Chloroplast genome analysis and genetic transformation of *Parachlorella kessleri*-I a potential marine alga for biofuel production

Prachi Nawkarkar

International Centre For Genetic Engineering and Biotechnology New Delhi

Sagrika Chugh

International Centre For Genetic Engineering and Biotechnology New Delhi

Amita Tanwar

International Centre For Genetic Engineering and Biotechnology New Delhi

Surbhi Sharma

Jawaharlal Nehru University

Charli Kaushal

Jamia Hamdard

Malik Zainul Abdin

Jamia Hamdard

Mukesh Jain

Jawaharlal Nehru University

Shashi Kumar (✉ skrhode@icgeb.res.in)

International Centre For Genetic Engineering and Biotechnology <https://orcid.org/0000-0001-7192-1526>

Research article

Keywords: Chloroplast, genome analysis, genetic transformation, *Parachlorella kessleri*-I a, biofuel

Posted Date: July 29th, 2019

DOI: <https://doi.org/10.21203/rs.2.12052/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: *Parachlorella kessleri*-I produces higher biomass and lipid content suitable for commercial production of biofuels. Sequencing complete chloroplast genome will be instrumental in the constructing species specific chloroplast transformation vectors and generating chloroplast transgenic microalga with the desired traits and greater productivity, essential for commercial sustainability of microalgae based biofuel production. **Results:** Complete chloroplast genome sequence (cpDNA) of *P. kessleri*-I was annotated. The 109,642 bp chloroplast genome 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 117 unique genes, with 70 predicted protein coding genes, 35 tRNAs, 4 rRNAs. The cpDNA provided essential information like codons, UTRs and flank sequences for homologous recombination to make a species specific chloroplast transformation vector that facilitated chloroplast transformation of *P. kessleri*-I. To optimize chloroplast transformation, two antibiotic resistance makers aminoglycoside adenine transferase (*aadA*) conferring resistance to spectinomycin and *Sh-ble* gene from bacteria that conferred resistance to zeocin were tested. Using a *aadA* gene, transgenic colonies were retrieved on TAP medium containing 400 mg/l spectinomycin. However, no transgenic colonies were recovered in the zeocin supplemented medium. The spectinomycin resistant algal cell lines were analyzed by PCR. Southern blotting confirmed the stable transgenes integration into the chloroplast genome of *P. kessleri*-I via homologous recombination. **Conclusion:** The complete chloroplast genome analysis may provide valuable resources for population and evolutionary studies of *Parachlorella* species and identifying the related species. The chloroplast genome of *P. kessleri*-I was assembled as a quadripartite structure of 109,642 bp with defined IR regions. Its complete sequencing has provided essential information like codons, UTRs and flanking sequences to generate the species specific chloroplast transformation vector and obtaining the successful site-specific chloroplast transformation in *P. kessleri*-I via homologous recombination. The optimized chloroplast transformation in marine alga *P. kessleri*-I should open a new possibilities like controlling the chain lengths of fatty acids for biofuel application, manipulating the RubisCo enzyme for improving photosynthetic process, introducing carbon concentration mechanism (CCM) for fixing higher CO₂ that may be instrumental in producing economically viable biofuel molecules.

Background

Parachlorella is a unicellular [green algae](#) belonging to the class Trebouxiophyceae in the order [Chlorellales](#). The classes Trebouxiophyceae, Prasinophyceae, Ulvophyceae and Chlorophyceae belong to the phylum Chlorophyta (1,2) among which Prasinophyceae shows the most basal divergence (3) while the Trebouxiophyceae emerged before the Ulvophyceae and Chlorophyceae as suggested from data of chloroplast and mitochondrial genome (4–6)

Chloroplasts were captured early during the evolution of a eukaryotic cell as they are considered to have originated from cyanobacteria through endosymbiosis. During the course of evolution, extensive rearrangements occurred within the chloroplast genomes or plastome. However, as compared to that of

free-living cyanobacteria the size of the chloroplast genome is subsequently reduced, but still many DNA sequences of chloroplast resemble to the cyanobacterial genome (7). It is generally believed that land plants have evolved from green algae (8) and the chloroplast genome is a highly conserved among plant species (9). Most land plant plastomes have two identical copies of inverted repeat region (IRA and IRB) separating a large single copy (LSC) region and a small single copy (SSC) region (10,11). Therefore, insightful information on repeated sequences, intergenic regions, and pseudogenes in the chloroplast DNA may be very helpful in deciphering the process of chloroplast genome evolution.

The genetic transformation of industrial microalgae may be useful for an economic viability to overexpress the desired important traits for the commercial feedstock development and sustainable biofuels production (12–16). Chloroplast transformation is routine in *Nicotiana tabacum* plant and *Chlamydomonas reinhardtii* alga and first stable chloroplast transformation was achieved in the same species, but it has been challenging in other plant species and algae. A nuclear genome based genetic transformation method was reported for *Parachlorella* sp. (17). However, the plastome genetic engineering tools are not reported yet for this spp. The chloroplast organelle in algae with several copies of genome, occupies about two thirds of the cellular space in an organism. The nature of the chloroplast genome is prokaryotic that enables site-specific recombination between homologous DNA sequences providing an unique opportunity for overexpression of transgenes and genetic modifications (18). Chloroplast transformation exhibits higher level of transgene expression, thus higher level of biofuel molecules and value added products can be produced in algae due to the presence of multiple copies of chloroplast transgenes per cell. Moreover, transgenes integration in chloroplast genome is precise and unaffected by phenomenon such as pre or post-transcriptional silencing despite transcripts accumulating at 169-fold higher levels than in nuclear transformation (20). There is no deletions and rearrangements of transgene DNA, at the site of insertion. This is advantageous over nuclear transformation that usually leads to random integration of transgenes (19). Moreover, in chloroplast genome transformation, transgene integrated into one IR region gets duplicated via *RecA* enzyme into the other IR as well (10). Chloroplast genes in green microalga (*C. reinhardtii*) are known to transfer via maternal inheritance (21). The uniparental inheritance of chloroplast traits in green algae (21, 22) can be the key to generate genetically modified species with lower environmental risks (23–25). This would be beneficial in biological containment of genetically modified strains when cultured outdoors in large volumes e.g. raceway ponds and photobioreactors. In brief, chloroplast transformation in algae may be useful to overexpress thioesterases to produce fatty acids of desired chain lengths, overexpression of carbon concentration mechanism (CCM) to fix higher CO₂ and reduce photorespiration, improvement of photosynthetic efficiency in alga by directly expressing Rubisco transgenes in the chloroplast genome.

Researchers have been studying algal fuels extensively since the 1970s, but no commercially-viable strain has yet been isolated (26). In the past, some programmes had to be shut down due to the higher cost of algal biofuels as compared to fossil fuels. The use of recombinant DNA techniques offer the greatest range of options to improve upon the performance of wild strains. Therefore, the transgenic approach has been investigated as a potential method of increasing the productivity of algae and to compete with relatively cheaper fossil fuels. *P. kessleri* accumulates higher starch and lipid content in

laboratory conditions as well as on a semi-industrial scale in outdoor photobioreactors (27) and is considered an ideal microalgal species for biofuel production. It occurs in both freshwater and marine environments.

To obtain optimal expression of transgenes in the chloroplast, identification of spacer regions and endogenous regulatory sequences for integration of transgenes is required. The homologous flanking sequences are attached on either side of the transgene cassette while designing the chloroplast transformation vectors to facilitate double recombination, thereby eliminating the concerns of position effect. A novel marine strain *P. Kessleri* was isolated from the Indian ocean, which contains high lipids on a dry weight basis (14). Thus, to gain more insights, we have undertaken complete sequencing and annotation of the chloroplast genome of the oleaginous marine alga *P. kessleri*. Phylogenetic analysis indicated that, *P. kessleri* shares common ancestry with the freshwater *P. kessleri*. The chloroplast genome sequence information will be used for designing appropriate species specific vector using codon optimization, internal UTRs and other elements essential for higher expression of transgenes in *P. kessleri*.

Results

2.1. Chloroplast genomic organization and features

The whole genomic DNA of *P. kessleri* was sequenced using Illumina techniques. A total of 17498340 of paired end reads were generated with a read length of 150 bp. NOVOPlasty v3.0 (28) was used to assemble the circular chloroplast genome of *P. kessleri*. The chloroplast DNA sequence of *P. kessleri* was assembled as a quadripartite structure of 109,642 bp (Fig. 1) 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 (28). The GC and AC content of whole plastome were 29.5% and 70.5% respectively.

In the chloroplast genome of *P. kessleri*, total 116 genes were predicted including 23 photosynthetic genes (6 for PSI and 17 for PS II), 5 cytochrome complexes, 6 ATP synthases, 4 atp binding and one large subunit of Rubisco. There were 75 protein synthesis genes including 32 tRNA genes, 5 rRNA genes, 11 small ribosomal unit genes, 8 large ribosomal unit genes, 4 RNA polymerase subunit genes. The genome also has 4 hypothetical genes. (Table 1) The plastome of marine alga *P. kessleri* (MN013385) strain has 4 extra genes viz. mbpX, pbf1, psb30 (PSII protein) and trnM-cau in comparison to the freshwater strain *P. kessleri* (NC_012978.1). However, both strains do not have the trnR(CCG) but having 10 large ribosomal subunit encoding genes (*rpl20*, *rpl19*, *rpl16*, *rpl14*, *rpl5*, *rpl32*, *rpl23*, *rpl2*, *rpl12*, *rpl36*).. Both the plastomes have 5 hypothetical genes (*ycf1*, *ycf3*, *ycf4*, *ycf47* and *ycf20*).. Conserved gene clusters in LSC region were observed among marine *P. kessleri* and freshwater *P. kessleri* with little or no change in the genes arrangement (29).

2.2. Comparative Analysis

Genomic alignment was carried out using MAUVE software (30) in order to identify evolutionary changes in the DNA by alignment of homologous regions of sequence. Both the marine *P. kessleri*-I and freshwater *P. kessleri* plastome were laid out horizontally. The colored blocks represented homologous segments that are connected across genomes. The blocks that lie above the centre line are aligned in the forward orientation relative to the first genome sequence i.e *P. kessleri*-I. Blocks below the centre line represented the region that aligns in reverse complement orientation. Comparative analyses showed that both the chloroplast genomes of marine and freshwater algae have a significant homology as shown in *Fig. 2A* Mauve colored region of the similarity plot represented the conserved part in both the genomes. *Fig. 2B* Average level of conservation in a particular region was shown by the height of similarity profile.

2.3. Chloroplast transformation and selection of transgenic cultures

The growth of *P. kessleri*-I was optimized using a semisolid TAP medium and varying concentrations of zeocin and spectinomycin were used to determine the antibiotic sensitivity for *P. kessleri*-I cells growth. The 20 mgL⁻¹ zeocin and 400 mgL⁻¹ spectinomycin were observed to be ceased to growth of wildtype algal cells and used as an optimum antibiotics concentrations for selecting the transgenic alga after introducing the PkCpV cassette via bombardment. The chloroplast transformation vector containing PkCpV cassette of *Sh-Ble* gene confers resistance to bleomycin, phleomycin, and Zeocin (Sh ble protein: QBQ65853) was placed under 16S-rn PEP promoter (AC# KY426960.1) along with 5'UTR-g10 translator of *E.coli* (AC# AF176637). The *aadA* gene (AC# AY442171) confers resistance to spectinomycin was placed under the light regulated promoter psbA along with 5'UTR-psbA translator and the *aadA* gene product was terminated with 3'UTR-psbA (similar to AC# NC_001879). The PkCpV cassette was designed *in-silico* as shown in supplementary figure S1 and DNA vector was synthesized (BioBasic Inc. Canada).

Following the reported protocol (31), *P. kessleri*-I cells lawn was prepared and bombarded with the chloroplast transformation vector PkCpV. There were 60 plates were bombarded using 1350 psi rupture disk at the distance of 6 cm while 50 plates were bombarded at the distance of 9 cm. The bombarded algal plates were incubated in the dark for two nights. Thereafter, approximately equal number of cells (1×10^6) were plated on the selection medium containing zeocin and spectinomycin on the third day. A single green colony was observed bombarded at 6 cm distance while two green colonies when bombarded at 9 cm after 3–4 weeks on the TAP medium containing 400 mg/l spectinomycin. However, no transgenic colonies appeared on the TAP medium containing zeocin (20 mg/l), even after six weeks after the bombardment. The spectinomycin resistant colonies were further subcultured after every two weeks on antibiotic containing TAP medium for 4 months. Transgenes integration was analysed by both PCR and Southern blot.

2.4. Analysis of transgene integration in chloroplast genome

The three spectinomycin resistant transgenic colonies (*Fig. 3B*), 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 extracting the total genomic DNA. The presence of *aadA* gene integrated into the chloroplast genome was determined by PCR as described in Material and Methods. The expected PCR amplicon of size ~850 bp of *aadA* gene was observed, which confirmed the transgenes integration into *P. kessleri-I* cells (*Fig. 3C*). There was no PCR-amplification was observed in the WT cells.

To ascertain if the targeted DNA cassette has indeed been integrated at the specified position in the chloroplast genome between 16S-trnI and trnA–23S locus; genomic DNA of transgenic 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. 3A*. Southern blot analysis was performed on the digested genomic DNA isolated from transgenic cells line T1, T2 and T3. The plasmid PkCpV was digested with PvuI and NotI to get the *aadA* fragment (~ 0.4 kb) used as a probe. Genomic DNA was digested with NcoI and HindIII and hybridized with the probe as shown in *Fig. 3A*. All the transgenic samples showed hybridization signals with probe (*Fig. 3D*). The transformed algal cell lines thus confirmed transgenes integration at specified sites of chloroplast genome of *P. kessleri-I* via homologous recombination.

Transgenic algal cell line T, T2 and T3 were maintained as a frozen cultures in minus –70°C as reported (14) were not exposed or disposed into the environment at the end of the experiment.

2.5. Comparative growth analysis of wild type and transgenic culture

The total biomass and lipid content were determined as dry cell weight (DCW) after harvesting one-liter *P. kessleri-I* transgenic lines selected on TAP medium containing spectinomycin and WT cells grown without antibiotic. The dry cell weight was observed about same in transgenic cells T1 (315±0.21 mg/L), T2 (290±0.62 mg/L) and T3 (319±0.53 mg/L) compared to WT (342±0.35 mg/L). Similarly, total lipid yield was observed in transgenic cells T1 (97±0.2 mg/L), T2 (65±0.80 mg/L) and T3 (78±0.46 mg/L) compared to WT (88±0.4 mg/L). These observations indicate that there was no adverse impact on the transgenic culture due to the transgenes load on the genetically modified chloroplast genome.

Discussion

Chloroplasts are members of a class of organelles known as plastids and are found in plant cells and eukaryotic algae. As the site of photosynthesis, chloroplasts are the primary source of the world's food productivity and they sustain life on the Earth. Other important activities that occur in plastids include evolution of oxygen, sequestration of carbon, production of starch, synthesis of amino acids, fatty acids, and pigments, and key aspects of sulphur and nitrogen metabolism. Chloroplast genomes or plastomes are circular DNA molecules having approximately 100–200 genes, most of which encode photosynthetic, transcription and translational apparatus. The chloroplastic gene structure and expression are prokaryotic; hence the genes are usually arranged as operons. The genes are transcribed by eubacterial-type RNA Polymerase and the mRNA translated on 70S ribosome (32). Sequence information about the chloroplast genome can pave way to establish toolbox for its efficient engineering and further the use of microalgae as cell factories for the production of useful biomolecules. The whole genome of marine *P. kessleri*-I (I named after ICGEB) strain was sequenced through the Illumina Platform. The complete *P. kessleri*-I chloroplast genome is 109,642 bp in length (Gen Bank Accession No. MN013385), with 117 unique genes was assembled as a quadripartite structure. The decoded DNA sequence of the IR *i.e.* 16S-trnI-trnA-23S of *P. kessleri*-I was used as flanking region to integrate the transgenes cassette via homologous recombination. Thus, successful chloroplast transformation was optimized for oil-bearing marine alga *P. kessleri*-I.

The most important part relevant to chloroplast genome was DNA sequence present in the IR. 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 (MN013385) consists of 109,642 bp while freshwater *P. kessleri* (NC_012978.1) has 123,994 bp with 70% AT rich region (29). *P. kessleri*-I contains 117 genes (Table 1) whereas *P. kessleri* freshwater strain contains 112 genes. The additional genes present in marine strain are *pbf1* (photosynthesis biogenesis factor), *mbpX* (similar to *cysA* for ATP binding), *Psb30* (PSII reaction centre subunit). The size of *P. kessleri*-I chloroplast genome was smaller than its freshwater counterpart, 109,642 bp and 123,994 bp respectively. Genome progressive alignment was carried out in order to ascertain if *P. kessleri*-I has undergone any genome rearrangement, gene gain, loss, and duplication. MAUVE software was used to align the two genomes and it was noticed that maximum region was conserved in both the genome sequences *Fig. 2B*.. This would indicate a common ancestry for both the marine and freshwater species that may have acclimatized according to their diverse habitat.

Chloroplasts are ideal hosts for expression of transgenes (10). 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 (33). Also, protein folding and disulphide bond formation occurs readily in the chloroplasts making it ideal platform for production of proteins with multiple domains or subunits. Moreover, the chloroplast provides a sub-cellular compartment for accumulation of the recombinant proteins. 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 RecA protein. RecA also helps in maintaining chloroplast DNA integrity by participating in homologous recombination DNA repair (34). Complete homology of the plastid DNA flanking sequences ensures highly efficient chloroplast transformation. The PkCpV cassette was flanked by endogenous 16S rDNA-trnI sequence on the left and trnA–23S rDNA sequence on the right. 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 (35), hence this region was chosen as site specific integration of transgenes attributing to its high transcriptional activity. Using the NcoI site and HindIII restriction enzyme digestion of transgenic genome, we have ensured that transgenes *Sh-Ble* and *aadA* have precisely integrated between the *trnI* and *trnA* region. The HindIII restriction enzyme cuts outside the flanking region of vector, which confirmed that transgenes are integrated at the specified site. Also spectinomycin was observed to give better selection after transformation than zeocin. This may be because of heightened sensitivity of *P. kessleri-I* to the antibiotic zeocin in absence of any salts as increasing NaCl beyond 0.2M decreases the antibiotic sensitivity in microalgae (36). Similar observation is reported by Muñoz et al, 2019 (37) where zeocin was tested for sensitivity against *Acutodesmus obliquus* and *Neochloris oleoabundans* but was not used for selection of transformants.

The comparative growth analysis of WT and transgenic cell cultures have produced around same amount of biomass and lipid that indicate that there was no detrimental impact of transgenes, which have been integrated into chloroplast genome of *P. kessleri-I*. Similar growth and lipid production was observed in transgenic and WT cells. Chloroplast transformation of other microalgae has resulted in a variety of bio products such as vaccines, monoclonal antibodies, biocatalysts etc. *Chlamydomonas reinhardtii* chloroplast has been extensively exploited for production of subunit vaccines, monoclonal antibodies, immunotoxins and cancer cell therapeutics (33). Chloroplast of another marine microalga *Dunelliella tertiolecta* was transformed for production of enzymes such as xylanase, α -galactosidase, phytase, phosphatase, and β -mannanase (38).

Conclusion

Marine *P. kessleri-I* (CGEB strain) is an oil bearing robust strain and industrially important for producing biofuel molecules. We studied the whole chloroplast genome of *P. kessleri-I* strain, isolated 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 in order to confirm the repeat regions using NCBI Blast. The complete *P. kessleri-I* chloroplast genome is 109,642 bp in length, with 117 unique genes. The trnI-GAU, trnA-UGC, 16S and 23S ribosomal DNA unique sites helped in obtaining precise integration and expression of transgenes. The optimized chloroplast transformation via homologous recombination at the specific should help in improving the *P. kessleri-I* strain for enhancing biofuel properties. This study should open a new gateway in chloroplast engineering of *Parachlorella* spp and improve the biofuel values via carbon concentration mechanism, overexpression of RuBisCo enzyme for higher photosynthesis, expression of short chain fatty acid thioesterases to make biofuel suitable for drop in jet-fuel or producing value added products.

Abbreviations

IRA & IRB: Inverted Repeat Region A and B

LSC: Large Single Copy

SSC: Small Single Copy

DNA: Deoxyribonucleic acid

CCM: Carbon Concentration Mechanism

UTRs: Untranslated Regions

PSI & PSII: Photosystem I and II

ATP: Adenosine triphosphate

WT: Wild Type

TAP: Tris Acetate Phosphate Medium

PCR: Polymerase Chain Reaction

DCW: Dry Cell Weight

pbf1: photosynthesis biogenesis factor

TSP: Total Soluble Protein

DTT: Dithiothreitol

CTAB: Cetyl Trimethyl Ammonium Bromide

RuBP: Rubisco Bis-phosphate Carboxylase/Oxygenase

RuBisCO: Ribulose–1,5-bisphosphate carboxylase oxygenase

NCBI: National Center for Biotechnology Information

BLAST: Basic Local Alignment Search Tool

DOGMA: Dual Organellar GenoMe Annotator

OGDRAW: Organellar Genome DRAW

Methods

5.1. Strain and growth conditions

The *P. kessleri*-I algal strain was isolated from Indian Ocean by Indian Institute of Technology, Madras and kindly provided by Mr. Shrikumar Suryanaran and further it was characterized in our laboratory as reported (14) as per the guidelines of National Biodiversity Authority (NBA) of India. All the isolation and The axenic culture of marine alga was maintained on agar plates containing f/2 medium (pH 8.0) (40) under 2000 Lux light intensity for 16:8 light:dark conditions and 25°C temp. Optical density at 750 nm was used to assess the growth of algal cells (14).

5.2. DNA Extraction: Qualitative and Quantitative analysis of gDNA

P. kessleri-I strain was scaled up using liquid f/2 medium. The culture was centrifuged at 3600xg and the pellet was resuspended in 3ml ice cold breaking buffer containing 5mM dithiothreitol (DTT) and 5mM sodium ascorbate. Algal cells were sonicated on ice using a cell disruptor (Sonics, 3 mm Tapered microtip probe) at frequency of 1 KHz, 5 sec/ 5 sec off pulse at 30% amplitude for 2 min. The algal cells were collected by centrifugation and pellet was resuspended in 1 ml buffer containing 10 mM HEPES, 600 mM sorbitol, 50 mM MgCl₂ and 0.1% bovine serum albumin (BSA), pH 7.5–7.8 at 4°C. The autofluorescence of intact chloroplasts was monitored using Fluorescence Brightner 28 (Sigma Aldrich). The fluorescence from stained cellular components was observed at 650 nm and autofluorescence 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 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 at the middle of the tube (41). The plastomic DNA was extracted using CTAB (Cetyl Trimethyl Ammonium Bromide) buffer method (42). The quality of genomic DNA was checked on 1% agarose gel for a single intact band. 1µl of 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.

5.3. 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 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.

5.4. Chloroplast genome Assembly and Annotation

The chloroplast genome of *P. kessleri*-I was assembled *de novo* employing NOVOPlasty v3.0, using untrimmed reads and Rubisco-bis-phosphate oxygenase (RuBP) as the seed sequence (28). The other specified parameters included genome range of 100000–220000, K-mer value of 39, reference sequence of reported fresh water *P. kessleri* chloroplast genome ([NC_012978.1](#)) and the paired-end reads option. A web based annotation tool, GeSeq (part of the CHLOROBX toolbox) designed for annotation of organellar sequences (43) was used for annotating the chloroplast genome. Annotation was confirmed manually as well as by using NCBI BLAST (39) and DOGMA (44). Visualization of the cp genome was done using OGDRAW (45). All transfer RNA sequences (tRNA) encoded in the cp genome were verified using ARAGORN v1.2.38 (46) with default parameters.

5.5. Comparative analysis

The plastome of *P. kessleri*-IMN013385) and *Parachlorella kessleri* (NC_012978.1) was compared using MAUVE software: Multiple alignment of conserved genomic sequence with rearrangements (30). Progressive MAUVE genome alignment algorithm was used with default parameters in order to compare the two genomes so as to observe any gene loss, duplication, genome rearrangement, and horizontal transfer caused due to recombination.

5.6. Study of antibiotics sensitivity

Prior to chloroplast transformation, algal cells were adapted to grow on TAP medium (47) without addition of NaCl (salt) so that sensitivity to antibiotics can be accurately determined. In the presence of

salt, the algal cells may become tolerant to antibiotics (36). 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) and spectinomycin (100, 200, 300, 400, 500 and 600 mg/L) were tested using TAP medium solidified with 1.2% agar. Approximately 1.5×10^6 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 $25 \pm 1^\circ\text{C}$. The final concentrations of 20 mgL^{-1} zeocin and 400 mgL^{-1} spectinomycin were observed to completely cease the growth of *P. kessleri*-I cells. Therefore, 20 mgL^{-1} zeocin and 400 mgL^{-1} spectinomycin were taken into account as optimized antibiotic concentrations for exploring the chloroplast transformation in *P. kessleri*-I.

5.7. Chloroplast transformation vector

The chloroplast genome sequences were used for generating the flanking regions *in-silico*. Both left and right flanks contained the plastome fragments approx. 0.8 Kb of each, comprising of 16S rDNA and *trnI* and *trnA* and 23S rDNA respectively. Between the right and left flanks, transgenes cassette PkCpV contained 16S Prn 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. 3A. The whole DNA cassette of PkCpV vector prepared *in-silico* and DNA vector was synthesised with the help of 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.

5.8. Genetic transformation of *P. kessleri*-I and antibiotics selection

Prior to bombardment, algal cells lawn was prepared, containing high concentration of cells in the middle area of the Petri-plates. The chloroplast transformation vector DNA specific to *P. kessleri*-I (PkCpV) was coated on gold carrier particles S550d following vendor's instruction (Seashell Technology, CA, USA). About 10 μg of plasmid DNA (stock 1mg/ml) was added to 3 mg of gold carrier for 5 transfections. Thereafter, an equal volume of precipitation buffer (50 μL) was added. The total mixture (100 μL) was briefly vortex and incubated for 3 minutes. The mixture was spun (10,000 rpm for 10 sec) to pellet the DNA coated on gold particles. The supernatant was removed and 500 μL of 100% chilled ethanol was added. Again mixture was briefly vortex and spun (10,000 rpm for 10 sec) to pellet DNA-coated gold particles. Supernatant was discarded and 50 μL of 100% ethanol was added to DNA-coated gold particles. The pellet was homogenised in ethanol by sonication and 10 μL each of the suspension of DNA-coated gold particles was spotted in the middle of microcarrier (10 discs) (Bio-Rad, Hercules, CA, USA). The DNA gold particles were bombarded on the algal cells lawn using 1350 psi rupture disc at the 6 and 9 cm distance respectively (31). After overnight incubation of cultures in dark, bombarded cells were divided equally on TAP medium containing 20 mgL^{-1} zeocin and 400 mgL^{-1} spectinomycin.

5.9. PCR and Southern blot analysis

Transgenic colonies appeared only spectinomycin (400 mg/L) selection medium. No transgenic colonies were observed on TAP medium containing 20 mgL⁻¹ zeocin. Spectinomycin resistant colonies were screened by colony PCR using gene specific primer pair F1 (5'-GTCGACATGGAACAGAAGTTGATT-3') and R1 (5'-GTGGACAAATTCTTCCAAGTCT-3'). Each PCR reaction was performed in 50µl reaction mixture containing 1X Taq Reaction buffer (HiMedia Labs, Mumbai, India), 0.1µg DNA template, 200 µM of each dNTPs, 1 Unit Taq polymerase (HiMedia Labs, Mumbai, India), and concentrations of F1 and R1 used 0.5 µM of each oligo. A BioRad MyCycler (BioRad, USA) was employed for PCR amplification as shown in *Fig. 3C* 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 mg/L spectinomycin. The amplified PCR products were obtained using 35 cycles at 95°C denaturation of DNA template, 59°C annealing and extension at 72°C for 1 min and final extension at 72°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 µg genomic DNA was digested with NcoI and HindIII (restriction enzyme located outside the flanking region of 23S) as shown in *Fig. 3A*. The digested genomic DNA was electrophoresed on 0.8 % agarose gel, denatured and blotted onto Hybond positive nylon membrane. Probe was prepared as per the AlkPhos Direct Labeling Reagents kit (GE Healthcare) using *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).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent to publish

All authors read the manuscript and approved to publish.

Availability of data and materials

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

Competing Interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Department of Biotechnology, Government of India, and Grant No. BT/PB/Centre/03/ICGEB/2011(II) to SK. This extended grant to carry the research and developed in the field of algae bioenergy.

Authors' contributions

PN carried out carried out the transformation optimization experiments and carried out molecular analysis on transgenics along with CK and AT. SC and SS carried out the assembly and annotation of the chloroplast genome. MJ and MZA assisted in interpreting the results, reviewed and gave suggestions on 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 results and wrote the final manuscript along with SK. All authors have read the final manuscript and approved for the communication and publication.

Acknowledgements:

We would like to acknowledge 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.

Authors' information

Prachi Nawkarkar, Sagrika Chugh, Amita Tanwar, Shashi Kumar are working at the International Centre for Genetic Engineering and Biotechnology (ICGEB), Aruna Asaf Ali Marg, New Delhi 110067. Charli Kaushal and Malik Zainul Abdin are from the Department of Biotechnology, Jamia Hamdard, New Delhi 110062. Mukesh Jain works at the School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi 110067 and Surbhi Sharma worked at ICGEB, New Delhi 110067.

References

1. Lewis LA, McCourt RM. Green algae and the origin of land plants. *Am J Bot.* 2004;91(10):1535–56.
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 dictyochloropsis reticulata and from members of the genus myrmecia (chlorophyta, trebouxiophyceae cl.) *J Phycol.* 1995;31(4):632–9.

3. Friedl T. The evolution of the Green Algae. In Springer, Vienna; 1997. p. 87–101.
4. Pombert J-F, Otis C, Lemieux C, Turmel M. The Complete Mitochondrial DNA Sequence of the Green Alga *Pseudendoclonium 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):922–35.
5. Pombert J-F, Otis C, Lemieux C, Turmel M. The Chloroplast Genome Sequence of the Green Alga *Pseudendoclonium akinetum* (Ulvophyceae) Reveals Unusual Structural Features and New Insights into the Branching Order of Chlorophyte Lineages. *Mol Biol Evol.* 2005;22(9):1903–18.
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.
7. Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, et al. Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc Natl Acad Sci U S A.* 2002;99(19):12246–51.
8. Graham LE. Green algae to land plants: An evolutionary transition. *J Plant Res.* 1996;109(3):241–51.
9. Henry RJ, editor. Plant diversity and evolution: genotypic and phenotypic variation in higher plants. Wallingford: CABI; 2005
10. Verma D, Daniell H. Chloroplast Vector Systems for Biotechnology Applications. *Plant Physiol.* 2007;145(4):1129–43.
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.
12. Gimpel JA, Specht EA, Georgianna DR, Mayfield SP. Advances in microalgae engineering and synthetic biology applications for biofuel production. *Curr Opin Chem Biol.* 2013;17(3):489–95.
13. Zhang C, Hu H. High-efficiency nuclear transformation of the diatom *Phaeodactylum tricornutum* by electroporation. *Mar Genomics.* 2014;16:63–6.
14. Ahmad I, Fatma Z, Yazdani SS, Kumar S. DNA barcode and lipid analysis of new marine algae potential for biofuel. *Algal Res.* 2013;2(1):10–5.
15. Wang Q, Lu Y, Xin Y, Li W, Huang S, Xu J. Genome editing of model oleaginous microalgae *Nannochloropsis* spp. by CRISPR/Cas9. 2016;
16. Cui Y, Qin S, Jiang P. Chloroplast Transformation of *Platymonas* (*Tetraselmis*) *subcordiformis* with the bar Gene as Selectable Marker. Webber A, editor. *PLoS One.* 2014;9(6):e98607.
17. Rathod JP, Prakash G, Pandit R, Lali AM. Agrobacterium-mediated transformation of promising oil-bearing marine algae *Parachlorella kessleri*. *Photosynth Res.* 2013;118(1–2):141–6.
18. Mayfield SP, Franklin SE. Expression of human antibodies in eukaryotic micro-algae. *Vaccine.* 2005;23(15):1828–32.
19. Doron L, Segal N, Shapira M. Transgene Expression in Microalgae—From Tools to Applications. *Front Plant Sci.* 2016;7:505.

20. Lee S-B, Kwon H-B, Kwon S-J, Park S-C, Jeong M-J, Han S-E, et al. Accumulation of trehalose within transgenic chloroplasts confers drought tolerance. *Mol Breed*. 2003;11(1):1–13.
21. Nishimura EK, Jordan SA, Oshima H, Yoshida H, Osawa M, Moriyama M, et al. Dominant role of the niche in melanocyte stem-cell fate determination. *Nature*. 2002;416(6883):854–60.
22. Kuroiwa T, Kawano S, Nishibayashi S, Sato C. Epifluorescent microscopic evidence for maternal inheritance of chloroplast DNA. *Nature*. 1982;298(5873):481–3.
23. Daniell H. Molecular strategies for gene containment in transgenic crops. *Nat Biotechnol*. 2002;20(6):581–6.
24. Daniell H. Transgene containment by maternal inheritance: Effective or elusive? *Proc Natl Acad Sci*. 2007;104(17):6879–80.
25. Kumar S. GM Algae for Biofuel Production: Biosafety and Risk Assessment. Vol. 9, Collection of Biosafety Reviews. 2015.
26. Vuttipongchaikij S, Brocklehurst D, Steele-King C, Ashford DA, Gomez LD, McQueen-Mason SJ. Arabidopsis GT34 family contains five xyloglucan α -1,6-xylosyltransferases. *New Phytol*. 2012;195(3):585–95.
27. Li X, Přibyl P, Bišová K, Kawano S, Cepák V, Zachleder V, et al. The microalga *Parachlorella kessleri* –A novel highly efficient lipid producer. *Biotechnol Bioeng*. 2013;110(1):97–107.
28. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Res*. 2016 Oct 24;45(4):gkw955.
29. 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 Oct;26(10):2317–31.
30. Darling ACE. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Res*. 2004;14(7):1394–403.
31. Tran M, Mayfield SP. Expression of Full Length Monoclonal Antibodies (mAb) in Algal Chloroplast. In: *Antibody Engineering*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010. p. 503–16.
32. Green BR. Chloroplast genomes of photosynthetic eukaryotes. *Plant J*. 2011 Apr;66(1):34–44.
33. Dyo YM, Purton S. The algal chloroplast as a synthetic biology platform for production of therapeutic proteins. *Microbiology*. 2018;164(2):113–21.
34. 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):516–26.
35. De Cosa B, Moar W, Lee SB, Miller M, Daniell H. Overexpression of the Bt cry2Aa2 operon in chloroplasts leads to formation of insecticidal crystals. *Nat Biotechnol*. 2001;19(1):71–4.
36. Anila N, Chandrashekar A, Ravishankar GA, Sarada R. Establishment of *Agrobacterium tumefaciens*-mediated genetic transformation in *Dunaliella bardawil*. *Eur J Phycol*. 2011;46(1):36–44.
37. Muñoz CF, Sturme MHJ, D'Adamo S, Weusthuis RA, Wijffels RH. Stable transformation of the green algae *Acutodesmus obliquus* and *Neochloris oleoabundans* based on *E. coli* conjugation. *Algal Res*.

38. Georgianna DR, Hannon MJ, Marcuschi M, Wu S, Botsch K, Lewis AJ, et al. Production of recombinant enzymes in the marine alga *Dunaliella tertiolecta*. *Algal Res.* 2013;2(1):2–9.
39. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* [Internet]. 1990 Oct 5 [cited 2019 Jun 6];215(3):403–10.
40. Guillard RRL, Reyther JH. Studies of marine planktonic diatoms *I. Cyclotella nana* Hustedt and *Detonula confervacea* Cleve. *Can. J. Microbiol.* 8 (1962); 229-239.
- 141Angelova A, Park SH, Kyndt J, Fitzsimmons K, Brown JK. Sonication-based isolation and enrichment of *Chlorella protothecoides* chloroplasts for Illumina genome sequencing. *J Appl Phycol.* 2014;26(1):209–18
42. Clarke JD. Cetyltrimethyl Ammonium Bromide (CTAB) DNA Miniprep for Plant DNA Isolation. *Cold Spring Harb Protoc.* 2009;(3):pdb.prot5177
43. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 2017;45(W1):W6–11
44. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics.* 2004 Nov;20(17):3252–5.
45. 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):267–74.
46. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 2004;32(1):11–6.
47. Gorman DS, Levine RP. Photosynthetic Electron Transport Chain of *Chlamydomonas reinhardtii*. V. Purification and Properties of Cytochrome 553 and Ferredoxin. *Plant Physiol.* 1966;41(10):1643–7.

Table

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, psaI, 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, trnFM-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, ycf1</i>

Figure 1

Figure 1

Chloroplast genome map of *Parachlorella kessleri*-I. The predicted genes are shown and colours represent genes code, which are given at bottom left . Gene shown on the outside of the circle are transcribed

clockwise and on inside of the circle are transcribed anticlockwise.

Figure 2

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 genome sequence of marine *Parachlorella kessleri-I* genome that aligned to a corresponding part of the freshwater *Parachlorella kessleri* genome, and lines connect blocks of putative homology. Within the blocks the colored bars indicate the level of sequence similarities. b. Progressive Mauve was used to align the marine and freshwater *Parachlorella kessleri* genome. Backbone display mode shows the conserved regions between two genomes drawn in mauve color.

Figure 3

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* control the expression *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 confirmed the

transgenes integration at the specific location of chloroplast genome, Southern blot was performed on genomic DNA digested with NcoI and HindII (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.