Martens, R.B. Draaisma, R.H. Wijffels, Edible oils from microalgae: insights in TAG accumulation, Trends Biotechnol. 32 (2014) 521–528. https://doi.org/10.1016/j.tibtech.2014.07.004.[2] P. Spolaore, C. Joannis-Casse, E. Duran, A. Isambert, Commercial applications of microalgae, J Biosci Bioeng. 101 (2) (2006) https://doi.org/10.1263/jbb.101.87.

[3] V. Mimouni, L. Ulmann, E. Hansen, V. Pasquet, M. Mathieu, L. Picot, G. Bougaran, J-P. Cadoret, A. Morant-Manceau, B. Schoefs, The potential of microalgae for the production of bioactive molecules of pharmaceutical interest, Curr. Pharma. Biotechnol., 13 (2012). https://doi.org/10.2174/138920112804724828

[4] C. Lauritano, J.H. Andersen, E. Hansen, M. Albrigtsen, L. Escalera, F. Esposito, K. Helland, K.Ø. Hanssen, G. Romano, A. Ianora, Bioactivity Screening of Microalgae for Antioxidant, Anti-Inflammatory, Anticancer, Anti-Diabetes, and Antibacterial Activities, Front. Mar. Sci. 3 (2016). https://doi.org/10.3389/fmars.2016.00068.

[5] W. Fu, A. Chaiboonchoe, B. Khraiwesh, D.R. Nelson, D. Al-Khairy, A. Mystikou, A. Alzahmi, K. Salehi-Ashtiani, Algal Cell Factories: Approaches, Applications, and Potentials, Mar. Drugs. 14 (2016). https://doi.org/10.3390/md14120225.

[6] Z. Yi, Y. Su, M. Xu, A. Bergmann, S. Ingthorsson, O. Rolfsson, K. Salehi-Ashtiani, S. Brynjolfsson, W. Fu, Chemical Mutagenesis and Fluorescence-Based High-Throughput Screening for Enhanced Accumulation of Carotenoids in a Model Marine Diatom Phaeodactylum tricornutum, Mar. Drugs. 16 (2018). https://doi.org/10.3390/md16080272.

[7] A. Ghosh, S. Khanra, M. Mondal, G. Halder, O.N. Tiwari, S. Saini, T.K. Bhowmick, K. Gayen, Progress toward isolation of strains and genetically engineered strains of microalgae for production of biofuel and other value added chemicals: A review, Energy Convers. Manag. 113 (2016) 104–118. https://doi.org/10.1016/j.enconman.2016.01.050.

[8] R.T. Rowlands, Industrial strain improvement: Mutagenesis and random screening procedures, Enzyme Microb. Technol. 6 (1984) 3–10. https://doi.org/10.1016/0141-0229(84)90070-X.

[9] T.T.Y. Doan, J.P. Obbard, Enhanced intracellular lipid in Nannochloropsis sp. via random mutagenesis and flow cytometric cell sorting, Algal Res. 1 (2012) 17–21. https://doi.org/10.1016/j.algal.2012.03.001.

[10] D. Nojima, Y. Ishizuka, M. Muto, A. Ujiro, F. Kodama, T. Yoshino, Y. Maeda, T. Matsunaga, T. Tanaka, Enhancement of Biomass and Lipid Productivities of Water Surface-Floating Microalgae by Chemical Mutagenesis, Mar. Drugs. 15 (2017). https://doi.org/10.3390/md15060151.

[11] G. Perin, A. Bellan, A. Segalla, A. Meneghesso, A. Alboresi, T. Morosinotto, Generation of random mutants to improve light-use efficiency of Nannochloropsis gaditana cultures for biofuel production, Biotechnol. Biofuels. 8 (2015) 161. https://doi.org/10.1186/s13068-015-0337-5.

[12] R. Chaturvedi, Y. Fujita, Isolation of enhanced eicosapentaenoic acid producing mutants of Nannochloropsis oculata ST-6 using ethyl methane sulfonate induced mutagenesis techniques and their characterization at mRNA transcript level, Phycol. Res. 54 (2006) 208–219. https://doi.org/10.1111/j.1440-1835.2006.00428.x.

[13] T. Flynn, M.L. Ghirardi, M. Seibert, Accumulation of O2-tolerant phenotypes in H2-producing strains of Chlamydomonas reinhardtii by sequential applications of chemical mutagenesis and selection, Int. J. Hydrog. Energy. 27 (2002) 1421–1430. https://doi.org/10.1016/S0360-3199(02)00117-9.

[14] B. Xie, D. Stessman, J.H. Hart, H. Dong, Y. Wang, D.A. Wright, B.J. Nikolau, M.H. Spalding, L.J. Halverson, High-throughput fluorescence-activated cell sorting for lipid hyperaccumulating Chlamydomonas reinhardtii mutants, Plant Biotechnol. J. 12 (2014) 872–882. https://doi.org/10.1111/pbi.12190.

[15] J.W. Davey, P.A. Hohenlohe, P.D. Etter, J.Q. Boone, J.M. Catchen, M.L. Blaxter, Genome-wide genetic marker discovery and genotyping using next-generation sequencing, Nat. Rev. Genet. 12 (2011) 499–510. https://doi.org/10.1038/nrg3012.

[16] R.J. Elshire, J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, S.E. Mitchell, A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species, PLOS ONE. 6 (2011) e19379. https://doi.org/10.1371/journal.pone.0019379.

[17] Y.-F. Niu, M.-H. Zhang, D.-W. Li, W.-D. Yang, J.-S. Liu, W.-B. Bai, H.-Y. Li, Improvement of Neutral Lipid and Polyunsaturated Fatty Acid Biosynthesis by Overexpressing a Type 2 Diacylglycerol Acyltransferase in Marine Diatom Phaeodactylum tricornutum, Mar. Drugs. 11 (2013) 4558–4569. https://doi.org/10.3390/md11114558.

[18] Z.-K. Yang, Y.-F. Niu, Y.-H. Ma, J. Xue, M.-H. Zhang, W.-D. Yang, J.-S. Liu, S.-H. Lu, Y. Guan, H.-Y. Li, Molecular and cellular mechanisms of neutral lipid accumulation in diatom following nitrogen deprivation, Biotechnol. Biofuels. 6 (2013) 67. https://doi.org/10.1186/1754-6834-6-67.

[19] M.L. Hamilton, R.P. Haslam, J.A. Napier, O. Sayanova, Metabolic engineering of Phaeodactylum tricornutum for the enhanced accumulation of omega-3 long chain polyunsaturated fatty acids, Metab. Eng. 22 (2014) 3–9. https://doi.org/10.1016/j.ymben.2013.12.003.

[20] R. Radakovits, P.M. Eduafo, M.C. Posewitz, Genetic engineering of fatty acid chain length in Phaeodactylum tricornutum, Metab. Eng. 13 (2011) 89–95. https://doi.org/10.1016/j.ymben.2010.10.003.

[21] P.R. Walne, Experiments in the large-scale culture of the larvae of Ostrea edulis L., H.M.S.O., London, 1966.

[22] J.-Y. Cho, J.-S. Choi, I.-S. Kong, S.-I. Park, R.G. Kerr, Y.-K. Hong, A procedure for axenic isolation of the marine microalga Isochrysis galbana from heavily contaminated mass cultures, J. Appl. Phycol. 14 (2002) 385–390. https://doi.org/10.1023/A:1022199332224.

[23] 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 (2012) 2737–2745. https://doi.org/10.1002/bit.24560.

[24] H. Sonah, M. Bastien, E. Iquira, A. Tardivel, G. Légaré, B. Boyle, É. Normandeau, J. Laroche, S. Larose, M. Jean, F. Belzile, An Improved Genotyping by Sequencing (GBS) Approach Offering Increased Versatility and Efficiency of SNP Discovery and Genotyping, PLoS ONE. 8 (2013). https://doi.org/10.1371/journal.pone.0054603.

[25] B. Ewing, P. Green, Base-calling of automated sequencer traces using phred. II. Error probabilities, Genome Res. 8 (1998) 186–194.

[26] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, EMBnet.Journal. 17 (2011) 10–12. https://doi.org/10.14806/ej.17.1.200.

[27] W.-P. Lee, M.P. Stromberg, A. Ward, C. Stewart, E.P. Garrison, G.T. Marth, MOSAIK: A Hash-Based Algorithm for Accurate Next-Generation Sequencing Short-Read Mapping, PLOS ONE. 9 (2014) e90581. https://doi.org/10.1371/journal.pone.0090581.

[28] F.E. Dewey, M.E. Grove, C. Pan, B.A. Goldstein, J.A. Bernstein, H. Chaib, J.D. Merker, R.L. Goldfeder, G.M. Enns, S.P. David, N. Pakdaman, K.E. Ormond, C. Caleshu, K. Kingham, T.E. Klein, M. Whirl-Carrillo, K. Sakamoto, M.T. Wheeler, A.J. Butte, J.M. Ford, L. Boxer, J.P.A. Ioannidis, A.C. Yeung, R.B. Altman, T.L. Assimes, M. Snyder, E.A. Ashley, T. Quertermous, Clinical interpretation and implications of whole-genome sequencing, JAMA. 311 (2014) 1035–1045. https://doi.org/10.1001/jama.2014.1717.

[29] F. Dahmani-Mardas, C. Troadec, A. Boualem, S. Lévêque, A.A. Alsadon, A.A. Aldoss, C. Dogimont, A. Bendahmane, Engineering melon plants with improved fruit shelf life using the TILLING approach, PloS One. 5 (2010) e15776. https://doi.org/10.1371/journal.pone.0015776.

[30] M. Dalmais, J. Schmidt, C. Le Signor, F. Moussy, J. Burstin, V. Savois, G. Aubert, V. Brunaud, Y. de Oliveira, C. Guichard, R. Thompson, A. Bendahmane, UTILLdb, a Pisum sativum in silicoforward and reverse genetics tool, Genome Biol. 9 (2008) R43. https://doi.org/10.1186/gb-2008-9-2-r43.

[31] C.M. McCallum, L. Comai, E.A. Greene, S. Henikoff, Targeted screening for induced mutations, Nat. Biotechnol. 18 (2000) 455–457. https://doi.org/10.1038/74542.

[32] R. Loppes, Ethyl methanesulfonate: an effective mutagen in Chlamydomonas reinhardi, Mol. Gen. Genet. MGG. 102 (1968) 229–231. https://doi.org/10.1007/bf00385978.

[33] W.L. Russell, E.M. Kelly, P.R. Hunsicker, J.W. Bangham, S.C. Maddux, E.L. Phipps, Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse., Proc. Natl. Acad. Sci. U. S. A. 76 (1979) 5818. https://doi.org/10.1073/pnas.76.11.5818.
[34] M. Krasovec, R.E.M. Rickaby, D.A. Filatov, Evolution of Mutation Rate in Astronomically Large Phytoplankton Populations, Genome Biol. Evol. 12 (2020) 1051–1059. https://doi.org/10.1093/gbe/evaa131.
[35] M. Krasovec, S. Sanchez-Brosseau, G. Piganeau, First Estimation of the Spontaneous Mutation Rate in Diatoms, Genome Biol. Evol. 11 (2019) 1829–1837. https://doi.org/10.1093/gbe/evz130.
[36] M. Krasovec, A. Eyre-Walker, S. Sanchez-Ferandin, G. Piganeau, Spontaneous Mutation Rate in the Smallest Photosynthetic Eukaryotes, Mol. Biol. Evol. 34 (2017) 1770–1779. https://doi.org/10.1093/molbev/msx119.
[37] R.W. Ness, A.D. Morgan, N. Colegrave, P.D. Keightley, Estimate of the Spontaneous Mutation Rate in Chlamydomonas reinhardtii, Genetics. 192 (2012) 1447–1454. https://doi.org/10.1534/genetics.112.145078.
[38] U. Tripathi, G. Venkateshwaran, R. Sarada, G.A. Ravishankar, Studies on Haematococcus pluvialis for improved production of astaxanthin by mutagenesis, World J. Microbiol. Biotechnol. 17 (2001) 143–148. https://doi.org/10.1023/A:1016609815405.
[39] S. Tano, H. Yamaguchi, Effects of several nitroso compounds on the induction of somatic mutations in Tradescantia with special regard to the dose response and threshold dose, Mutat. Res. Mol. Mech. Mutagen. 148 (1985) 59–64. https://doi.org/10.1016/0027-5107(85)90207-6.
[40] G.J.S. Jenkins, S.H. Doak, G.E. Johnson, E. Quick, E.M. Waters, J.M. Parry, Do dose response thresholds exist for genotoxic alkylating agents?, Mutagenesis. 20 (2005) 389–398. https://doi.org/10.1093/mutage/gei054.
[41] K.L. Dobo, R.D. Fiedler, W.C. Gunther, C.J. Thiffeault, Z. Cammerer, S.L. Coffing, T. Shutsky, M. Schuler, Defining EMS and ENU dose–response relationships using the Pig-a mutation assay in rats, Mutat. Res. Toxicol. Environ. Mutagen. 725 (2011) 13–21. https://doi.org/10.1016/j.mrgentox.2011.06.005.
[42] O. Godfroy, A.F. Peters, S.M. Coelho, J.M. Cock, Genome-wide comparison of ultraviolet and ethyl methanesulphonate mutagenesis methods for the brown alga Ectocarpus, Mar. Genomics. 24 (2015) 109–113. https://doi.org/10.1016/j.margen.2015.03.007.
[43] E. Gocke, L. Müller, In vivo studies in the mouse to define a threshold for the genotoxicity of EMS and ENU, Mutat. Res. Toxicol. Environ. Mutagen. 678 (2009) 101–107. https://doi.org/10.1016/j.mrgentox.2009.04.005.
[44] S.H. Doak, G.J.S. Jenkins, G.E. Johnson, E. Quick, E.M. Parry, J.M. Parry, Mechanistic influences for mutation induction curves after exposure to DNA-reactive carcinogens, Cancer Res. 67 (2007) 3904–3911. https://doi.org/10.1158/0008-5472.CAN-06-4061.
[45] S. Casabianca, L. Cornetti, S. Capellacci, C. Vernesi, A. Penna, Genome complexity of harmful microalgae, Harmful Algae. 63 (2017) 7–12. https://doi.org/10.1016/j.hal.2017.01.003.
[46] S.M. An, D.H. Choi, H. Lee, J.H. Lee, J.H. Noh, S.M. An, D.H. Choi, H. Lee, J.H. Lee, J.H. Noh, Next-generation sequencing reveals the diversity of benthic diatoms in tidal flats, ALGAE. 33 (2018) 167–180. https://doi.org/10.4490/algae.2018.33.4.3.
[47] J.A. Visco, L. Apothéloz-Perret-Gentil, A. Cordonier, P. Esling, L. Pillet, J. Pawlowski, Environmental Monitoring: Inferring the Diatom Index from Next-Generation Sequencing Data, Environ. Sci. Technol. 49 (2015) 7597–7605. https://doi.org/10.1021/es506158m.

[48] Y.-F. Huang, J.A. Poland, C.P. Wight, E.W. Jackson, N.A. Tinker, Using Genotyping-By-Sequencing (GBS) for Genomic Discovery in Cultivated Oat, PLOS ONE. 9 (2014) e102448. https://doi.org/10.1371/journal.pone.0102448.

[49] J. Poland, J. Endelman, J. Dawson, J. Rutkoski, S. Wu, Y. Manes, S. Dreisigacker, J. Crossa, H. Sánchez-Villeda, M. Sorrells, J.-L. Jannink, Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing, Plant Genome. 5 (2012) 103–113. https://doi.org/10.3835/plantgenome2012.06.0006.

[50] M.D. Donato, S.O. Peters, S.E. Mitchell, T. Hussain, I.G. Imumorin, Genotyping-by-Sequencing (GBS): A Novel, Efficient and Cost-Effective Genotyping Method for Cattle Using Next-Generation Sequencing, PLOS ONE. 8 (2013) e62137. https://doi.org/10.1371/journal.pone.0062137.

[51] K.L. Silva-Brandão, O.A.B.N. e Silva, M.M. Brandão, C. Omoto, F.A.H. Sperling, Genotyping-by-sequencing approach indicates geographic distance as the main factor affecting genetic structure and gene flow in Brazilian populations of Grapholita molesta (Lepidoptera, Tortricidae), Evol. Appl. 8 (2015) 476–485. https://doi.org/10.1111/eva.12257.

[52] C. Li, G. Waldbieser, B. Bosworth, B.H. Beck, W. Thongda, E. Peatman, SNP discovery in wild and domesticated populations of blue catfish, Ictalurus furcatus, using genotyping-by-sequencing and subsequent SNP validation, Mol. Ecol. Resour. 14 (2014) 1261–1270. https://doi.org/10.1111/1755-0998.12272.

[53] P. Ciminiello, C. Dell'Aversano, E. Dello Iacovo, E. Fattorusso, M. Forino, L. Tartaglione, G. Benedettini, M. Onorari, F. Serena, C. Battocchi, S. Casabianca, A. Penna, First finding of Ostreopsis cf. ovata toxins in marine aerosols, Environ. Sci. Technol. 48 (2014) 3532–3540. https://doi.org/10.1021/es405617d.

Highlights

- A literature data-based modeling approach was used to assess the best mutagenesis conditions aiming to maximize the genetic diversity.
- Survival rate of EMS mutagenized-population of *P. tricornutum* is an helpfull criterion for mutagenesis optimization
- Mutagenesis experiments on *P. tricornutum* showed that the survival rate (83 to 12%) has a negative linear trend as a function of the EMS dose applied (3.65 to 4.00 M).
- Allelic frequency showed a threshold effect with EMS dose, with the maximum frequency of mutations obtained for 26% survival rate (3.92 M EMS dose).