

## RESEARCH ARTICLE

# Characterization of the Chloroplast Genome Facilitated the Transformation of *Parachlorella kessleri*-I, A Potential Marine Alga for Biofuel Production

Prachi Nawkarkar<sup>1,\*</sup>, Sagrika Chugh<sup>1</sup>, Surbhi Sharma<sup>1</sup>, Mukesh Jain<sup>2</sup>, Sachin Kajla<sup>3</sup> and Shashi Kumar<sup>1,\*</sup>

<sup>1</sup>International Centre for Genetic Engineering and Biotechnology, New Delhi 110067, India; <sup>2</sup>School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi 110067, India; <sup>3</sup>Tata Steel Limited, Research & Development, P O Burmamines, Jamshedpur 831007, India

**Abstract: Introduction:** The microalga *Parachlorella kessleri*-I produces high biomass and lipid content that could be suitable for producing economically viable biofuel at a commercial scale. Sequencing the complete chloroplast genome is crucial for the construction of a species-specific chloroplast transformation vector.

**Methods:** In this study, the complete chloroplast genome sequence (cpDNA) of *P. kessleri*-I was assembled; annotated and genetic transformation of the chloroplast was optimized. For the chloroplast transformation, we have tested two antibiotic resistance makers, aminoglycoside adenine transferase (*aadA*) gene and *Sh-ble* gene conferring resistance to spectinomycin and zeocin, respectively. Transgene integration and homoplasty determination were confirmed using PCR, Southern blot and Droplet Digital PCR.

**Results:** The chloroplast genome (109,642 bp) exhibited a quadripartite structure with two reverse repeat regions (IRA and IRB), a long single copy (LSC), and a small single copy (SSC) region. The genome encodes 116 genes, with 80 protein-coding genes, 32 tRNAs and 4 rRNAs. The cpDNA provided essential information like codons, UTRs and flank sequences for homologous recombination to make a species-specific vector that facilitated the transformation of *P. kessleri*-I chloroplast. The transgenic algal colonies were retrieved on a TAP medium containing 400 mg. L<sup>-1</sup> spectinomycin, but no transgenic was recovered on the zeocin-supplemented medium. PCR and Southern blot analysis ascertained the transgene integration into the chloroplast genome, via homologous recombination. The chloroplast genome copy number in wildtype and transgenic *P. kessleri*-I was determined using Droplet Digital PCR.

**Conclusion:** The optimization of stable chloroplast transformation in marine alga *P. kessleri*-I should open a gateway for directly engineering the strain for carbon concentration mechanisms to fix more CO<sub>2</sub>, improving the photosynthetic efficiency and reducing the overall biofuels production cost.

**Keywords:** Chloroplast genome, genetic engineering, homologous recombination, photosynthetic organism, microalgae biofuels, *parachlorella*.

## 1. INTRODUCTION

*Parachlorella* is a unicellular green alga belonging to the class Trebouxiophyceae in the order *Chlorellales*. The classes Trebouxiophyceae, Prasinophyceae, Ulvophyceae and Chlorophyceae belong to the phylum Chlorophyta [1, 2], amongst which Prasinophyceae show the most basal divergence [3] while Trebouxiophyceae emerged before Ulvophyceae and Chlorophyceae as evident from the reported data of chloroplast and mitochondrial genomes [4-6].

Chloroplasts were captured early during the evolution of the eukaryotic cell, as they are considered having originated from cyanobacteria through endosymbiosis. During the course of evolution, extensive rearrangements occurred within the chloroplast genomes or (plastomes). However, as compared to that of free-living cyanobacteria, the size of the chloroplast genome has subsequently reduced in the past, but still, many DNA sequences of chloroplast resembled the cyanobacterial genome [7]. It is generally believed that land plants have been evolved from green algae [8] and most of the chloroplast genomes are highly conserved among plant species [9]. Most land plant plastomes have two identical copies of inverted repeat regions (IRA and IRB) separating a large single copy (LSC) region and a small single copy

\*Address correspondence to these authors at the International Centre for Genetic Engineering and Biotechnology, New Delhi 110067, India; Tel: +91 8750663713; E-mails: [nprachi2503@gmail.com](mailto:nprachi2503@gmail.com); [skrhode@icgeb.res.in](mailto:skrhode@icgeb.res.in)

(SSC) region [10, 11]. Therefore, insightful information on repeated sequences, intergenic regions, and pseudogenes in the chloroplast DNA might be very helpful in deciphering the process of chloroplast genome evolution.

The genetic transformation of industrial microalgae may be useful for overexpressing the desired traits, leading to the production of sustainable biofuels feedstocks [12, 13]. Chloroplast genome transformation is routine in *Nicotiana tabacum* and *Chlamydomonas reinhardtii*. The first stable chloroplast transformation was achieved in *C. reinhardtii*. However, chloroplast transformation is still challenging in other plant species and algae. Transformation of the nuclear genome was reported for *Parachlorella* sp. [14], but no chloroplast transformation has been reported. In algae, chloroplast organelle occupies about two-thirds of the cellular space and carries up to 100 copies of the plastome. The chloroplast genome is prokaryotic in nature, which enables site-specific recombination between homologous DNA sequences and provides a unique opportunity to overexpress the important traits via plastome modifications [15]. There are no deletions and rearrangements of transgene DNA reported at the site of insertion. This is advantageous over the nuclear transformation that usually leads to the random integration of transgenes [16]. Moreover, transgenes integration in the chloroplast genome is precise and unaffected by phenomena such as pre or post-transcriptional silencing despite transcripts accumulating at 169-fold higher levels comparative to nuclear genome transformation [17]. Due to multiple copies of plastome in an algal cell, chloroplast transgenics may exhibit higher levels of transgene expression, thus a higher level of biofuel molecules and value-added products can be produced. Further expression levels can be enhanced if transgene is integrated into one IR region, which duplicates via the *RecA* enzyme into the other IR as well [10].

The genes of the chloroplast genome are transferred via maternal inheritance in green alga [18]. The uniparental inheritance of chloroplast traits in green algae [18, 19] can be the key to generate genetically modified species and mitigate the environmental risks [20, 21]. This would be beneficial in the biological containment of genetically modified strains when cultured outdoors in large volumes, e.g., raceway ponds and photobioreactors. The chloroplast transformation may be useful to overexpress the genes like thioesterases to produce fatty acids of desired chain lengths, it can be implemented to fix higher CO<sub>2</sub> and reduce photorespiration via carbon concentration mechanism (CCM), as well as improving the photosynthetic efficiency by the overexpression of RuBisCO genes.

Researchers are studying algal-based biofuels since the 1970s, yet no commercially viable strain has been claimed so far. In the past, some programs were shut down due to the higher cost of algal-based biofuels compared to fossil fuels. The advancement of recombinant DNA techniques and synthetic biology offers the greatest range of options to improve the performance of wild strains. Therefore, different transgenic approaches need to be investigated as a potential method for increasing the productivity of algae and enabling them to compete with relatively cheaper fossil fuels. The microalgal strain *P. kessleri*-I accumulates high starch and lipid in laboratory conditions as well as on a semi-industrial

scale in outdoor photobioreactors [22] and considered an ideal strain for biofuel production. It occurs in both freshwater and marine environments.

An oleaginous marine algal strain *P. kessleri*-I, isolated from the Indian Ocean, that accumulates higher lipid amount on a dry weight basis [23] compared to the previously reported freshwater *P. kessleri* strain CCALA 225 [22] was explored for optimization of the chloroplast transformation using *aadA* and *Sh-ble* gene conferring resistance to spectinomycin and zeocin respectively. To obtain an optimal expression of transgenes in the chloroplast, identification of spacer regions and endogenous regulatory sequences are required. Thus, to gain more insights, we have undertaken complete sequencing and annotation of the chloroplast genome of marine alga *P. kessleri*-I. Phylogenetic analysis showed that *P. kessleri*-I shares a common ancestry with the freshwater *P. kessleri*. The chloroplast genome sequence information was used for designing appropriate species-specific vectors using codon optimization, internal UTRs and other elements essential for a higher expression of transgenes in *P. kessleri*-I. For site-specific integration, homologous flanking sequences were attached on either side of the genes cassette in a species-specific chloroplast transformation vector. The successful transgenic colonies were recovered on a spectinomycin-supplemented medium. Transgene integration in the chloroplast genome of alga was confirmed with the help of and PCR and Southern blot analysis.

## 2. MATERIALS AND METHODS

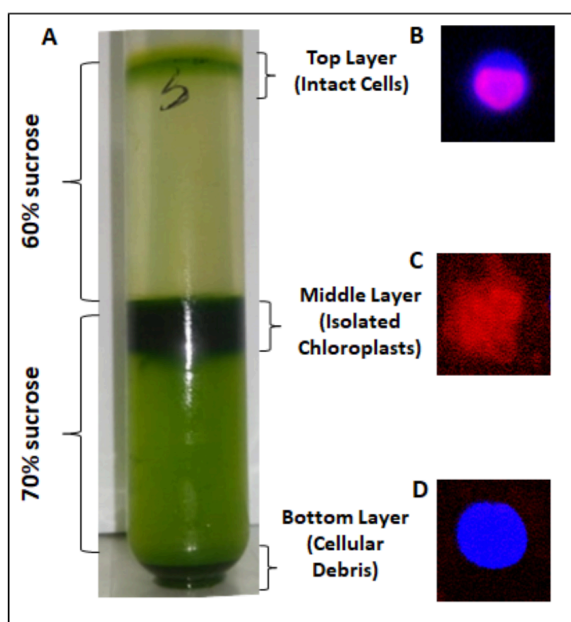
### 2.1. Strain and Growth Conditions

The algal strain, *P. kessleri*-I, was isolated from the Indian Ocean by the Indian Institute of Technology Madras and kindly provided by Mr. Shrikumar Suryanaran. Furthermore, it was characterized in our laboratory as reported [23] as per the guidelines of the National Biodiversity Authority (NBA) of India. All the isolation and the axenic culture of marine alga were maintained on agar plates containing f/2 medium (pH 8.0) [24] under 2000 Lux light intensity for 16:8 light: dark conditions and temperature of 25°C. Optical density at 750 nm was used to assess the growth of algal cells [23].

### 2.2. DNA Extraction: Qualitative and Quantitative Analysis of cpDNA

*P. kessleri*-I strain was scaled up using liquid f/2 medium, centrifuged at 3600xg and the pellet was re-suspended in 3ml ice cold breaking buffer containing 5mM dithiothreitol (DTT) and 5mM sodium ascorbate. Algal cells were sonicated on ice using cell disruptor (Sonics, 3 mm Tapered microtip probe) at a frequency of 1 kHz, 5 sec/5 sec off pulse at 30% amplitude for 2 min. The algal cells were collected by centrifugation and the pellet was re-suspended in 1 ml buffer containing 10 mM HEPES, 600 mM sorbitol, 50 mM MgCl<sub>2</sub> and 0.1% bovine serum albumin (BSA), pH 7.5-7.8 at 4°C. The auto-fluorescence of intact chloroplasts was monitored using Fluorescence Brightener 28 (Sigma Aldrich). The fluorescence from stained cellular components was observed at 650 nm and auto-fluorescence from intact chloroplast was seen at 420 nm. The presence or absence of fluorescence was observed using a NIKON fluorescence microscope. After confirmation of chloroplast release, the enriched chloroplast

was layered on a 60-70% sucrose gradient. The gradients were allowed to equilibrate overnight at 4°C. Gradient layered with cell lysate was centrifuged at 197120xg for 60 min at 10°C using the Ultracentrifuge (Beckman Coulter Optima XE-100). The chloroplast fraction was observed as a single dark green band almost in the middle of the tube [25] (Fig. 1A). The presence of intact cells, released chloroplasts and the cellular debris in the top, middle and bottom layers, respectively, was confirmed by staining with Fluorescence Brightener 28 (Fig. 1B-D). The plastomic DNA was extracted using the CTAB (Cetyl Trimethyl Ammonium Bromide) buffer method [26]. The quality of genomic DNA was checked on 1% agarose gel for a single intact band. 1 µl of the sample was used for determining the DNA concentration by NanoDrop™ 2000 Spectrophotometer (Thermo Scientific™). The purified cpDNA was supplied to Xcelris Genomics (Ahmedabad, India) for sequencing the chloroplast genome of *P. kessleri-I*.



**Fig. (1).** (A-D) *P. kessleri-I* algal cells stained with Fluorescence Brightener 28 after ultracentrifugation in sucrose gradient. (A) After sonication, *P. kessleri-I* cells were layered on 60–70% sucrose gradient and subjected to separation by ultracentrifugation. Three layers obtained were stained with Fluorescence Brightener 28 (Sigma Aldrich) to confirm cell rupture to release chloroplasts. (B) The top layer consisted of enriched intact, non-ruptured cells. The cell wall appears bright blue from the stain under the UV light. The bean-shaped chloroplast emitted red auto-fluorescence. (C) The middle layer shows intact chloroplasts showing the isolated chloroplasts as red-orange points. (D) The bottom layer contains cellular debris, dead and fractured cells that sediment to the bottom of the tube, indicated by blue fluorescence. (A higher resolution / colour version of this figure is available in the electronic copy of the article).

### 2.3. Genome Sequencing and Data Pre-processing

The chloroplast genome sequencing of *P. kessleri-I* was carried out with the help of Xcelris Genomics (Ahmedabad, India). The paired-end sequencing libraries were prepared

using the Truseq Nano DNA Library prep kit and sequenced on Illumina NexSeq (2 X 150 PE). A high-fidelity amplification step was performed using HiFi PCR Master Mix to ensure maximum yield.

### 2.4. Chloroplast Genome Assembly and Annotation

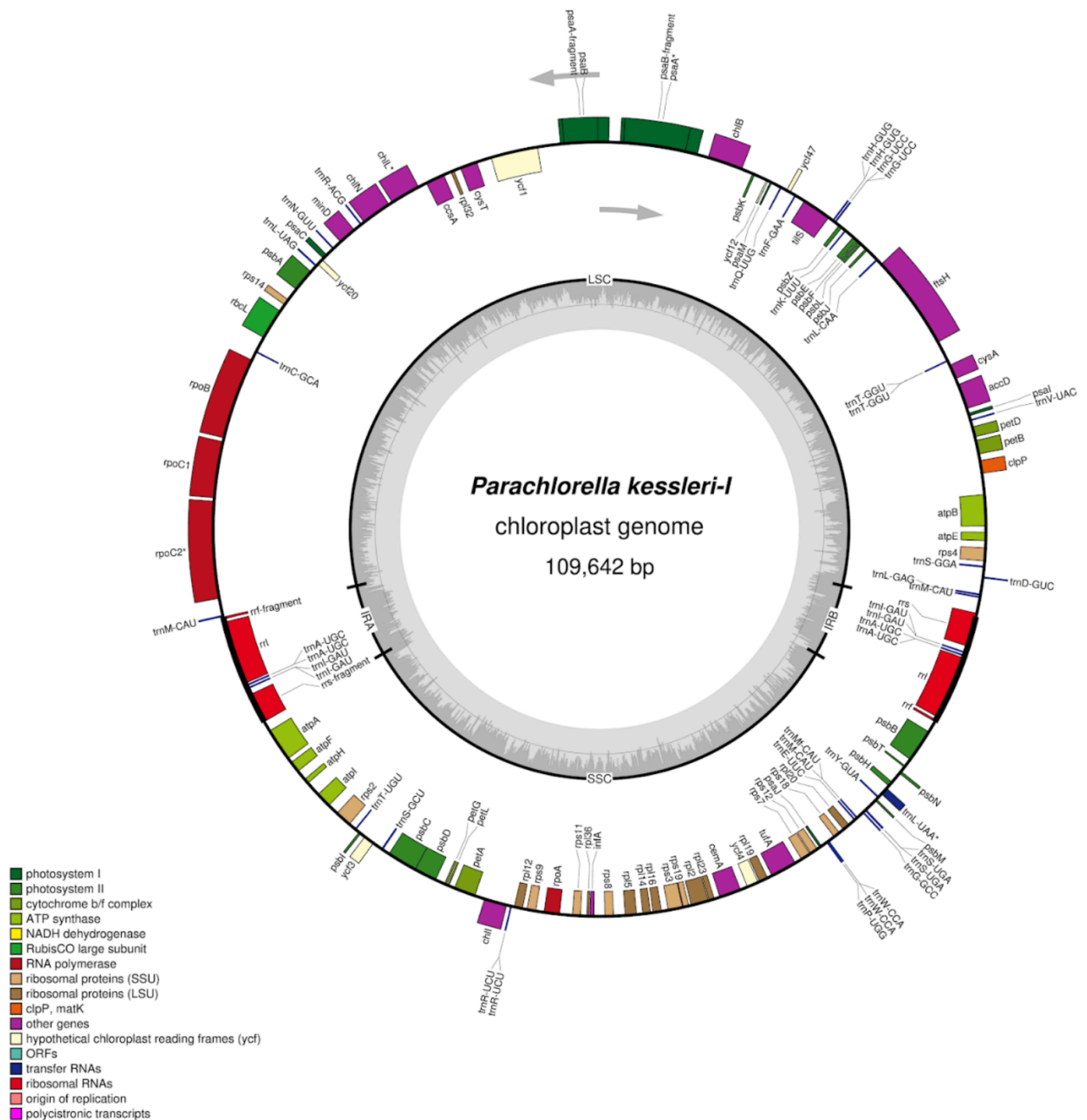
The chloroplast genome of *P. kessleri-I* was assembled using NOVOPlasty v3.0, using untrimmed reads and Ru-bisco-bis-phosphate carboxylase/oxygenase (RuBisCo) as the seed sequence [27]. The other specified parameters included a genome range of 100000-220000 bp, K-mer value of 39, reference sequence of reported freshwater *P. kessleri* chloroplast genome (NC\_012978.1) and the paired-end reads option. A web-based annotation tool, GeSeq (part of the CHLOROBOX toolbox), designed for annotation of organellar sequences [28], was used for annotating the chloroplast genome. Annotation was confirmed manually and by using NCBI BLAST [29] and DOGMA [30]. Visualization of the cp genome was done using OGDRAW [31] (Fig. 2). All transfer RNA sequences (tRNA) encoded in the cp genome were verified using ARAGORN v1.2.38 [32] with default parameters.

### 2.5. Comparative Analysis

The plastome of *P. kessleri-I* (MN013385) and *P. kessleri* (NC\_012978.1) was compared using MAUVE software: Multiple alignments of the conserved genomic sequence with rearrangements [33]. Progressive MAUVE genome alignment algorithm was used with default parameters to compare the two genomes to observe any gene loss, duplication, genome rearrangement, and horizontal transfer caused due to recombination. The whole chloroplast genome was also compared using BRIG to visualize the approximate similarity between the genomes using BLASTn (Fig. 3A and B). Hypothetical genes were studied using NCBI Conserved Domain Search against CDD v3.18 – 55570 PSSMs with an expected value threshold of 0.010000 [34]. Subcellular localization of the proteins was predicted using a deep learning algorithm DeepLoc [35].

### 2.6. Study of Antibiotic Sensitivity

Prior to chloroplast transformation, algal cells were adapted to grow on the TAP medium [36] without the addition of NaCl (salt), so that sensitivity to antibiotics can be accurately determined. In the presence of salt, the algal cells may be tolerant to antibiotics [37]. In order to identify an optimal concentration of the antibiotics zeocin (Invitrogen) and spectinomycin (Sigma Aldrich), different concentrations of zeocin (10, 20, 40, 60, 80 and 100 mg.L<sup>-1</sup>) and spectinomycin (100, 200, 300, 400, 500 and 600 mg.L<sup>-1</sup>) were tested using TAP medium solidified with 1.2% agar. Approximately  $1.5 \times 10^6$  cells of *P. kessleri-I* were plated onto 90 mm Petri plates and incubated with 16:8 h light (irradiance of 2000 lux) and dark conditions at 25±1°C. The final concentrations of 20 mg.L<sup>-1</sup> zeocin and 400 mg.L<sup>-1</sup> spectinomycin were observed to completely cease the growth of *P. kessleri-I* cells. Therefore, 20 mg.L<sup>-1</sup> zeocin and 400 mg.L<sup>-1</sup> spectinomycin were taken into account as optimized antibiotic concentrations for exploring the chloroplast transformation in *P. kessleri-I*.

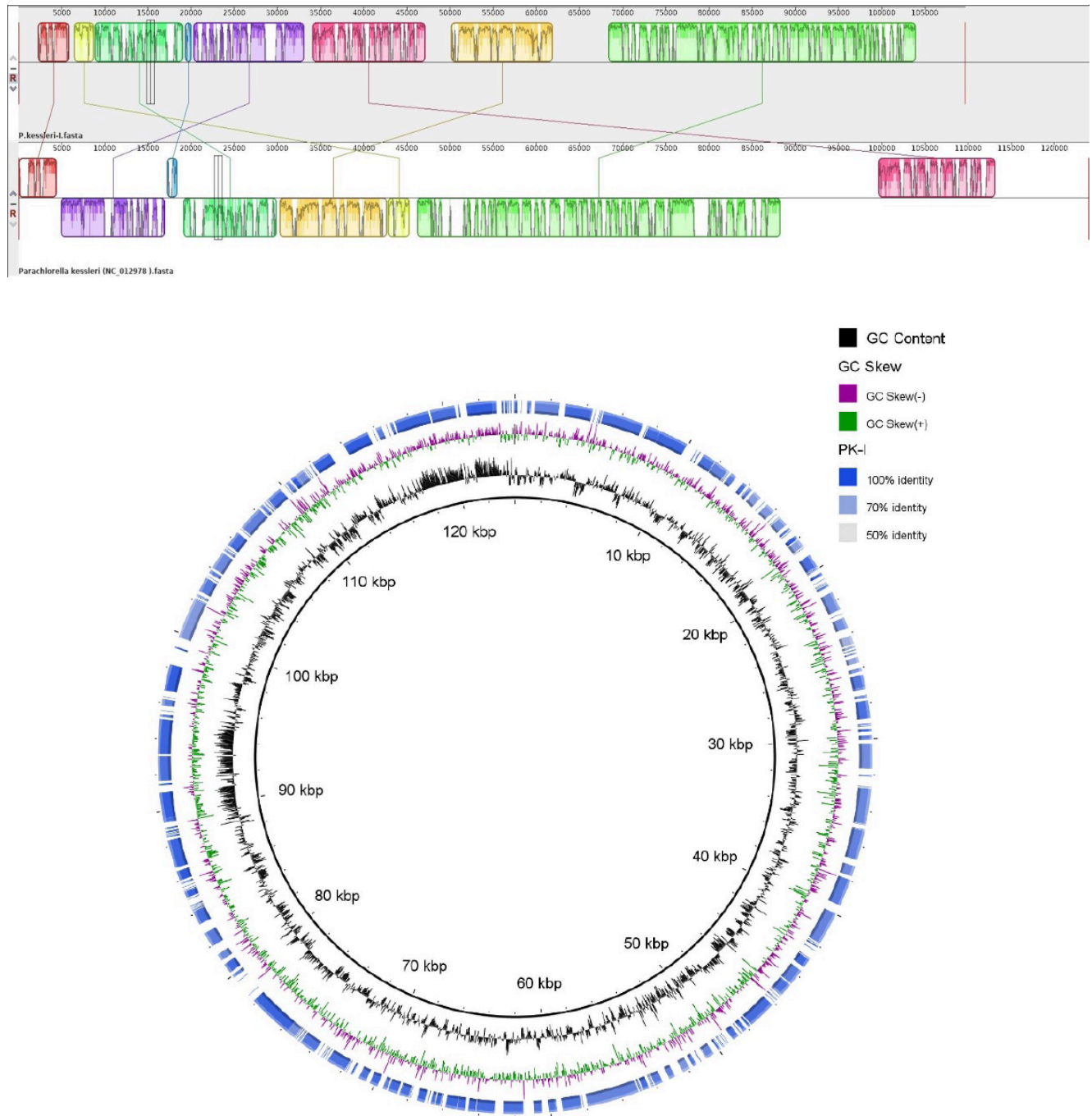


**Fig. (2).** Chloroplast genome map of *Parachlorella kessleri-I*. The predicted genes are shown and colors represent genes code, which are given at the bottom left. Genes shown on the outside of the circle are transcribed clockwise and on the inside of the circle are transcribed anti-clockwise. (A higher resolution / colour version of this figure is available in the electronic copy of the article).

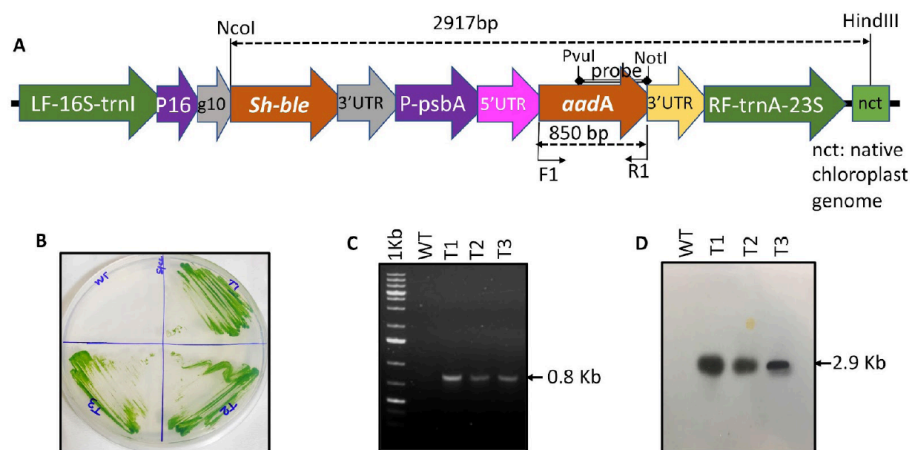
**2.7. Chloroplast Transformation Vector**

Particular regions from the chloroplast genome sequence, namely 16S rDNA, *trnI*, *trnA* and 23S rDNA were extracted and used for the construction of vector *in-silico*. Both left and right flanking regions which contained the plastome fragments of approximately 0.8 Kb each, consisting of 16S rDNA -*trnI* and *trnA* - 23S rDNA, respectively, were used for homologous recombination. Between the right and left

flanks, transgenes cassette contained 16S Prm promoter-g10-*ShBle* gene-3'UTR-Trps. Promoter *psbA*-5'UTR *psbA-aadA* gene-3'UTR *psbA* was cloned into the pUC19 plasmid, as shown in Fig. (4A). The whole DNA cassette of the PkCpV vector prepared *in-silico* and DNA vector was synthesized with the help of the vendor (BioBasic Inc., Canada). The PkCpV chloroplast transformation vector was transformed into *E. Coli* (Top-10) competent cells to maintain and bulk up the plasmid.



**Fig. (3).** **A.** Synteny analysis between the chloroplast genomes of marine *Parachlorella kessleri-I* and freshwater *Parachlorella kessleri*, using MAUVE alignments. Colored and outlined blocks surround regions of the chloroplast genome sequence of marine *P. kessleri-I* that are aligned to a corresponding part of the freshwater *P. kessleri* genome, and lines connect blocks of putative homology. Within the blocks, the colored bars indicate a high level of sequence similarities. The blocks that lie above the centerline are aligned in the forward orientation relative to the first genome sequence, i.e., *P. kessleri-I*. Blocks below the centerline represent the region that aligns in reverse complement orientation. White areas indicate low identity regions between strains. Regions with the same color indicate high-similarity syntenic blocks and are connected by the same color bars. The mauve colored region of the similarity plot represents the conserved part in both genomes. The average level of conservation in a particular region is shown by the height of the similarity profile. **B.** Comparative circular genome (BLAST) visualization of *Parachlorella kessleri-I* and *Parachlorella kessleri*. From inside to outside Ring 1: GC content, Ring 2: GC Skew, Ring 3: BLAST comparison of *P. kessleri-I* with *P. kessleri* on a sliding scale according to percentage identity (100%, 75%, or 50%) along with GC content and skew. Image generated by BRIG. The outermost ring comprises major segments with 100% and 70% sequence similarity to freshwater strain; however, regions of about 50% sequence similarity are also observed. Regions with high GC content are also clearly observed from the innermost ring depicting the GC content of the chloroplast genome. (A higher resolution / colour version of this figure is available in the electronic copy of the article).



**Fig. (4).** A-D. Physical map of chloroplast transformation vector of *Parachlorella kessleri*-I. (A) The PkCpV cassette contained the left and right flanks (native chloroplast DNA sequences of *P. kessleri*-I) and Promotor *rrn16S* controls the expression of the *Sh ble* gene and promoter *psbA* drive *aadA* expression. (B) The PkCpV cassette cloned in pUC19 plasmid was bombarded on algal culture lawn plated over 1% agar containing TAP medium. Transgenic colonies were selected on TAP medium either containing zeocin or spectinomycin. (C) The spectinomycin resistant green colonies were further ascertained for gene integration via PCR analysis producing 0.8kb amplicon using *aadA* specific primers. (D) To confirm the integration of transgenes at the specific location of the chloroplast genome, Southern blot was performed on genomic DNA digested with *NcoI* and *HindIII* (restriction enzyme cut outside the flanking region of 23S), yielding 2.9 kb signal in transgenic cell lines named as T1, T2 and T3. No signal was detected in WT. (A higher resolution / colour version of this figure is available in the electronic copy of the article).

## 2.8. Genetic Transformation of *P. kessleri*-I and Antibiotics Selection

Prior to bombardment, an algal cell lawn was prepared, containing a high concentration of cells in the middle area of the Petri plates. The chloroplast DNA transformation vector specific to *P. kessleri*-I (PkCpV) was coated on gold particles S550d following the vendor's instruction (Seashell Technology, CA, USA). About 10  $\mu\text{g}$  of plasmid DNA (stock  $1\text{mg}\cdot\text{ml}^{-1}$ ) was added to 3 mg of a gold carrier for 5 transformation experiments. Thereafter, an equal volume of precipitation buffer (50  $\mu\text{L}$ ) was added. The total mixture (100  $\mu\text{L}$ ) was briefly vortex and incubated for 3 minutes. The mixture was spun (10,000 rpm for 10 sec) to pellet the DNA coated-gold particles. The supernatant was removed and 500  $\mu\text{L}$  of 100% chilled ethanol was added. Again, the mixture was briefly vortex and spun (10,000 rpm for 10 sec) to pellet DNA-coated gold particles. The supernatant was discarded and 50  $\mu\text{L}$  of 100% ethanol was added to DNA-coated gold particles. The pellet was homogenized in ethanol by sonication and 10  $\mu\text{L}$  each of the suspension of DNA-coated gold particles was spotted in the middle of the microcarrier (5 discs) (Bio-Rad, Hercules, CA, USA). The DNA coated-gold particles were bombarded on the algal lawn using 1350-psi rupture discs at 6 and 9 cm distance, respectively [38]. After overnight incubation of cultures in the dark, bombarded cells were divided equally on TAP medium containing 20  $\text{mg}\cdot\text{L}^{-1}$  zeocin and 400  $\text{mg}\cdot\text{L}^{-1}$  spectinomycin (Fig. 4B).

## 2.9. PCR and Southern Blot Analysis

Transgenic colonies appeared only in the spectinomycin (400  $\text{mg}\cdot\text{L}^{-1}$ ) containing the selection medium. Spectinomycin resistant colonies were screened by colony PCR using gene-specific primer pair: SP\_F1 (5'-GTCGACATGGAA CAGAAGTTGATT-3') and SP\_R1 (5'-GTGGACAAA TTCTTCCAAGTATCT-3'). Each PCR reaction was performed in 50 $\mu\text{L}$  reaction mixture containing 1X Taq Reaction

buffer (HiMedia Labs, Mumbai, India), 0.1 $\mu\text{g}$  DNA template, 200  $\mu\text{M}$  of each dNTPs, 1 Unit Taq polymerase (HiMedia Labs, Mumbai, India), and concentrations of F1 and R1 used 0.5  $\mu\text{M}$  of each oligo. A BioRad MyCycler (BioRad, USA) was employed for PCR amplification, as shown in Fig. (4C), to amplify about 850 bp amplicon. A PkCpV plasmid DNA was used as a control (25-ng) along with wild type algal cells and transformed colonies selected on 400  $\text{mg}\cdot\text{L}^{-1}$  spectinomycin. The amplified PCR products were obtained using 35 cycles at 95 $^{\circ}\text{C}$  denaturation of DNA template, 59 $^{\circ}\text{C}$  annealing and extension at 72 $^{\circ}\text{C}$  for 1 min and a final extension at 72 $^{\circ}\text{C}$  for 10 min. PCR products were analyzed on agarose gel (1.0 %) electrophoresis (BioRad, USA).

The transgenes integration was further confirmed by Southern blot analysis. Genomic DNA of transformed and untransformed cells was extracted using the DNeasy plant mini kit (Qiagen). About 6  $\mu\text{g}$  genomic DNA was digested with *NcoI* and *HindIII* (restriction enzyme located outside the flanking region of 23S), as shown in Fig. (4A). The digested genomic DNA was electrophoresed on a 0.8 % agarose gel, denatured and blotted onto Hybond positive nylon membrane. The probe was prepared as per the AlkPhos Direct Labeling Reagents kit (GE Healthcare) using the *aadA* gene around 400 bp fragment isolated from PkCpV plasmid after digestion with *NotI* and *PvuI* enzymes. Southern hybridization was carried out according to AlkPhos kit instructions (GE Healthcare, Little Chalfont, United Kingdom).

## 2.10. Determination of Copy Number (ploidy) of the Chloroplast Genome Using Droplet Digital PCR

The copy number (ploidy) of the chloroplast genome was determined by using Droplet Digital PCR where actin (ACT) was used as a reference gene and a segment in the IR region of *P. kessleri*-I chloroplast was used as a 'target' gene (IR). Similarly, the spectinomycin resistance *aadA* gene (SP) was

used as a target gene to decipher a copy number in the transgenic *P. kessleri-I* (Supplementary Table 1). A total reaction of 20 microliters was prepared (according to manufacturer's instruction) containing ddPCR EvaGreen Supermix (BioRad), primers and template DNA (predigested with NcoI & HindIII) and pipetted into the compartments of the Droplet Generator DG8™ Cartridge (BioRad), and 70 µL of the Droplet Generation Oil for EvaGreen Supermix (BioRad) was added to the appropriate wells. The cartridges were covered with DG8™ Gaskets (BioRad) and placed in a QX100™ Droplet Generator (BioRad) to generate the droplets. Afterward, the droplets were gently transferred to a semi-skirted 96-well PCR plate (BioRad) using a Pipet-lite™ XLS+ manual 8-channel pipette with the range 5–50 µL (Rainin). The PCR plate was sealed with pierceable foil (BioRad) using a PX1™ PCR Plate Sealer (BioRad). After sealing, the PCR plate was placed in a C1000 Touch™ Thermal Cycler (BioRad) for PCR amplification. PCR cycling conditions were recommended by the manufacturer: 5 min initial denaturation at 95°C followed by 40 cycles of 30 s at 95°C and extension for 1 min at a temperature gradient of 51–60°C with a ramp speed of 2°C /sec; one cycle of stabilization of 5 min at 4°C and 5 min at 90°C. The PCR reaction was held at 12°C. Individual droplets were analysed with the QX 100 Droplet reader (BioRad) and the acquired data analysis was performed using the QuantaSoft™ Analysis Pro software package (BioRad version 1.0.596). Samples with more than 10,000 total droplets and good separation between positive and negative droplets were selected for analysis. The absolute amount of gene for individual samples was calculated by the software under default settings and the actual copy number of each target gene was estimated using the following formula [39]:

$$\text{Copy of Gene (Target)} = \frac{\text{Absolute amount of Gene (Target)}}{\text{Absolute amount of Gene (Reference)}} \times \text{Copy of Gene (Reference)}$$

### 3. RESULTS

#### 3.1. Chloroplast Genome Organization and Features

The whole genomic DNA of *P. kessleri-I* was sequenced using the Illumina truseq paired-end sequencing. A total of 17498340 paired-end reads were generated with a read length of 150 bp. NOVOPlasty v3.0 [27] was used to assemble the circular chloroplast genome of *P. kessleri-I*. The chloroplast DNA sequence of *P. kessleri-I* was assembled as a quadripartite structure of 109,642 bp (Fig. 2) with LSC (41,443 bp), SSC (35,669 bp) and a pair of IRs (10,216 bp). The organelle genome percent obtained was 6.09% and average organelle coverage was 1539 using NOVOPlasty V3.0 [27]. The GC and AT content of the whole plastome was 29.5% and 70.5%, respectively.

In the chloroplast genome of *P. kessleri-I*, a total of 116 genes were identified, including 22 photosynthetic genes (6 for PSI and 16 for PS II), 5 subunits of cytochrome complex, 6 subunits of ATP synthase, 4 ATP binding genes, 3 chlorophyll biosynthesis genes and one large subunit of RuBisCo. Protein synthesis and DNA replication genes include 32 tRNAs, 4 rRNAs, 11 small ribosomal unit genes, 10 large ribosomal unit genes and 4 RNA polymerase subunit genes. The genome also has 4 hypothetical genes, *ycf3*, *ycf4*, *ycf47*

and *ycf20* (Table 1). *Ycf3* is the only member of the *ycf3* superfamily. It is essential for the assembly of the photosystem I (PSI) complex. *Ycf4* superfamily consists of hypothetical *Ycf4* proteins from various chloroplast genomes. It has been suggested that *Ycf4* is involved in the assembly and/or stability of the photosystem I complex in chloroplasts. These two genes have been characterized by the green alga *Chlamydomonas reinhardtii* [40]. *Ycf47* and *Ycf20* have domains of unknown function based on conserved domain search [34]. Subcellular localization was predicted by DeepLoc v1.0 [35], which suggests that *Ycf3* and *Ycf4* are soluble proteins while *Ycf47* and *Ycf20* are membrane-bound proteins.

#### 3.2. Comparative Analysis

The chloroplast genome alignment generated in MAUVE showed eleven local collinear blocks (LCB), revealing the synteny between the two genomes when laid horizontally. The overall architecture of marine and freshwater *P. kessleri* chloroplast genomes suggests that there have been rearrangements of genes even though the synteny is conserved (Fig. 3). Comparative analyses showed that both the chloroplast genome of marine and freshwater algae have a significant homology, as shown in Fig. (3A). Conserved gene clusters in the LSC region have close similarity in the genes' arrangement in both the marine and freshwater *Parachlorella* species [41]. The chloroplast genome of *P. kessleri-I* was compared with *P. kessleri* by using BLAST and the comparative image was generated by the BRIG software (BLAST Ring Image Generator, v0.95), [42] which showed a high similarity between the sequences of two strains denoted by concentric colored rings (Fig. 3B).

#### 3.3. Chloroplast Transformation and Selection of Transgenic Cultures

The growth of *P. kessleri-I* was optimized using the TAP medium (solidified with agar) and varying concentrations of zeocin and spectinomycin were used to determine the antibiotic sensitivity for the culture. The growth of wildtype (WT) algal cells was ceased at 20 mg. L<sup>-1</sup> zeocin and 400 mg. L<sup>-1</sup> spectinomycin. These were used as optimum antibiotics concentrations for the selection of the transgenic algae colonies after introducing the PkCpV cassette *via* bombardment. The chloroplast transformation vector containing PkCpV cassette with *Sh-Ble* gene (*Sh ble* protein: QBQ65853) that confers resistance to bleomycin, phleomycin, and zeocin was placed under the *16S-rrn* PEP promoter (AC# KY426960.1) along with 5'UTR-g10 translator of *E. coli* (AC# AF176637). The *aadA* gene (AC# AY442171) that confers resistance to spectinomycin was placed under the light-regulated promoter *psbA* along with 5'UTR-*psbA* translator and 3'UTR-*psbA* terminator sequence (similar to AC# NC\_001879). The PkCpV cassette was designed *in-silico*, as shown in supplementary Fig. S1 and the whole DNA vector was synthesized (BioBasic Inc. Canada).

Following the reported protocol [38], *P. kessleri-I* cell lawn was prepared and bombarded with the chloroplast transformation vector PkCpV. A total of 60 plates were bombarded using 1350 psi rupture disk at a distance of 6 cm, while 50 plates were bombarded at a distance of 9 cm. The

Table 1. Categorical description of the 117 genes in the chloroplast genome of *Parachlorella kessleri*-I.

Category	Group of Genes	Name of Genes	
Photosynthesis	Photosystem I	<i>psaA, psaB, psaC, psal, psaJ, psam</i>	
	Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ, ycf12, psb30</i>	
	Cytochrome complex	<i>petA, petB, petD, petG, petL</i>	
	ATP synthase	<i>atpE, atpF, atpI, atpA, atpH, atpB, atpD</i>	
	ATP binding	<i>cysA, ftsH, cysT, minD</i>	
	Large subunit of Rubisco	<i>rbcL</i>	
	Protein synthesis and DNA replication genes	Ribosomal RNAs	<i>rrs, rrf, rrl</i>
Transfer RNAs		<i>trnA-UGC (*2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnM-CAU, trnG-UCC, trnG-GCC, trnH-GUG, trnI-CAU, trnI-GAU (*2), trnK-UUU, trnL-CAA, trnL-UAA, trnL-UAG, trnL-GAG, trnM-CAU (*3), trnN-GUU, trnP-UGG, trnQ-UUG, trnR-ACG, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-UAC, trnW-CCA, trnY-GUA, tilS</i>	
Small ribosomal unit		<i>rps7, rps11, rps18, rps4, rps19, rps3, rps2, rps8, rps14, rps12, rps9</i>	
Large ribosomal unit		<i>rpl14, rpl16, rpl32, rpl36, rpl23, rpl19, rpl12, rpl20, rpl2, rpl5</i>	
RNA polymerase subunits		<i>rpoA, rpoB, rpoC1, rpoC2</i>	
Miscellaneous group		Protease	<i>clpP</i>
		Acetyl-CoA-carboxylase subunit	<i>accD</i>
	Envelope membrane protein	<i>cemA</i>	
	Component of TIC complex	<i>ycf1</i>	
	c-type cytochrome synthesis	<i>ccsA</i>	
	Unknown	Hypothetical genes	<i>ycf3, ycf4, ycf47, ycf20</i>



bombarded algal plates were incubated in the dark for two nights. Thereafter, approximately equal numbers of cells ( $1 \times 10^6$ ) were plated on the selection medium containing zeocin and spectinomycin on the third day. A single green colony was observed on plates bombarded at 6 cm distance while two green colonies when bombarded at 9 cm after 3-4 weeks on the TAP medium containing  $400 \text{ mg. L}^{-1}$  spectinomycin. However, no transgenic colonies appeared on the TAP medium containing zeocin ( $20 \text{ mg. L}^{-1}$ ), even six weeks after the bombardment. The spectinomycin resistant colonies were further sub-cultured after every two weeks on the TAP medium supplemented with antibiotics for 4 months. Transgenes integration was analyzed by PCR and Southern blot.

### 3.4. Analysis of Transgene Integration in the Chloroplast Genome

The three spectinomycin resistant transgenic colonies (Fig. 4B), namely T1, T2, and T3, were tested for transgene integration in the chloroplast genome. The transformed cell lines of *P. kessleri-I* and wildtype (WT) cells were grown on liquid TAP medium to bulk up the biomass for total genomic DNA extraction. The presence of the *aadA* gene integrated into the chloroplast genome was determined by PCR, as described in Materials and Methods. The expected PCR amplicon of size  $\sim 850 \text{ bp}$  of the *aadA* gene was observed, which confirmed the integration of the transgenes into *P. kessleri-I* cells (Fig. 4C). There was no PCR-amplification observed in the WT cells.

To confirm if the targeted DNA cassette was indeed integrated at the specified position in the chloroplast genome, that is, between 16S-*trnI* and *trnA-23S* locus, Southern blot analysis was performed. The genomic DNA of transgenic (T1, T2, and T3) and WT cell lines was extracted and digested by *NcoI* and *HindIII* restriction enzymes. *NcoI* site was located within the vector cassette, while *HindIII* digested the DNA outside the Right Flank (*trnA-23S*) region, as shown in Fig. (4A). The plasmid PkCpV was digested with *PvuI* and *NotI* to get the *aadA* fragment ( $\sim 0.4 \text{ kb}$ ) used as a probe. Genomic DNA digested with *NcoI* and *HindIII* was hybridized with the probe, as shown in Fig. (4A). All the transgenic samples showed a hybridization signal with probes (Fig. 4D), while no signal was observed in the WT. The transformed algal cell lines thus confirmed transgenes integration at specified sites of the chloroplast genome of *P. kessleri-I* via homologous recombination.

Transgenic algal cell lines T, T2, and T3 were maintained as frozen cultures at  $-70^\circ\text{C}$  as reported [23] and were not exposed or disposed of into the environment at the end of the experiment.

### 3.5. Determination of a Copy Number of the Chloroplast Genome in *P. kessleri-I*

The ddPCR experiment was performed to estimate the number of copies of the chloroplast genome in *P. kessleri-I* based on the method described by Sun *et al.* [39]. Assuming actin to be a single copy gene in the nuclear genome of *P. kessleri-I*, it was used as a reference for the determination of the copy number (ploidy) of the chloroplasts. A segment of the inverted repeat region (IR) was taken as a target (marker)

to determine the copy number of the chloroplast genome in *P. kessleri-I*. Similarly, spectinomycin resistant *aadA* marker gene was used to assess the homoplasmic status of the transgenic lines T1-T3. The absolute copies of the chloroplast markers were calculated using the QuantaSoft™ Analysis Pro software package (BioRad version 1.0.596) (Supplementary Fig. S2 and S3). To determine the actual copy number of the chloroplast genomes, the values deduced from ddPCR were divided by two, since the chloroplast markers (both IR and SP) fall in the inverted repeat regions, which is repeated twice in the chloroplast genome of *P. kessleri-I*. Thus, the copy number of the chloroplast genomes in *P. kessleri-I* was observed between 90 to 100. Whereas the copy number of transformed chloroplast genomes carrying the *aadA* gene was found between 74-83 in the transgenic cell lines; the highest copy number was observed in the transgenic T1 line (Supplementary Table 2).

### 3.6. Comparative Growth Analysis Transgenics Microalga Lines

The total biomass and lipid content in algal cells was determined as dry cell weight (DCW) after harvesting a *P. kessleri-I* transgenic cultures selected on the TAP medium (containing spectinomycin) and WT cells (grown without antibiotic). The DCW was observed as follows: transgenic cells T1 ( $315 \pm 0.21 \text{ mg. L}^{-1}$ ), T2 ( $290 \pm 0.62 \text{ mg. L}^{-1}$ ) and T3 ( $319 \pm 0.53 \text{ mg. L}^{-1}$ ) compared to WT ( $342 \pm 0.35 \text{ mg. L}^{-1}$ ). Similarly, total lipid yield was observed in transgenic cells T1 ( $97 \pm 0.2 \text{ mg. L}^{-1}$ ), T2 ( $65 \pm 0.80 \text{ mg. L}^{-1}$ ) and T3 ( $78 \pm 0.46 \text{ mg. L}^{-1}$ ) compared to WT ( $88 \pm 0.4 \text{ mg. L}^{-1}$ ). These observations indicate that there was no adverse impact on the transgenic cultures due to the transgenes load on the genetically modified chloroplast genome (Supplementary Table 3).

## 4. DISCUSSION

Chloroplasts are members of a class of organelles known as plastids found in plant cells and eukaryotic algae that carry out photosynthesis and are the primary site for the world's food production. Other important activities that occur in plastids include oxygen evolution, carbon sequestration, starch production, synthesis of amino acids, fatty acids, and pigments, and also key aspects of sulphur and nitrogen metabolism. Chloroplast organelles have their own genomes and genetic system. Chloroplast genomes or plastomes are circular DNA molecules having approximately 100-200 genes, most of which encode photosynthetic, transcription, and translational apparatus. The chloroplast genes are usually arranged in operon-like clusters of similar polarity, resulting in large polycistronic primary transcripts that are later processed to oligo-/monocistronic mRNAs by editing and splicing. The genes are transcribed by a eubacterial-type RNA Polymerase and the mRNAs are translated on the 70S ribosome [38]. Detailed DNA sequence information of the chloroplast genome should help in establishing a toolbox for efficient engineering of this specialized subcellular compartment and producing useful biomolecules.

The whole genome of marine *P. kessleri-I* (I stands for ICGEB) strain was sequenced through the Illumina Platform. The complete *P. kessleri-I* chloroplast genome is 109,642 bp

(GenBank Accession No. MN013385) and assembled as a quadripartite structure. *P. kessleri*-I contains 116 genes (Table 1) wherein 3 were ribosomal RNA genes (5s, 16s and 23s), 32 transfer RNA genes, 21 genes for ribosomal subunits (11 for small and 10 for large) and four genes code for DNA directed RNA polymerase (*rpoA*, *rpoB*, *rpoC1*, *rpoC2*). There were 6 and 16 genes respectively for photosystem I (*psa*) and photosystem II (*psb*). There were 7 subunits for ATP synthase, one large unit for ribulose biphosphate carboxylase, 5 subunits for cytochrome complex and one subunit of Acetyl-CoA-carboxylase. The most important part relevant to the chloroplast genome was the DNA sequence present in the inverted repeat (IR) region. The chloroplast organization and gene order of *P. kessleri*-I exhibited a quadripartite structure with GC and AT content of 29.5% and 70.5%, respectively. The circular chloroplast genome of marine alga *P. kessleri*-I (Accession no. MN013385) consists of 109,642 bp while freshwater *P. kessleri* (Accession no. NC\_012978.1) has 123,994 bp with a 70% AT-rich region [41]. *P. kessleri*-I encodes 116 genes (Table 1), whereas the freshwater *P. kessleri* strain contains 120 genes. However, the size of the *P. kessleri*-I chloroplast genome was found to be smaller than that of freshwater strain. Genome progressive alignment was carried out to ascertain if there is any genome rearrangement, gene gain, loss, and duplication. The MAUVE tool was used to align the two genomes, and it was noticed that the maximum region was conserved in both genome sequences, as visualized using BRIG (Fig. 3B). This indicates a common ancestry for both the marine and freshwater species and they might have acclimatized over time into the diverse habitat. The decoded DNA sequence of the IR, *i.e.*, 16S-trnI-trnA-23S part of the chloroplast genome of *P. kessleri*-I, was crucial for generating the transgenics. This part was selected as a flanking region to integrate the transgenes cassette *via* homologous recombination. Using the 16S-trnI-trnA-23S flanking regions, successful chloroplast transformation was optimized for the oleaginous marine alga *P. kessleri*-I.

Chloroplasts are ideal hosts for the expression of transgenes [10]. Genetic engineering of the chloroplast genome provides a promising platform for the production of high-value targets because it allows manipulation of metabolic processes in ways that would be prohibitively difficult through traditional approaches [13]. Genes in the chloroplast genome are organized in the form of operons often transcribed from the same promoter. Chloroplasts have the necessary machinery for cutting and processing the polycistronic transcripts into individual readable transcripts, which provides the opportunity to express multiple genes, enabling the engineering metabolic pathways where the expression of multiple genes is required. A major advantage of chloroplast transformation is the ability to accumulate large amounts of foreign protein. The reported yields of recombinant proteins in the algal chloroplasts are generally in the range of 0.1 to 5% total soluble protein (TSP) of the cell [43]. Also, protein folding and disulphide bond formation occur readily in the chloroplasts making this system an ideal platform for the production of proteins with multiple domains or subunits. Moreover, as a cellular compartment, chloroplasts can act as a safe subcellular compartment for the accumulation of foreign protein(s) without disturbing the biology of the rest of

the cell. Another major advantage that chloroplast engineering offers is the precise insertion of transgenes into specific sites in the chloroplast genome *via* homologous recombination and is not random, unlike the nuclear transgene integration. Homologous recombination is very common in the plastid genomes and is mediated by the RecA protein. *RecA* also helps in maintaining chloroplast DNA integrity by participating in homologous recombination DNA repair [44]. Complete homology of the plastid DNA flanking sequences ensures highly efficient chloroplast transformation. Overall, transplastomic technology represents an important and feasible approach for the production of high-value targets on a large scale. Several recombinant proteins, therapeutics, commercial enzymes, and metabolites have been produced in microalgae by transforming the chloroplast genome [13].

Chloroplasts have been originated from a series of endosymbiotic events leading to marked homology with the current cyanobacterial genome and the gene expression also reflects the prokaryotic nature of the chloroplast. During this transition, a drastic shrinkage of the cyanobacterial genome occurred where the genes disappeared or got translocated to the nuclear genome [45]. The chloroplast transcripts are mostly polycistronic (similar to prokaryotic) with promoters and untranslated regions at both 5' and 3' ends. The mechanism of gene expression in chloroplasts is more complex than the prokaryotic system and is regulated by the plastid-encoded plastid-RNA polymerase (PEP), which shares similarity with the bacterial RNA polymerases. However, the chloroplast retained the genes for RNA polymerase core subunit (*rpoA*, *rpoB*, *rpoC1*, and *rpoC2*), genes for the regulatory sigma factor were transferred to the nuclear genome [46]. For example, in *C. reinhardtii*, the PEP core complex interacts with a single sigma factor encoded by the nuclear genome- *RPOD* [47]. Other studies also show the presence of interorganellar interaction between nuclear genes controlling the expression of chloroplast localized proteins [48]. Various nuclear factors have been shown to protect chloroplast transcripts from nucleolytic degradation in *C. reinhardtii* [49]. Transgene expression driven by *psaA*-exon1 promoter was found to be further increased by a nuclear mutation, which involves factors required for *psaA* RNA splicing [50], thus regulating the expression of chloroplast genes. Some reports also indicate that the expression of certain nuclear genes encoding chloroplast-localized proteins is dependent on the functional state of the plastid, *via* a process known as retrograde signaling [48]. For example, two chloroplast essential genes (*rps12* and *rpoA*) have shown to affect the expression of the nucleus-encoded proteins, suggesting retrograde signaling between plastid and nucleus in *C. reinhardtii*, similar to land plants [51]. Therefore, although chloroplasts are semi-autonomous organelles, their transcription is mediated by a hybrid of both prokaryotic and eukaryotic systems.

The translational machinery of this specialized subcellular compartment also reflects the evolutionary origin of plastids; the chloroplastic ribosomes are more prokaryotic in origin than the eukaryotic ribosomes. The translational control of chloroplast gene's expression is regulated by light, redox poise and transcription factors with repeats of amino acid sequence motifs [52]. Translational regulation among microalgae is mostly studied in *C. reinhardtii* because of its

amenability to diverse experimental approaches. Translation regulation in the mature chloroplast of *Chlamydomonas* varies during the light and dark phases. Several photosynthetic proteins synthesis was documented in the light phase only and not in the dark conditions. Proteins of PSII reaction centre such as D1, encoded by *psbA*; D2, encoded by *psbD*; CP43, encoded by *psbC*; CP47, encoded by *psbB*; and of the PSI reaction center such as PsaA, encoded by *psaA*; and PsaB, encoded by *psaB*; the LSU of Rubisco, encoded by *rbcL*, were synthesized during the light phase. Whereas the translation of *tufA* mRNA encoding translation elongation factor EF-Tu was recorded higher in the dark than light [52]. Intracellular localization of the mRNAs also regulates the activation or repression of their translation [53]. For example, the activation of translation of PSII subunits and the LSU of Rubisco by light correlates with the localization of its mRNAs (encoding at least two PSII subunits) and the LSU to a specific region of the *Chlamydomonas* chloroplast around the pyrenoid called the translation (T) zone [54]. Also, translational regulation of the biogenesis of photosynthetic complexes occurs by the “Control by Epistasis of synthesis” (CES). In *Chlamydomonas*, CES involves repression of translation in the biogenesis of PSI, PSII, and the cytochrome b6/f complex. It prevents the synthesis of additional unassembled copies by repressing translation of the mRNA encoding it; when a subunit of one of these complexes is present in excess [52].

The PkCpV cassette was flanked by an endogenous 16S rDNA-*trnI* sequence on the left and *trnA*-23s rDNA sequence on the right (Fig. 4A). Integration of transgenes in the intergenic region between *trnI-GAU* and *trnA-UGC* genes has resulted in high levels of expression in microalgae as well as in plants [55], hence this region was chosen for site-specific integration of transgenes attributing to its high transcriptional activity. Using the NcoI and HindIII restriction enzymes for digestion of transgenic genome, we have ensured that transgenes *Sh-Ble* and *aadA* have been precisely integrated between the *trnI* and *trnA* region of the native chloroplast. The HindIII restriction enzyme cuts outside the flanking region of the vector (into the native chloroplast genome), which confirms that transgenes are integrated at the specified site. Both the promoters used in the study, namely 16Srrn and *psbA*, are known to drive successful expression of the transgene in both algae and plants. The g-10 from *E. coli* is a strong RBS and greatly enhances the translation of the foreign gene. The enhancing effect relies on the regulation of mRNA binding and release of the ribosome S30 subunit [56], especially to 16S rRNA. The addition of the g-10 sequence was thought to improve the 5'UTR function under the 16S-rrn promoter. The *psbA* promoter is light-driven and has shown maximum protein expression in the algae chloroplast [57], hence it was preferred.

Accurate estimation of the copy number of specific genes is important for several applications. Digital PCR has provided a platform for sequence-specific detection and quantification of nucleic acids. The measurement accuracy of the Digital PCR relies on partitioning the conventional PCR reaction into a large number of individual reaction volumes. The distribution of the target gene sequences in these partitions follows a binomial function such that parts of the reac-

tion volume do not contain a copy of the target gene. The partitions amplifying copies of the target gene are labelled positive and counted, along with those without the target gene (labelled negative), after an endpoint PCR reaction. The proportions of positive partitions are used to calculate the copy number concentration according to the statistics for binomial distribution [58]. In our study, we have used the Droplet Digital (dd) PCR from BioRad to determine the chloroplast genome copy number using actin (ACT) as a reference and inverted repeat (IR) and *aadA* (SP) as target genes. Conventional actin is a single copy gene in *C. reinhardtii* [59]. Primers based on this gene sequence were used to amplify actin in *P. kessleri*-I that was considered as a single copy gene. The actual number of chloroplast genomes in *P. kessleri*-I was calculated to be between 90-100, which is slightly higher in range than that reported in *C. reinhardtii* (5-80) by Staub and Maliga [60]. This may be because of the variance in algal species. Furthermore, we also checked the state of homoplasmy attained in the transformed chloroplasts of transgenic lines with the help of ddPCR with ACT/SP as the reference/target gene. It was observed that an average of 13% of chloroplasts might not contain the transgene *aadA* (SP). Such a minor difference could not be detected by Southern blot, which is the conventional method to assess the homoplastic state in chloroplast transformation events. However, the ddPCR being a more accurate and sensitive technique could reveal the state of ploidy between the transformed *versus* untransformed chloroplasts.

Spectinomycin was observed to give a better selection of transgenic algae after transformation than zeocin. This may be because of the heightened sensitivity of *P. kessleri*-I to the antibiotic zeocin in the absence of any salts; as increasing NaCl beyond 0.2M decreases the antibiotic sensitivity in microalgae [37]. A similar observation is reported by Muñoz *et al.*, [61], where zeocin was tested for sensitivity against *Acutodesmus obliquus* and *Neochloris oleoabundans* but was not used for the selection of transformed microalgae.

In the comparative growth analysis, WT and transgenic cell cultures have produced around the same amount of biomass and lipid. This indicates that there was no detrimental impact of transgenes integrated into the chloroplast genome on the growth physiology of *P. kessleri*-I. Chloroplast transformation of other microalgae has resulted in a variety of bioproducts such as vaccines, monoclonal antibodies, biocatalysts, *etc.* *C. reinhardtii* chloroplast has been extensively exploited for the production of subunit vaccines, monoclonal antibodies, immunotoxins and cancer cell therapeutics [43]. The chloroplast of other marine microalgae, *Dunelliella tertiolecta*, was transformed for the production of enzymes such as xylanase,  $\alpha$ -galactosidase, phytase, phosphatase, and  $\beta$ -mannanase [61].

## CONCLUSION

Marine *P. kessleri*-I (ICGEB strain) is an oleaginous strain and has great potential for producing biofuel molecules. We studied the whole chloroplast genome of isolated microalga from the Indian Ocean. It was sequenced through Illumina Platform and assembled using NOVOPlasty v3.0 and annotated using GeSeq MPI chlorobox. Manual annotation was done to confirm the repeat regions using NCBI

BLAST [39]. The complete chloroplast genome of *P. kessleri-I* is 109,642 bp in length, with 116 genes. The detailed study of the *trnI*-GAU and *trnA*-UGC region of *P. kessleri-I* was used for the chloroplast transformation. These unique sites have allowed integration and expression of highly efficient transgenes, precisely occurred on 16s and 23s ribosomal DNA sites. The first report of site-specific optimization of the chloroplast transformation in a new marine strain should help further to improve its biofuel properties *via* synthetic biology approach.

#### LIST OF ABBREVIATIONS

ATP	=	Adenosine triphosphate
BLAST	=	Basic Local Alignment Search Tool
CCM	=	Carbon Concentration Mechanism
CTAB	=	Cetyl Trimethyl Ammonium Bromide
DCW	=	Dry Cell Weight
DNA	=	Deoxyribonucleic acid
DOGMA	=	Dual Organellar GenoMe Annotator
DTT	=	Dithiothreitol
IRA & IRB	=	Inverted Repeat Region A & B
LSC	=	Large Single Copy
NCBI	=	National Center for Biotechnology Information
OGDRAW	=	Organellar Genome DRAW
PCR	=	Polymerase Chain Reaction
PSI & PSII	=	Photosystem I & II
RuBisCo	=	Rubisco Bis-phosphate Carboxylase/Oxygenase
SSC	=	Small Single Copy
TAP	=	Tris Acetate Phosphate Medium
TSP	=	Total Soluble Protein
UTRs	=	Untranslated Regions
WT	=	Wild Type

#### AUTHORS' CONTRIBUTIONS

PN carried out the transformation optimization experiments and carried out the molecular analysis on transgenic algae. SC and SS carried out the assembly and annotation of the chloroplast genome. MJ & Sachin Kajla assisted in interpreting the results, reviewed and gave suggestions on the study, and helped in preparing the manuscript draft. SK conceptualized the study and supervised the whole experimental research, data and analysis of results. PN and SC interpreted the results and wrote the final manuscript along with SK. All the authors have read the final manuscript and approved for communication and publication.

#### ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

#### HUMAN AND ANIMAL RIGHTS

No Animals/Humans were used for studies that are the basis of this research.

#### CONSENT FOR PUBLICATION

Not applicable.

#### AVAILABILITY OF DATA AND MATERIALS

All data generated or analyzed during this study are included in this article, along with additional files.

#### FUNDING

This work was supported by the Department of Biotechnology, Government of India, and Grant No. BT/PB/Centre/03/ICGEB/2011-II Phase to SK.

#### CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

#### ACKNOWLEDGEMENTS

We would like to acknowledge the Department of Biotechnology, Government of India and International Centre for Genetic Engineering and Biotechnology, New Delhi, for the financial and technical support for accomplishing this work.

#### SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's website along with the published article.

#### REFERENCES

- [1] Lewis, L.A.; McCourt, R.M. Green algae and the origin of land plants. *Am. J. Bot.*, **2004**, *91*(10), 1535-1556. <http://dx.doi.org/10.3732/ajb.91.10.1535> PMID: 21652308
- [2] Friedl, T. Inferring taxonomic positions and testing genus level assignments in coccoid green lichen algae: a phylogenetic analysis of 18S ribosomal RNA sequences from *Dictyochochloropsis reticulata* and from members of the genus *Myrmecia* (*Chlorophyta*, *Trebouxiophyceae* cl. nov.). *J. Phycol.*, **1995**, *31*(4), 632-639. <http://dx.doi.org/10.1111/j.1529-8817.1995.tb02559.x>
- [3] Friedl, T. The evolution of the green algae. *Origins of Algae and their Plastids*; Springer: Vienna, **1997**, pp. 87-101. [http://dx.doi.org/10.1007/978-3-7091-6542-3\\_4](http://dx.doi.org/10.1007/978-3-7091-6542-3_4)
- [4] Pombert, J.F.; Otis, C.; Lemieux, C.; Turmel, M. The complete mitochondrial DNA sequence of the green alga *Pseudoclonium akinetum* (Ulvophyceae) highlights distinctive evolutionary trends in the chlorophyta and suggests a sister-group relationship between the Ulvophyceae and Chlorophyceae. *Mol. Biol. Evol.*, **2004**, *21*(5), 922-935. <http://dx.doi.org/10.1093/molbev/msh099> PMID: 15014170
- [5] Pombert, J.F.; Otis, C.; Lemieux, C.; Turmel, M. The chloroplast genome sequence of the green alga *Pseudoclonium akinetum* (Ulvophyceae) reveals unusual structural features and new insights into the branching order of chlorophyte lineages. *Mol. Biol. Evol.*, **2005**, *22*(9), 1903-1918. <http://dx.doi.org/10.1093/molbev/msi182> PMID: 15930151
- [6] Pombert, J.F.; Lemieux, C.; Turmel, M. The complete chloroplast DNA sequence of the green alga *Oltmannsiellopsis viridis* reveals a distinctive quadripartite architecture in the chloroplast genome of early diverging ulvophytes. *BMC Biol.*, **2006**, *4*(1), 3. <http://dx.doi.org/10.1186/1741-7007-4-3> PMID: 16472375
- [7] Martin, W.; Rujan, T.; Richly, E.; Hansen, A.; Cornelsen, S.; Lins, T.; Leister, D.; Stoebe, B.; Hasegawa, M.; Penny, D. Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. USA*, **2002**, *99*(19), 12246-12251. <http://dx.doi.org/10.1073/pnas.182432999> PMID: 12218172
- [8] Graham, L.E. Green algae to land plants: an evolutionary transition. *J. Plant Res.*, **1996**, *109*(3), 241-251. <http://dx.doi.org/10.1007/BF02344471>

- [9] Henry, R.J. *Plant diversity and evolution: genotypic and phenotypic variation in higher plants*; Cabi Publishing, **2005**.  
http://dx.doi.org/10.1079/9780851999043.0000
- [10] Verma, D.; Daniell, H. Chloroplast vector systems for biotechnology applications. *Plant Physiol.*, **2007**, *145*(4), 1129-1143.  
http://dx.doi.org/10.1104/pp.107.106690 PMID: 18056863
- [11] Daniell, H.; Lin, C.S.; Yu, M.; Chang, W.J. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.*, **2016**, *17*(1), 134.  
http://dx.doi.org/10.1186/s13059-016-1004-2 PMID: 27339192
- [12] Gimpel, J.A.; Specht, E.A.; Georgianna, D.R.; Mayfield, S.P. Advances in microalgae engineering and synthetic biology applications for biofuel production. *Curr. Opin. Chem. Biol.*, **2013**, *17*(3), 489-495.  
http://dx.doi.org/10.1016/j.cbpa.2013.03.038 PMID: 23684717
- [13] Cui, Y.; Qin, S.; Jiang, P. Chloroplast transformation of *Platymonas* (*Tetraselmis*) subcordiformis with the bar gene as selectable marker. *PLoS One*, **2014**, *9*(6), e98607.  
http://dx.doi.org/10.1371/journal.pone.0098607 PMID: 24911932
- [14] Rathod, J.P.; Prakash, G.; Pandit, R.; Lali, A.M. Agrobacterium-mediated transformation of promising oil-bearing marine algae *Parachlorella kessleri*. *Photosynth. Res.*, **2013**, *118*(1-2), 141-146.  
http://dx.doi.org/10.1007/s1120-013-9930-2 PMID: 24097049
- [15] Mayfield, S.P.; Franklin, S.E. Expression of human antibodies in eukaryotic micro-algae. *Vaccine*, **2005**, *23*(15), 1828-1832.  
http://dx.doi.org/10.1016/j.vaccine.2004.11.013 PMID: 15734050
- [16] Doron, L.; Segal, N.; Shapira, M. Transgene expression in microalgae-from tools to applications. *Front. Plant Sci.*, **2016**, *7*, 505.  
http://dx.doi.org/10.3389/fpls.2016.00505 PMID: 27148328
- [17] Lee, S.B.; Kwon, H.B.; Kwon, S.J.; Park, S.C.; Jeong, M.J.; Han, S.E.; Byun, M.O.; Daniell, H. Accumulation of trehalose within transgenic chloroplasts confers drought tolerance. *Mol. Breed.*, **2003**, *11*(1), 1-3.  
http://dx.doi.org/10.1023/A:1022100404542
- [18] Kuroiwa, T.; Kawano, S.; Nishibayashi, S.; Sato, C. Epifluorescent microscopic evidence for maternal inheritance of chloroplast DNA. *Nature*, **1982**, *298*(5873), 481-483.  
http://dx.doi.org/10.1038/298481a0 PMID: 7088193
- [19] Daniell, H. Transgene containment by maternal inheritance: effective or elusive? *Proc. Natl. Acad. Sci. USA*, **2007**, *104*(17), 6879-6880.  
http://dx.doi.org/10.1073/pnas.0702219104 PMID: 17440039
- [20] Daniell, H. Molecular strategies for gene containment in transgenic crops. *Nat. Biotechnol.*, **2002**, *20*(6), 581-586.  
http://dx.doi.org/10.1038/nbt0602-581 PMID: 12042861
- [21] Kumar, S. GM algae for biofuel production: biosafety and risk assessment. *Collect Biosaf Rev.*, **2015**, *9*, 52-75.
- [22] Li, X.; Příbyl, P.; Bišová, K.; Kawano, S.; Cepák, V.; Zachleder, V.; Čížková, M.; Brányiková, I.; Vitová, M. The microalga *Parachlorella kessleri*-a novel highly efficient lipid producer. *Biotechnol. Bioeng.*, **2013**, *110*(1), 97-107.  
http://dx.doi.org/10.1002/bit.24595 PMID: 22766749
- [23] Ahmad, I.; Fatma, Z.; Yazdani, S.S.; Kumar, S. DNA barcode and lipid analysis of new marine algal species potential for biofuel. *Algal Res.*, **2013**, *2*(1), 10-15.  
http://dx.doi.org/10.1016/j.algal.2012.10.003
- [24] Guillard, R.R.; Ryther, J.H. Studies of marine planktonic diatoms. I. *Cyclotella nana* Hustedt, and *Detonula confervacea* (Cleve) Gran. *Can. J. Microbiol.*, **1962**, *8*(2), 229-239.  
http://dx.doi.org/10.1139/m62-029 PMID: 13902807
- [25] Angelova, A.; Park, S.H.; Kyndt, J.; Fitzsimmons, K.; Brown, J.K. Sonication-based isolation and enrichment of *Chlorella protothecoides* chloroplasts for Illumina genome sequencing. *J. Appl. Phycol.*, **2014**, *26*(1), 209-218.  
http://dx.doi.org/10.1007/s10811-013-0125-1
- [26] Clarke, J.D. Cetyltrimethyl ammonium bromide (CTAB) DNA miniprep for plant DNA isolation. *Cold Spring Harb. Protoc.*, **2009**, *2009*(3), pdb.prot5177.
- [27] Dierckxens, N.; Mardulyn, P.; Smits, G. NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Res.*, **2017**, *45*(4), e18.  
PMID: 28204566
- [28] Tillich, M.; Lehwark, P.; Pellizzer, T.; Ulbricht-Jones, E.S.; Fischer, A.; Bock, R.; Greiner, S. GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.*, **2017**, *45*(W1), W6-W11.  
http://dx.doi.org/10.1093/nar/gkx391 PMID: 28486635
- [29] Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.*, **1990**, *215*(3), 403-410.  
http://dx.doi.org/10.1016/S0022-2836(05)80360-2 PMID: 2231712
- [30] Wyman, S.K.; Jansen, R.K.; Boore, J.L. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*, **2004**, *20*(17), 3252-3255.  
http://dx.doi.org/10.1093/bioinformatics/bth352 PMID: 15180927
- [31] Lohse, M.; Drechsel, O.; Bock, R. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.*, **2007**, *52*(5-6), 267-274.  
http://dx.doi.org/10.1007/s00294-007-0161-y PMID: 17957369
- [32] Laslett, D.; Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.*, **2004**, *32*(1), 11-16.  
http://dx.doi.org/10.1093/nar/gkh152 PMID: 14704338
- [33] Darling, A.C.; Mau, B.; Blattner, F.R.; Perna, N.T. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.*, **2004**, *14*(7), 1394-1403.  
http://dx.doi.org/10.1101/gr.2289704 PMID: 15231754
- [34] Marchler-Bauer, A.; Bo, Y.; Han, L.; He, J.; Lanczycki, C.J.; Lu, S.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; Gwadz, M.; Hurwitz, D.I.; Lu, F.; Marchler, G.H.; Song, J.S.; Thanki, N.; Wang, Z.; Yamashita, R.A.; Zhang, D.; Zheng, C.; Geer, L.Y.; Bryant, S.H. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.*, **2017**, *45*(D1), D200-D203.  
http://dx.doi.org/10.1093/nar/gkw1129 PMID: 27899674
- [35] Almagro Armenteros, J.J.; Sønderby, C.K.; Sønderby, S.K.; Nielsen, H.; Winther, O. DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics*, **2017**, *33*(21), 3387-3395.  
http://dx.doi.org/10.1093/bioinformatics/btx431 PMID: 29036616
- [36] Gorman, D.S.; Levine, R.P. Photosynthetic electron transport chain of *Chlamydomonas reinhardtii*. V. Purification and properties of cytochrome 553 and ferredoxin. *Plant Physiol.*, **1966**, *41*(10), 1643-1647.  
http://dx.doi.org/10.1104/pp.41.10.1643 PMID: 16656452
- [37] Anila, N.; Chandrashekar, A.; Ravishankar, G.A.; Sarada, R. Establishment of *Agrobacterium tumefaciens*-mediated genetic transformation in *Dunaliella bardawil*. *Eur. J. Phycol.*, **2011**, *46*(1), 36-44.  
http://dx.doi.org/10.1080/09670262.2010.550386
- [38] Tran, M.; Mayfield, S.P. Expression of full length monoclonal antibodies (mAb) in algal chloroplast. *Antibody engineering*; Springer: Berlin, Heidelberg, **2010**, pp. 503-516.  
http://dx.doi.org/10.1007/978-3-642-01144-3\_32
- [39] Sun, Y.; Joyce, P.A. Application of droplet digital PCR to determine copy number of endogenous genes and transgenes in sugarcane. *Plant Cell Rep.*, **2017**, *36*(11), 1775-1783.  
http://dx.doi.org/10.1007/s00299-017-2193-1 PMID: 28849385
- [40] Boudreau, E.; Takahashi, Y.; Lemieux, C.; Turmel, M.; Rochaix, J.D. The chloroplast *ycf3* and *ycf4* open reading frames of *Chlamydomonas reinhardtii* are required for the accumulation of the photosystem I complex. *EMBO J.*, **1997**, *16*(20), 6095-6104.  
http://dx.doi.org/10.1093/emboj/16.20.6095 PMID: 9321389
- [41] Turmel, M.; Otis, C.; Lemieux, C. The chloroplast genomes of the green algae *Pedinomonas minor*, *Parachlorella kessleri*, and *Oocystis solitaria* reveal a shared ancestry between the *Pedinomonadales* and *Chlorellales*. *Mol. Biol. Evol.*, **2009**, *26*(10), 2317-2331.  
http://dx.doi.org/10.1093/molbev/msp138 PMID: 19578159
- [42] Alikhan, N.F.; Petty, N.K.; Ben Zakour, N.L.; Beatson, S.A. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics*, **2011**, *12*(1), 402.  
http://dx.doi.org/10.1186/1471-2164-12-402 PMID: 21824423
- [43] Green, B.R. Chloroplast genomes of photosynthetic eukaryotes. *Plant J.*, **2011**, *66*(1), 34-44.  
http://dx.doi.org/10.1111/j.1365-313X.2011.04541.x PMID: 21443621
- [44] Dyo, Y.M.; Purton, S. The algal chloroplast as a synthetic biology platform for production of therapeutic proteins. *Microbiology (Reading)*, **2018**, *164*(2), 113-121.

- http://dx.doi.org/10.1099/mic.0.000599 PMID: 29297850
- [45] Adem, M.; Beyene, D.; Feyissa, T. Recent achievements obtained by chloroplast transformation. *Plant Methods*, **2017**, *13*(1), 30. <http://dx.doi.org/10.1186/s13007-017-0179-1> PMID: 28428810
- [46] Yagi, Y.; Shiina, T. Recent advances in the study of chloroplast gene expression and its evolution. *Front. Plant Sci.*, **2014**, *5*, 61. <http://dx.doi.org/10.3389/fpls.2014.00061> PMID: 24611069
- [47] Idoine, A.D.; Boulouis, A.; Rupprecht, J.; Bock, R. The diurnal logic of the expression of the chloroplast genome in *Chlamydomonas reinhardtii*. *PLoS One*, **2014**, *9*(10), e108760. <http://dx.doi.org/10.1371/journal.pone.0108760> PMID: 25272288
- [48] Surpin, M.; Larkin, R.M.; Chory, J. Signal transduction between the chloroplast and the nucleus. *Plant Cell*, **2002**, *14*(Suppl.), S327-S338. <http://dx.doi.org/10.1105/tpc.010446> PMID: 12045286
- [49] Eberhard, S.; Drapier, D.; Wollman, F.A. Searching limiting steps in the expression of chloroplast-encoded proteins: relations between gene copy number, transcription, transcript abundance and translation rate in the chloroplast of *Chlamydomonas reinhardtii*. *Plant J.*, **2002**, *31*(2), 149-160. <http://dx.doi.org/10.1046/j.1365-3113X.2002.01340.x> PMID: 12121445
- [50] Michelet, L.; Lefebvre-Legendre, L.; Burr, S.E.; Rochaix, J.D.; Goldschmidt-Clermont, M. Enhanced chloroplast transgene expression in a nuclear mutant of *Chlamydomonas*. *Plant Biotechnol. J.*, **2011**, *9*(5), 565-574. <http://dx.doi.org/10.1111/j.1467-7652.2010.00564.x> PMID: 20809927
- [51] Ramundo, S.; Rahire, M.; Schaad, O.; Rochaix, J.D. Repression of essential chloroplast genes reveals new signaling pathways and regulatory feedback loops in *chlamydomonas*. *Plant Cell*, **2013**, *25*(1), 167-186. <http://dx.doi.org/10.1105/tpc.112.103051> PMID: 23292734
- [52] Sun, Y.; Zerges, W. Translational regulation in chloroplasts for development and homeostasis. *Biochim. Biophys. Acta*, **2015**, *1847*(9), 809-820. <http://dx.doi.org/10.1016/j.bbabi.2015.05.008> PMID: 25988717
- [53] Weis, B.L.; Schleiff, E.; Zerges, W. Protein targeting to subcellular organelles via mRNA localization. *Biochim. Biophys. Acta*, **2013**, *1833*(2), 260-273. <http://dx.doi.org/10.1016/j.bbamer.2012.04.004> PMID: 23457718
- [54] Uniacke, J.; Zerges, W. Photosystem II assembly and repair are differentially localized in *Chlamydomonas*. *Plant Cell*, **2007**, *19*(11), 3640-3654. <http://dx.doi.org/10.1105/tpc.107.054882> PMID: 18055604
- [55] Odahara, M.; Inouye, T.; Nishimura, Y.; Sekine, Y. RECA plays a dual role in the maintenance of chloroplast genome stability in *Physcomitrella patens*. *Plant J.*, **2015**, *84*(3), 516-526. <http://dx.doi.org/10.1111/tpj.13017> PMID: 26340426
- [56] Takahashi, S.; Furusawa, H.; Ueda, T.; Okahata, Y. Translation enhancer improves the ribosome liberation from translation initiation. *J. Am. Chem. Soc.*, **2013**, *135*(35), 13096-13106. <http://dx.doi.org/10.1021/ja405967h> PMID: 23927491
- [57] Rasala, B.A.; Muto, M.; Sullivan, J.; Mayfield, S.P. Improved heterologous protein expression in the chloroplast of *Chlamydomonas reinhardtii* through promoter and 5' untranslated region optimization. *Plant Biotechnol. J.*, **2011**, *9*(6), 674-683. <http://dx.doi.org/10.1111/j.1467-7652.2011.00620.x> PMID: 21535358
- [58] Deprez, L.; Corbisier, P.; Kortekaas, A.M.; Mazoua, S.; Beaz Hidalgo, R.; Trapmann, S.; Emons, H. Validation of a digital PCR method for quantification of DNA copy number concentrations by using a certified reference material. *Biomol Detect. Quantif.*, **2016**, *9*, 29-39. <http://dx.doi.org/10.1016/j.bdq.2016.08.002> PMID: 27617230
- [59] Hirono, M.; Uryu, S.; Ohara, A.; Kato-Minoura, T.; Kamiya, R. Expression of conventional and unconventional actins in *Chlamydomonas reinhardtii* upon deflagellation and sexual adhesion. *Eukaryot. Cell*, **2003**, *2*(3), 486-493. <http://dx.doi.org/10.1128/EC.2.3.486-493.2003> PMID: 12796293
- [60] Staub, J.M.; Maliga, P. Long regions of homologous DNA are incorporated into the tobacco plastid genome by transformation. *Plant Cell*, **1992**, *4*(1), 39-45. PMID: 1356049
- [61] Muñoz, C.F.; Sturme, M.H.; D'Adamo, S.; Weusthuis, R.A.; Wijffels, R.H. Stable transformation of the green algae *Acutodesmus obliquus* and *Neochloris oleoabundans* based on *E. coli* conjugation. *Algal Res.*, **2019**, *39*, 101453. <http://dx.doi.org/10.1016/j.algal.2019.101453>