Algal Research 35 (2018) 85–90

Contents lists available at ScienceDirect

Algal Research

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

Spontaneous mutation rate as a source of diversity for improving desirable traits in cultured microalgae

Marc Krasovec[1], Sophie Sanchez-Brosseau, Nigel Grimsley, Gwenael Piganeau*

Sorbonne Universités, UPMC Univ Paris 06, CNRS, Biologie Intégrative des Organismes Marins (BIOM), Observatoire Océanologique, F-66650 Banyuls/Mer, France

ARTICLE INFO

Keywords:
Mutation rate
Mutation accumulation
Green algae biotechnology
Picochlorum
Domestication

ABSTRACT

Mutations are the main origin of the biodiversity and biological innovations across the tree of life. The number of mutations in a population depends of the mutation rate, noted μ, a key parameter for understanding the evolutionary and adaptive capacity of a species. New mutations are submitted to selection and drift and their probability of fixation in a population depends on their advantageous, deleterious or neutral fitness effect. This process occurs in natural populations, but also in any lab cultures. In this context, the role of spontaneous mutations in the generation of genetic diversity in cultured algae has so far been overlooked, despite its influence on the acquisition and maintenance of desirable phenotypic traits. Several algal species that have a high biotechnological potential, such those producing high-value molecules, might be improved by domestication and oriented selection by experimental evolution. Here, we provide the first estimation of the spontaneous mutation rate, μ, in *Picochlorum costavermella* (Trebouxiophyceae), a green alga with many potential biotechnological applications. Its spontaneous mutation rate is $\mu = 10.12 \times 10^{-10}$ (CI Poisson distribution, $\mu = 6.3–15.4 \times 10^{-10}$) mutations per nucleotide per genome per generation. This is one of the highest mutation rates reported for a unicellular eukaryote.

1. Introduction

Natural selection enables species to adapt from standing genetic variation, powered by new mutations, which constitute the main source of genetic diversity in a population. Quantifying the rate of mutations and their effects is thus of primary importance to better understand the evolution and adaptive potential of a species. Beyond the evolutionary importance of mutations, generation of genetic diversity is also relevant for the domestication process, as one single mutation can trigger a novel trait of interest in a species [1]. Domestication of crops and animals began several thousand years ago and involved manual selection and breeding for thousands of generations [2,3] and more recently genetic engineering [4]. In contrast to domesticated macro-organisms, which have long generation times, obligate sexual reproduction, and small population sizes, microorganisms have short generation times, asexual reproduction and huge population sizes. As a consequence, rare events, like spontaneous mutations, can be observed and monitored within a short time lag, typically within a few months in microbial cultures [5]. In the context of the global challenge caused by the scarcity of natural resources, the domestication of novel species and the development of biotechnologies are evoked as promising strategies [6]. Green microalgae have been brought into the spotlight as favoured candidates for biotechnological development [7]. Among other advantages, they have a short asexual life cycle, many can be grown in marine water, many are transformable and their biotechnological potential is manifold such as the production of biofuels [8–13], the production of proteins and high value molecules for health food or cosmetics [14–16], the production of hydrogen [17], and their photosynthetic production can be optimized [18].

Phenotypic traits of interest may be further increased by experimental evolution [19] and directional selection of algal strains for biotechnological applications. Any natural population harbours a standing genetic diversity [20], which enables adaptation to environmental changes and selection pressures. Traits of interest may be selected from this standing genetic diversity, but very little data about natural diversity in microalga is available yet [21,22]. Alternatively, adaptation may also occur from new mutations. Domestication of land plants revealed several cases in which one single nucleotide mutation induced the trait of interest ([23] for a review). The estimation of the spontaneous mutation rates is paramount for investigating the possibility of generating phenotypic trait of interest by mutations. The probability of obtaining a desirable genotype may be also increased by

* Corresponding author.
E-mail address: gwenael.piganeau@obs-banyuls.fr (G. Piganeau).
[1] Present address: Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, United Kingdom.

https://doi.org/10.1016/j.algal.2018.08.003
Received 27 April 2018; Received in revised form 6 July 2018; Accepted 6 August 2018

mutagenesis. Different protocols have been developed, such as heavy-ion irradiation in *Parachlorella kessleri* [24] and *Desmodesmus* spp. [25], gamma rays in *Chlorella* spp. [26,27] or UV irradiation in *Chlorella sorokiniana* [28,29]. However, mutagenesis has three pitfalls. First, it dramatically increases the deleterious mutation load. Algal populations must therefore be carefully exposed to a mutagen in order to not compromise their survival capacity. In the case of ultraviolet (UV) irradiation, the survival rate may reach only 10% in *Chlorella sorokiniana* [28,29]. Second, it indirectly selects for cells resistant to the mutagenic factor. Third, microorganisms derived from chemical or radiation mutagenesis are included under the Genetically Modified Organisms umbrella and are therefore submitted to many commercial and political conditions [30].

The spontaneous mutation rate of any organisms can be estimated by mutation accumulation (MA) experiments, and this ideally requires a high quality reference genome sequence [31–33]. The development of high throughput genome sequencing technologies led to a significant improvement in the practicability of whole genome sequencing, mirrored in an increase in the number of available complete and annotated reference genomes in microalga in the last decade [34]. The principle of MA experiment is to monitor genetic changes in MA lines maintained at a low population size by serial bottlenecks [35], as effective population size is inversely proportional to the strength of selection. Thus, a very low N_e drastically limits the efficiency of selection, allowing as many spontaneous mutations as possible to become fixed, including the deleterious ones, in order to estimate to the total number of spontaneous mutations [36]. These events are identified by comparing the MA genomes to the ancestral genome sequence. The monitoring of the number of cell divisions throughout the experiment is required to estimate the mutation rate per cell division. Mutation accumulation approaches have been applied to a large range of species, including bacteria and eukaryotes [32]. In green algae, MA studies have been conducted in a few model species and have allowed the fitness effects of mutations [5,37,38] and spontaneous mutation rates to be estimated in *Chlamydomonas reinhardtii* [39–41] and four Mamiellophyceae species; *Micromonas commoda* RCC299, *Bathycoccus prasinos* RCC1105, *Ostreococcus mediterraneus* RCC2590, and *O. tauri* RCC4221 [42].

In this study, we exemplify use of a mutation accumulation experiment in *Picochlorum* RCC4223, a green algal species of the Trebouxiophyceae class (Chlorophyta) [43]. We used the same protocol as previously described [5], where all cultures are maintained in liquid medium and cell count is estimated by flow cytometry. *Picochlorum* RCC4223 has a small haploid genome of ~13.5 Mb with 79.5% of coding sequences and 46% GC content [44]. Strains from the genus *Picochlorum* are versatile algae for large-scale culturing, capable of growing in a wide range of salinities and temperatures [45,46] with a few completely sequenced genomes already available, including the strain RCC4223 [46,47]. They also constitute interesting models in different fields, such as medicine [48], biofuels [49], food and aquaculture [50,51] and production of high value molecules [52].

2. Materials and methods

2.1. Mutation accumulation (MA) experiment

The *Picochlorum costavermella* RCC4223 strain was isolated from an estuary of the river “La Massane” (42°32′36 N, 3°03′09 E, France) and has been deposited in the Roscoff Culture Collection [53] (http://roscoff-culture-collection.org/).

Twelve MA lines were kept from a clonal ancestral population in 24-well plates, at 20 °C with a light cycle of 16 h–8 h dark-light in L1 medium. MA lines were inoculated as single cells after endpoint dilution and maintained by serial one-cell bottlenecks every 14 days, as described in [5]. At each bottleneck, cell concentration was measured with a FACSCanto II flow cytometer (Becton Dickinson, Franklin Lakes, NJ, U.S.A.) using natural chlorophyll fluorescence (670 nm used FL3 data) and SSC acquisitions. Effective population size, N_e, was estimated with the harmonic mean of cell number between bottlenecks and the number of generations (number of cell divisions) being provided by the following equation:

$$d = e^{\left[\ln\left(\frac{N_t}{N_0}\right)/t\right]} \quad (1)$$

where N_t is the total number of cells measured by flow cytometry, and t the time between two bottlenecks, *i.e* 14 days. Lines were maintained for 199 days and went through a series of 14 bottlenecks.

2.2. Sequencing and mutations identification

We extracted DNA using a CTAB protocol [54] and sequenced by Illumina MiSeq (200 bp paired-end reads) performed by GATC biotech® (Konstanz, Germany). Twelve MA lines and the ancestral type were sequenced. MiSeq reads were aligned to the reference genome with BWA *mem* [55], *bam* files were treated with *SAMtools* [56] and mutations were called with *GATK HaplotypeCaller* [57]. Afterwards, final *vcf* files and candidate mutations were obtained after following filtered steps: removal of low mapping quality sites (MQ < 40), low covered sites (DP < 10) and candidates shared by two MA lines. *SnpEff* program [58] permitted synonymous, non-synonymous, intronic and intergenic mutations to be identified using the annotation available on the ORCAE web site [59]. This mutation calling pipeline was used for base-substitution and insertions-deletions (indels).

The mutation rate was calculated by dividing the total number of *de novo* mutations by the total generations and the callable genome size to give the mutation rate per nucleotide per genome per generation (μ). The mutation rate per genome per generation is given by $U = G\text{x}\mu$, where G is the genome size.

Raw data are available in GenBank with Bioproject accessions PRJNA453760 and PRJNA389600 (Table S1).

2.3. Mutation spectrum

Pearson's chi-squared was used to test the distribution of observed mutations with the expected distribution, H_0, assuming that mutations appear randomly and independently in the genome. We compared the distribution of mutations between coding and non-coding regions; the level of expression of mutated sites using *STAR* [60]; the synonymous and non-synonymous base-substitution mutations; the direction of mutations from each nucleotide to others and the nucleotide context (between 2 and 10 nucleotides) around mutated sites.

The substitution spectrum between the 4 nucleotides is usually biased, and several previous studies have reported a bias from GC to AT mutations as compared to AT to GC mutations. This mutation bias has a consequence on the equilibrium GC content of the genome, noted GC_{eq}, which would be reached if the GC composition was the consequence of mutation processes alone. At equilibrium GC content, the number of mutations from GC to AT equals the number of mutations from AT to GC. The GC_{eq} can be estimated from the following equations [61]:

$$R_1 = \frac{(\text{GC} \rightarrow \text{AT})}{GC_n}, \quad R_2 = \frac{(\text{AT} \rightarrow \text{GC})}{AT_n}, \quad GCeq = \frac{R_2}{R_1 + R_2},$$

where GC_n and AT_n are the total GC and AT nucleotides of the genome; $GC \rightarrow AT$ and $AT \rightarrow GC$ are the number of *de novo* mutations.

3. Results and discussion

3.1. Mutational spectrum

The experiment lasted 199 days, corresponding to an average of 133 generations per MA line with an average effective population size of N_e ~6. *In silico* analysis identified a total of 21 new mutations in the 12 MA lines: 19 base-substitutions and 2 insertion-deletions (indels) (Table 1).