

# Developing Strain-Specific Simple Sequence Repeat (SSR) Markers for *Chlorella sorokiniana*

Mais Sweiss<sup>1\*</sup>, Maen Hasan<sup>2</sup>, and Nidal Odat<sup>3</sup>

<sup>1</sup>Department of Biotechnology, Faculty of Agricultural Technology, Al-Balqa Applied University, Al-Salt 19117, Jordan

<sup>2</sup>Department of Plant Production and Protection, Faculty of Agricultural Technology, Al-Balqa Applied University, Al-Salt 19117, Jordan

<sup>3</sup>Department Medical Laboratory Sciences, Faculty of Science, Al-Balqa Applied University, Al-Salt 19117, Jordan

***Chlorella sorokiniana* green microalga offers many environmentally friendly applications, including wastewater treatment, biofertilizers, animal feed, and biofuel production. Different strains of *C. sorokiniana* have unique properties that may suit one application but not another. There is a need to distinguish between the many available strains of *C. sorokiniana* to choose the one that best fits the application. Consequently, our research goal was to develop strain-specific simple sequence repeat (SSR) markers to differentiate between the different strains. Seventeen markers spanning ten out of the twelve chromosomes of the *C. sorokiniana* genome were developed and validated on eight different strains from culture collections and our lab, and were then analyzed by fragment analysis. The results demonstrate the potential of these polymorphic markers to detect the genetic differences between the strains of *C. sorokiniana*, and to serve as useful tools for the intra-species population genetic analysis and conservation genetics studies of *C. sorokiniana*.**

**Keywords:** *Chlorella*, simple sequence repeats, microsatellites, strain-specific

## Introduction

*Chlorella*, a widely known genus of microalgae primarily found in freshwater, but with some species inhabiting marine environments, soils, or living as endosymbionts [1], presents promising applications in various fields, such as biofuel production, due to its high growth rate and lipid content [2]. Some *Chlorella* species proved to be efficient in removing nutrients and metals from wastewater [3, 4], while others can be used as food supplements and animal feed [5, 6].

However, identifying *Chlorella* members at the species level based on the morphology is challenging due to limited morphological distinctions [1], exacerbated by environmental influences on morphology [7, 8]. For these reasons, polyphasic identification methods integrating morphological, ultrastructural, and molecular approaches are recommended [9, 10]. Since identifying *Chlorella* members at the species level is hard, it is even harder to identify different strains of the algae.

To address this challenge, microsatellite markers, a subgroup of tandem repeats in the genome known for their high mutation rates [11, 12], offer a solution for strain differentiation as they are very effective for selecting a target organism from its closely related species [13]. The discrimination power of the simple sequence repeat (SSR) markers depends on their polymorphic distribution in specific loci [13-15]. The specific flanking regions close to the microsatellite can provide accurate genotyping of intra- or inter-specific hybrids with a primer set [13, 16].

Selection of highly productive algal strains with favorable characteristics is a very important step to establish a successful, and economically viable, large-scale cultivation system. Understanding strain-specific differences in genome content and architecture is a significant issue in this field [17]. Microsatellite markers are a powerful tool for intraspecies level identification [13]. However, no microsatellite markers have been developed for *C. sorokiniana* strains yet. Thus, we sought in this research to develop SSR microsatellite markers to differentiate between *C. sorokiniana* strains.

## Materials and Methods

### *Chlorella* Strains

A total of nine *Chlorella* strains were used in this research. Three strains of *C. sorokiniana*, UTEX B 3016, UTEX 3010, and UTEX B 2805, were obtained from the culture collection of the University of Texas (UTEX) in the USA. Another two strains, *C. sorokiniana* SAG 211-31 and SAG 211-32, were obtained from the culture collection of

Received: April 16, 2024  
Accepted: July 5, 2024

First published online:  
July 18, 2024

\*Corresponding author  
E-mail: M.Swies@bau.edu.jo  
Phone: +00962779244122  
Fax: +009625557349

pISSN 1017-7825  
eISSN 1738-8872

Copyright © 2024 by the authors.  
Licensee KMB. This article is an  
open access article distributed  
under the terms and conditions  
of the Creative Commons  
Attribution (CC BY) license.

algae at Goettingen University (SAG) in Germany. One strain of *C. sorokiniana*, NIES-2169, was obtained from the microbial culture collection of the National Institute for Environmental Studies (NIES) in Japan and used as the positive control. Strain NIES-2169 is the same as strain UTEX 1230, which its genome was used to develop SSR markers. In addition, two strains of *C. sorokiniana* were available in our lab, Somerton-3, isolated from the United Kingdom, and Jordan-40, isolated from Jordan [18]. The strain *C. vulgaris* CCALA 269, used as a negative control, was obtained from the culture collection of autotrophic organisms in the Czech Republic. All the strains were cultured on Bold's Basal Medium (BBM) plates with 1% agar [19], and grown at 22°C with a light:dark cycle of 18:6 h at a light intensity of 60-80  $\mu\text{mol photon/m}^2/\text{s}$ .

#### DNA Isolation

A single colony from each strain was cultured in 10 mL BBM under the following conditions: light:dark cycle of 16:8 h, light intensity of 120-150  $\mu\text{mol photon/m}^2/\text{s}$ , 22°C, and mixing at 250 rpm. After the culture bloomed, the DNA was isolated following the protocol of Doyle and Doyle (1987) with minor modifications [20]. Around  $30 \times 10^6$  cells were collected and homogenized in liquid nitrogen. Then, 750  $\mu\text{l}$  of preheated (60°C) CTAB isolation buffer [2% hexadecyltrimethylammonium bromide (CTAB), 20 mM Ethylenediaminetetraacetic acid (EDTA), 1.4 M NaCl, 100 mM Tris-HCl pH 8.0] was used to lyse the cells. The samples were then incubated at 60°C for 30 min with gentle mixing, followed by purification with the same volume of chloroform, and centrifugation at  $8,000 \times g$  for 10 min. Next, the nucleic acids were precipitated by adding the same volume of ice-cold isopropanol, followed by incubation at -20°C for 30 min, and then collection by centrifugation for 15 min at  $15,000 \times g$ . The pellet was washed with 500  $\mu\text{l}$  of 80% ethanol and dried, and then it was resuspended in 40  $\mu\text{l}$  of 1X TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). Subsequently, the isolated genomic DNA was analyzed on 0.7% agarose gel in 0.5X TBE running buffer, stained with 0.5  $\mu\text{g/ml}$  ethidium bromide.

#### Amplification of Internal Transcribed Spacer (ITS) Region and the Plastid Elongation Factor Tu (*tufA*) Gene

To amplify the ITS1-5.8S-ITS2 region, the ITS-F 5'-GAAGTCGTAACAAGGTTTCC-3' and ITS-R 5'-TCCTGGTTAGTTTCTTTCC-3' primers were used [21]. Whilst, for the chloroplast *tufA* gene amplification, *tufAF* 5'-TGAAACAGAAAWCGTCATTATGC-3' and *tufAR* 5'-CCTTCNCGAATMGCRAAWCGC-3' primers were used [22]. The PCR mixes were prepared according to the instructions of the Phusion High-Fidelity DNA Polymerase Kit (New England BioLabs Inc., USA). In a total volume of 25  $\mu\text{l}$  containing 50-250 ng genomic DNA, dNTPs 240  $\mu\text{M}$ , from each primer 0.4  $\mu\text{M}$ ; 1X of 5X Phusion GC Buffer, 3% DMSO, and 0.5 U of the high-fidelity DNA polymerase. The thermal cycling conditions were initial denaturation at 94°C for 4 min, followed by 30 cycles of denaturation at 94°C for 45 s, annealing at 57°C for the ITS region, and 58°C for *tufA* gene for 45 s. The extension step was at 72°C for 30 s, followed by a final extension at 72°C for 7 min using Prime Thermal Cycler (Prime, UK). The results were analyzed on 1% agarose gel stained with 0.5  $\mu\text{g/ml}$  ethidium bromide.

#### Sequencing

The PCR products were purified using an agarose gel extraction kit from Jena Bioscience (Jena Bioscience, Germany), and sent for sequencing by Macrogen (Macrogen, Republic of Korea). For each PCR product, two sequences were obtained, one for the forward and one for the reverse reaction. After that, the low-quality sequences were removed by trimming these sequences from both the beginning and the end, and then the two sequences, the forward and the reverse complement of the reverse, were aligned to obtain the reliable sequence for each PCR product.

#### Phylogenetic Analysis based on the ITS Region and *tufA* Gene

The identity of the obtained strains was confirmed by doing a search using the Nucleotide Basic Local Alignment Search Tool (Nucleotide BLAST), provided by the National Center for Biotechnology Information (NCBI) [23, 24]. To elucidate the phylogenetic relationships between the *Chlorella* strains based on the two amplified regions, a maximum likelihood method based on the Tamura-Nei model was used to construct the phylogenetic trees using MEGA 11 [25, 26].

#### Searching for SSRs and Designing Strain-Specific Primers

The whole-genome sequence of *C. sorokiniana* strain UTEX 1230 (NIES-2169) was obtained from the National Center for Biotechnology Information (NCBI) database (<http://www.ncbi.nlm.nih.gov/>) under the GenBank assembly Accession No. GCA\_003130725.1. The candidates for the microsatellite motifs were screened using the Unipro UGENE program [27]. Strain-specific primers were designed based on the upstream/downstream flanking sequences of the chosen microsatellite motifs using Primer3Plus [28]. The criteria for primer design included the prediction of annealing temperature ( $T_m$ ) of 55°C-61°C, primer length ranging between 18-27 bp, PCR amplicon lengths of 200-1000 bp when possible, and an annealing temperature difference between both pairs of primers of under 1°C. Primers that are self-complementary and complementary to each other were not chosen. For fragment analysis, on the 5' end of the forward primers, the M13 complementary sequence was added to allow the attachment of the M13 6 FAM-labeled primer.

#### SSR PCR Amplification and Genotyping

The seventeen pairs of the SSR primers covered most of the chromosomes, except chromosomes 4 and 12, and were verified on eight strains of *C. sorokiniana* using touchdown PCR. The PCR reactions were performed in a

Prime Thermal Cycler (Prime) in a total volume of 25  $\mu$ l using AmpliTaq Gold 360 Master Mix Kit (Applied Biosystems, USA) according to the manufacturer's instructions containing 20-50 ng of genomic DNA, 1X AmpliTaq Gold 360 Master Mix, 1  $\mu$ l of 360 GC Enhancer, 0.2  $\mu$ M of reverse primer, and 0.2  $\mu$ M of the forward primer mixture (the primers forward 1:3 of the volume and an M13 6 FAM-labeled primer for the rest of the volume).

Touchdown PCR was applied to increase the specificity and sensitivity of the reaction, in which the initial denaturation step was at 95°C for 4 min. Stage 1 consisted of 5 cycles of denaturation at 95°C for 45 s, with annealing starting at 68°C for 2 min with decreases of 2°C per cycle, and an extension step at 72°C for 1 min. Stage 2 consisted also of 5 cycles of denaturation at 95°C for 45 s, with annealing starting at 58°C for 1 min with decreases of 2°C per cycle, and an extension step at 72°C for 1 min. The third stage consisted of 27 cycles of denaturation at 95°C for 45 s, annealing at 52°C for 1 min, an extension step at 72°C for 1 min, and a final extension step at 72°C for 10 min. The quality and quantity of PCR products were validated on a 1.2% agarose gel stained with 0.5  $\mu$ g/ml ethidium bromide.

### Fragment Analysis

To detect the precise size of the amplified PCR products, the PCR products were sent for fragment analysis by MacroGen (MacroGen) to be analyzed by 3730xl DNA (ABI, USA) Analyzer. The results were analyzed using Peak Scanner v1.0 (Applied Biosystems).

### Data Analysis

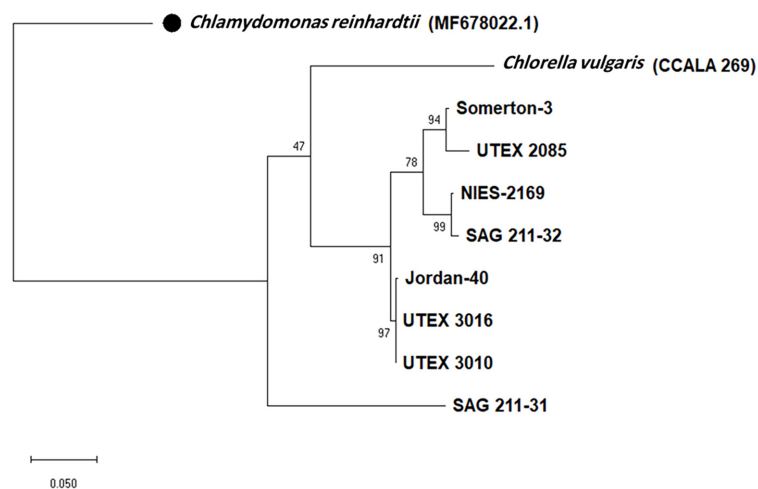
Polymorphic amplified SSR markers were classified as either present (1) or absent (0), and then transformed into a binary data matrix. To detect the relationship between the eight studied strains of *C. sorokiniana*, genetic similarity was estimated following the Dice similarity coefficient [29], which was calculated between all possible pairs of strains by using the SimQual method implanted in NTSYSpc (Version 2.01, Exeter Software, Setauket, NY, USA). Based on this similarity matrix, a tree showing the relationships among the different *C. sorokiniana* strains was constructed by the unweighted pair group method with arithmetic average (UPGMA) using the SAHN modules in the same software.

## Results

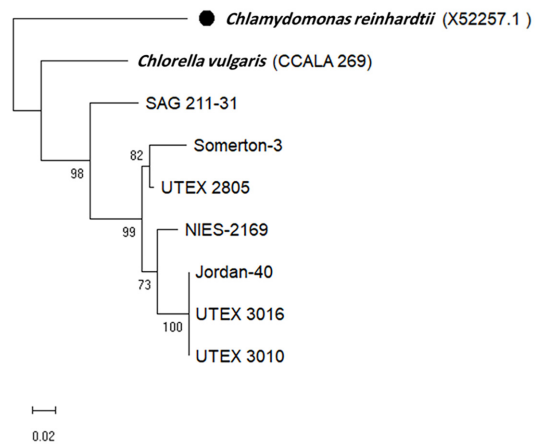
### Phylogenetic Analysis Based on the ITS Region and *tufA* Gene

To confirm the identity of the obtained *C. Sorokiniana* strains, two DNA barcoding markers were chosen, the ITS region and *tufA* gene, which were amplified and sent for sequencing. The ITS region was selected because its nuclear rDNA substitution rates are much higher than those of the rRNA genes, and it was also recommended as a marker for species identification of macro- and microalgae. In addition, *tufA* was chosen because it is recommended as a marker for DNA-based species delimitation and/or barcoding in green algae [30].

The identification of the *C. sorokiniana* strains based only on the BLAST search for their ITS and *tufA* sequencing was challenging, because of the limited sequences available in the database. The phylogenetic analysis for the obtained sequences for the ITS region and *tufA* gene can show differences at the strain levels, and the phylogenetic tree was rooted in the outgroup *Chlamydomonas reinhardtii* sequences which were obtained from GenBank. Some strains are closely related to each other and cluster together, such as UTEX 3010, UTEX 3016, and Jordan-40, as well as Somerton-3 and UTEX 2805 (Figs. 1 and 2), while SAG 211-31 stands alone on separate branches indicating that it is different from the rest of the strains based on the two loci. We remind the reader that



**Fig. 1. Maximum likelihood phylogenetic tree of ITS region.** The tree was constructed using MEGA11, the bootstrap value was 10000, and the black circle is for outgroup used to root the tree, which is *Chlamydomonas reinhardtii* (Accession No. MF678022.1).

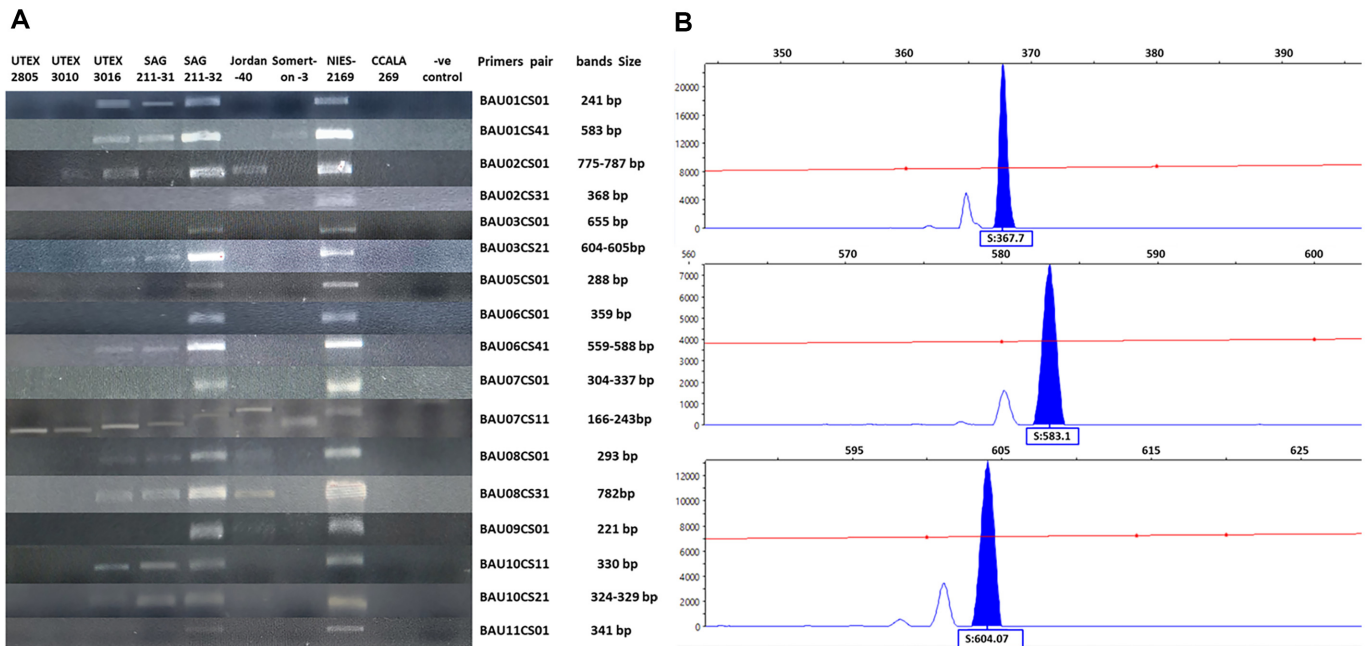


**Fig. 2. Maximum likelihood phylogenetic tree of *tufA* gene.** The tree was constructed using MEGA11, the bootstrap value was 10000, and the black circle is for the outgroup used to root the tree which is *Chlamydomonas reinhardtii* (Accession No. X52257.1).

the aim of these trees was not to study the genetic variation, but only to confirm the identity of the strains before we began the SSR marker validation.

**Validation of SSR Primer Design**

Seventeen pairs of primers that present in ten out of the twelve chromosomes of *C. sorokiniana* genome were designed and tested. These primers were validated on seven different *C. sorokiniana* strains and the positive control (NIES-2169), the genome sequence of which was used to design the primers, so it should show positive results, in addition to a negative control of a different *Chlorella* strain (CCALA 269) that should be negative (Fig. 3A). After the touchdown PCR, the PCR products were subjected to fragment analysis to detect the precise size of the amplified fragment (Fig. 3B). The results are summarized in Table 1. All these markers showed polymorphism among the eight examined strains (Table 1). Genetic similarity was assessed by the Dice similarity coefficient, and based on the similarity results, the Dice genetic similarity for all investigated strains ranged from 0.000 to 0.9375, with an overall mean of 0.191 (Table 2). The UPGMA tree represents the genetic similarity of the



**Fig. 3. SSR primer validation and fragment analysis.** (A) Touchdown PCR results were analyzed on 1.2% agarose gel stained with 0.5µg/ml ethidium bromide, NIES-2169, *C. sorokiniana* strain positive control; CCALA 269, *C. vulgaris* a negative control; -ve control, master mix without DNA. (B) Fragment analysis results for some samples, band size is shown in a box below the peak.

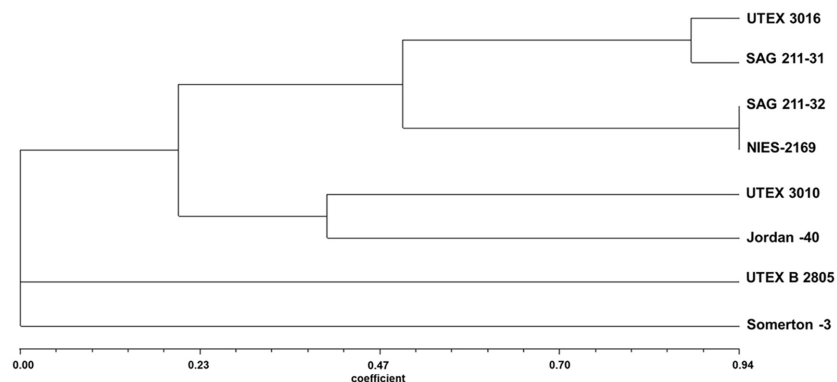
**Table 1. The designed SSR primers and their results after fragment analysis without the negative controls.**

	Primer name	Primer	Locus	Repeat motif	Size range (bp)	No. of non-amplifying samples
1	BAU01CS01_F BAU01CS01_R	GTAAAACGACGGCCAGCTCT TCAACATCTGCCATCACAAT AGGCGATGCTGGGTCTATG	Chromosome 1	(TGC)16	241	4
2	BAU01CS41_F BAU01CS41_R	GTAAAACGACGGCCAGAGC AGGGAGGAGGACGAGGACA GGCAGTGACATAGGG	Chromosome 1	(GCA)15	583	3
3	BAU02CS01_F BAU02CS01_R	GTAAAACGACGGCCAGGGG ATGGTTGAGGGTGCTAGGAT GGCTGCTCTGCTC	Chromosome 2	(GCT)115	775-787	2
4	BAU02CS31_F BAU02CS31_R	GTAAAACGACGGCCAGCCAG CCGGTCAATCAGATCAGGGG ACAAGAGCAAGGC	Chromosome 2	(TGC)10	368	6
5	BAU03CS01_F BAU03CS01_R	GTAAAACGACGGCCAGCATA CTGGGAGCGGTGCGGTCTC ACAGCAGCCTCAG	Chromosome 3	(GGC)5(GCT) 25	655	6
6	BAU03CS21_F BAU03CS21_R	GTAAAACGACGGCCAGGTG GATTGCTGCTGAGACGTGCT GGTATCCTGTGTAGCC	Chromosome 3	(TGC)20	604-605	4
7	BAU05CS01_F BAU05CS01_R	GTAAAACGACGGCCAGAAAC GACATCCAAAAGACCAACAA GACAGGAAGCAGATGGA	Chromosome 5	(TG)23	288	6
8	BAU06CS01_F BAU06CS01_R	GTAAAACGACGGCCAGAAA AGTCGCCAGAAATCCACCTG CTGGAAGTCTGAGGT	Chromosome 6	(GCA)21	359	6
9	BAU06CS41_F BAU06CS41_R	GTAAAACGACGGCCAGCTGT CAAAGCATCCCCATCGTGTT TGCTATTTGGGCGGAC	Chromosome 6	(CAG)22	559-588	4
10	BAU07CS01_F BAU07CS01_R	GTAAAACGACGGCCAGGCTG GACTTGATGCCCTGAGTGTA CTGGCTGGCGATCT	Chromosome 7	(GCT)48	304-337	6
11	BAU07CS11_F BAU07CS11_R	GTAAAACGACGGCCAGCCG ACAAATCTAACCGCCTCAG CAGCAGACGACCAGC	Chromosome 7	(CTG)24	166-243	0
12	BAU08CS01_F BAU08CS01_R	GTAAAACGACGGCCAGCTGT GGTCGGCAGGATTGAAATGA CGGGAGTGAGAATGG	Chromosome 8	(CTG)28	293	3
13	BAU08CS31_F BAU08CS31_R	GTAAAACGACGGCCAGGCCT CTCCACCTCCACACAGCGAA ACCACGACAACAG	Chromosome 8	(TGC)25	782	3
14	BAU09CS01_F BAU09CS01_R	GTAAAACGACGGCCAGGCAC GACTTGATCTTGGCAACCTT GGCGGTGGTTAATTGT	Chromosome 9	(TCT)19AC (TCT)4	221	5
15	BAU10CS11_F BAU10CS11_R	GTAAAACGACGGCCAGACCA ACCGTCACACTTCTCGAGC GAGGATTGAAGCAG	Chromosome 10	(CA)32	330	4
16	BAU10CS21_F BAU10CS21_R	GTAAAACGACGGCCAGGCG AGATGGGGTCAGGAGCTTAT TAGGGGCTGGGCAGA	Chromosome 10	(TG)55	324-329	4
17	BAU11CS01_F BAU11CS01_R	GTAAAACGACGGCCAGGCCT CGTGACCCAAACACCCAGC ACGACTTTCATACC	Chromosome 11	(TG)29	341	6

eight examined strains on 17 different loci distributed amongst the whole genome, and it showed that UTEX 2805 and Somerton-3 stand alone in different branches, indicating that they have the lowest similarity with the other strains based on the developed SSR markers (Fig. 4). On the other hand, SAG 211-32 and NIES-2169 have the highest genetic similarity observed (Table 2). Furthermore, SAG 211-31 is more similar to UTEX 3016 based on the developed SSR marker (Fig. 4).

**Table 2. Dice similarity matrix among the analyzed 8 genotypes based on SSR molecular data.**

	UTEX B 3016	UTEX 3010	UTEX 2805	SAG 211-31	SAG 211-32	Jordan -40	Somerton -3	NIES-2169
UTEX B 3016	1.000000							
UTEX 3010	0.222222	1.000000						
UTEX 2805	0.000000	0.000000	1.000000					
SAG 211-31	0.875000	0.000000	0.000000	1.000000				
SAG 211-32	0.521739	0.000000	0.000000	0.518518	1.000000			
Jordan -40	0.500000	0.400000	0.000000	0.307692	0.300000	1.000000		
Somerton -3	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	
NIES-2169	0.476190	0.000000	0.000000	0.480000	0.937500	0.315789	0.000000	1.000000

**Fig. 4. UPGMA Tree of eight *C. sorokiniana* strains based on UPGMA analysis of SSR markers.** The SSR markers represent 17 different loci of the whole genome. The tree was constructed using NTSYSpc software.

## Discussion

Eight different strains of *C. sorokiniana* were studied in this research. Six of these were obtained from culture collections while two came from our lab, and were isolated from wastewater from Jordan and the UK. Before starting the experiment, the strains were subjected to DNA barcoding and BLAST search to ensure that they were *C. sorokiniana*, and that there was no contamination or other problems. A phylogenetic tree for each DNA barcode was then constructed to confirm the identity results. These results represent the phylogenetic relationship among the isolated strains using one locus only (either ITS or *tufA*). To study the differences among the whole genome, 17 SSR markers were developed. SSR markers are a very valuable tool in studying the evolutionary process, population structure, and genetic mapping [11]. These primers were polymorphic and were able to show the genetic diversity between *C. sorokiniana* strains. The low similarity among some strains of *C. sorokiniana* in Fig. 4 has been reported by a previous comparative study on the genome of three *C. sorokiniana* strains, UTEX 1230 (NIES-2169), 1228, and UTEX 3016. The study found that less than 15% of the genome has nucleotide identity of 80%, which raises questions on the species taxonomy of the *C. sorokiniana* strains [17].

*C. sorokiniana* is a non-motile, single-cell organism that is directly in contact with the environment, thus the selection pressure is high among the cells. Moreover, *Chlorella* members are small in size, ranging from 2 to 15  $\mu\text{m}$  in diameter; they reproduce asexually within a few hours and can form large populations in a short period [31]. Besides that, they are haploid [32], and consequently, spontaneous mutations can be detected and traced within a few months in the cultured microalgae [33]. The genome sequence of *C. sorokiniana* strains revealed the presence of around 71 sex-related and meiosis-related genes, and a further 25 homologous recombination genes were found, which suggests that *C. sorokiniana* may be capable of dividing by meiosis and engage in homologous recombination [17]. In addition, genome sequencing discovered the presence of hotspots for genomic rearrangements and inversion in their genome [34]. The acquisition of DNA sequences from other organisms such as viruses has been documented in algae. More than 90,000 viral-origin sequences in 184 algal genomes have been detected, and thus viruses have played a role in the adaptation of their host to different environments [35]. These reasons and others have contributed to the high genetic diversity of *Chlorella* and allows them to adapt easily to selection pressure and environmental changes [33].

The increase of genetic diversity in a population is advantageous since it increases tolerance to harsh environmental conditions, and it is an important factor for maintaining ecological functions under varying environmental conditions [36]. This genetic diversity may explain the ability of different strains of *C. sorokiniana*, such as thermotolerance strains, to tolerate different environmental conditions. In addition to its thermotolerance, a *C. sorokiniana* strain that was isolated from water bodies around a steel plant in India can also tolerate high levels of carbon dioxide and nitric oxide [37]. Other thermotolerant strains, like *C. sorokiniana*

**Table 3. Summary of some of the microsatellites used to study genetic diversity in algae from literature listed in order from the oldest to the newest.**

Algae	Microsatellite marker	Main feature	Year/ Citation
<i>Gracilaria lemaneiformis</i> , <i>Gracilaria tenuistipitata</i> var. <i>liui</i> , <i>Gracilaria vermiculophylla</i> , <i>Gracilaria blodgettii</i> (Red macroalga)	Inter Simple Sequence Repeats (ISSR) <sup>†</sup>	Developed ISSR markers to investigate the genetic diversity of the genus <i>Gracilaria</i> (Rhodophyceae)	2003/ [55]
<i>Alexandrium tamarense</i> (Dinoflagellate microalgae)	Microsatellite *	Developed polymorphic microsatellite loci that provided microsatellite markers with high polymorphism	2004/ [49]
<i>Gracilaria lemaneiformis</i> (Red macroalga)	ISSR <sup>†</sup>	Used ISSR to distinguish and study the polymorphisms of the phases and sexes of <i>G. lemaneiformis</i>	2006/ [51]
<i>Alexandrium tamarense</i> (Dinoflagellate microalga)	Microsatellite*	Performed genetic analysis of <i>Alexandrium tamarense</i> populations from 10 sites along the Japanese and Korean coasts and also tried to detect the impact of natural and human-assisted dispersals on the genetic structure and gene flow	2007/ [48]
<i>C. vulgaris</i> , <i>C. pyrenoidosa</i> (Green microalgae)	ISSR	Used ISSR primers that were designed for <i>Triticum aestivum</i> to study genetic polymorphism and diversity of <i>Chlorella</i> for intra-species genetic analysis.	2008/ [44]
<i>Chondrus crispus</i> (Red macroalga)	ISSR <sup>‡</sup>	ISSR analysis was used to investigate genetic variations of haploid and diploid samples from nine North Atlantic	2008/ [56]
<i>Haematococcus pluvialis</i> (Green microalga)	ISSR	Using ISSR and Random Amplified Polymorphic DNA (RAPD) markers to study genetic diversity	2011/ [45]
<i>Ulva (Enteromorpha) prolifera</i> (Green macroalga)	ISSR <sup>‡</sup>	Using ISSR markers to provide information on the genetic variation of both attached and floating <i>U. prolifera</i> samples from China	2011/ [47]
<i>C. vulgaris</i> (Green microalga)	SSR	Developed and evaluated microsatellite markers along the chloroplast genome for strain-specific identification method	2014/ [13]
<i>Porphyra umbilicalis</i> (Red macroalga)	Microsatellite	Developed polymorphic microsatellite markers from enriched DNA libraries	2018/ [46]
<i>Prorocentrum donghaiense</i> (Dinoflagellate microalga)	SSR	Development of SSR marker based on transcriptome sequencing to study genetic diversity of the dinoflagellate blooms	2020/ [50]
Cladophoraceae species (Green macroalga)	SSRs	Designed a set of SSRs for species delimitation and insights into ploidy species of Cladophoraceae using high-throughput sequencing (HTS) data of three morphospecies	2020/ [52]
<i>Haematococcus pluvialis</i> (Green microalga)	Repeats and SSR	Identified among the chloroplast genome, potentially useful markers for detecting polymorphism used in evolutionary studies, which has high-degree variations within the same species	2021/ [54]
<i>Alaria esculenta</i> , <i>Pylaiella littoralis</i> (Brown macroalgae) <i>Calliblepharis jubata</i> , <i>Gracilaria gracilis</i> , <i>Gracilaria dura</i> , <i>Palmaria palmata</i> (Red macroalgae)	Microsatellites	Identified and characterised microsatellites in genomic sequences obtained using Double-Digest Restriction site Associated DNA (ddRAD), provided preliminary data about the genetic structure and reproduction mode of these six non-model species	2023/ [53]

\*Research used the same microsatellite marker

†Research used the same microsatellite marker

‡Research used the same microsatellite marke

LWG002615, which was isolated from Jeori thermal spring in India [38], and UTEX 2085, were found to grow at temperatures of 40–42°C and tolerate high light intensity of 2,500  $\mu\text{mol photon/m}^2/\text{s}$  for 5 h daily [39, 40]. Furthermore, some strains of *C. sorokiniana* can tolerate relatively high CO<sub>2</sub> concentrations [41, 42], and grow in wastewater. Three strains of *C. sorokiniana*, CS-01, UTEX 2714, and UTEX 1230(NIES-2169), behaved in different ways when they were grown in 10% diluted effluent of anaerobic digester from cattle manure [43]. The genetic diversity also enables some strains of *C. sorokiniana* to accumulate large amounts of lipid. In a project looking for top-performing green microalgae strains that would be suitable for biofuel feedstocks and performed

by the National Alliance for Advanced Biofuels and Bioproducts (NAABB) in the USA, *C. sorokiniana* strain DOE1412 (UTEX 3016) was a promising algal species for biofuel applications [2].

Several microalgal microsatellite markers have already been developed to study the genetic variation among the strains (Table 3) of some beneficial green microalgae. For instance, strain-specific SSR markers were developed for *C. vulgaris* and *C. pyrenoidosa* [13, 44]. Meanwhile, in case of *Haematococcus pluvialis*, the main astaxanthin-producing organism from different geographical regions, the strains were grouped based on their geographical locations [45]. In addition, seven microsatellite markers were identified and would be valuable for studying the genetic structure of the red alga *Porphyra umbilicalis* population [46]. On the other hand, microsatellites were used to study the genetic diversity of economically important harmful algae that cause economic loss, such as *Ulva (Enteromorpha) prolifera*, which causes green tides [47], or the toxic dinoflagellate *Alexandrium tamarense*, a cause of mass fish death in red tides [48, 49], and the dinoflagellate *Prorocentrum donghaiense*, whose blooms threaten coastal ecosystems [50]. Microsatellites were also applied to study the polymorphism of the different phases and sexes of the red alga *Gracilaria lemaneiformis* [51], and for species delimitation of *Cladophoraceae*, a species that is endemic to Lake Baikal [52]. A cost-effective microsatellite marker development method using Double-Digest Restriction site Associated DNA (ddRAD) was evaluated for six non-model species of red and brown seaweeds that have economic and ecological importance [53]. The microsatellite markers can be applied to study the chloroplast genome polymorphisms which is important for evolutionary studies. The presence of large numbers of SSRs may stimulate genome expansion, and a total of 401 SSRs were identified in the chloroplast genome of *H. pluvialis* and were dominated by mononucleotide repeats and tetranucleotide repeats [54].

## Conclusion

The genotyping of the eight *C. sorokiniana* strains based on the 17 developed SSR markers present in ten out of the twelve chromosomes of the genome suggests that while some strains are very closely related, such as SAG 211-32 and NIES-2169, there are two strains, Somerton-3 and UTEX 2805, that showed low similarity to the rest of the studied strains. This high genetic diversity among the strains may give these strains their unique characteristics, such as thermal tolerance. The developed SSR markers were able to successfully differentiate between the studied strains, and they could also provide a low-cost technique to further investigate the genetic diversity among the different *C. sorokiniana* strains.

## Acknowledgments

This research was funded by the Dean of Scientific Research and Innovation at Al-Balqa Applied University in Jordan.

## Conflict of Interest

The authors have no financial conflicts of interest to declare.

## References

- Bock C, Krienitz L, Proschold T. 2011. Taxonomic reassessment of the genus *Chlorella* (Trebouxiophyceae) using molecular signatures (barcodes), including description of seven new species. *Fottea* **11**: 293-312.
- Neofotis P, Huang A, Sury K, Chang W, Joseph F, Gabr A, et al. 2016. Characterization and classification of highly productive microalgae strains discovered for biofuel and bioproduct generation. *Algal Res.* **15**: 164-178.
- Kumar KS, Dahms H, Won E, Lee J, Shin K. 2015. Microalgae - promising tool for heavy metal remediation. *Ecotoxicol. Environ. Saf.* **113**: 329-352.
- Osundeko O, Davies H, Pittman JK. 2013. Oxidative stress-tolerant microalgae strains are highly efficient for biofuel feedstock production on wastewater. *Biomass Bioenergy* **56**: 284-294.
- Torres-Tiji Y, Fields FJ, Mayfield SP. 2020. Microalgae as a future food source. *Biotechnol. Adv.* **41**: 107536
- Ahmad MT, Shariff M, Md. Yusoff F, Goh YM, Banerjee S. 2020. Applications of microalga *Chlorella vulgaris* in aquaculture. *Rev. Aquac.* **12**: 328-346.
- Darienko T, Gustavs L, Eggert A, Wolf W, Proschold T. 2015. Evaluating the species boundaries of green microalgae (*Coccomyxa*, Trebouxiophyceae, Chlorophyta) using integrative taxonomy and DNA barcoding with further implications for the species identification in environmental samples. *PLoS One* **10**: e0127838.
- Luo W, Pflugmacher S, Proschold T, Walz N, Krienitz L. 2006. Genotype versus phenotype variability in *Chlorella* and *Micractinium* (Chlorophyta, Trebouxiophyceae). *Protist* **157**: 315-333.
- Krienitz L, Bock C. 2012. Present state of the systematics of planktonic coccoid green algae of inland waters. *Hydrobiologia* **698**: 295-326.
- Proschold T, Leliaert F. 2007. Systematics of the green algae: conflict of classic and modern approaches, pp 123-153, In Brodie J, Lewis J (eds.), *Unravelling The Algae: The Past, Present, and Future of Algal Systematics*. London, UK: Taylor and Francis, London, UK.
- Vieira MLC, Santini L, Diniz AL, Munhoz CDF. 2016. Microsatellite markers: what they mean and why they are so useful. *Genet. Mol. Biol.* **39**: 312-328.
- Gemayel R, Cho J, Boeynaems S, Verstrepen KJ. 2012. Beyond junk-variable tandem repeats as facilitators of rapid evolution of regulatory and coding sequences. *Genes* **3**: 461-480.
- Jo BH, Lee CS, Song HR, Lee HG, Oh HM. 2014. Development of novel microsatellite markers for strain-specific identification of *Chlorella vulgaris*. *J. Microbiol. Biotechnol.* **24**: 1189-1195.
- Li S, Zhang X, Yin T. 2010. Characteristics of microsatellites in the transcript sequences of the *Laccaria bicolor* genome. *J. Microbiol. Biotechnol.* **20**: 474-479.
- Kim J, Jo BH, Lee KL, Yoon ES, Ryu GH, Chung KW. 2007. Identification of new microsatellite markers in *Panax ginseng*. *Mol. Cells* **24**: 60-68.
- Chambers GK, MacAvoy ES. 2000. Microsatellites: consensus and controversy. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **126**: 455-476.



17. Hovde BT, Hanschen ER, Steadman Tyler CR, Lo CC, Kunde Y, Davenport K, *et al.* 2018. Genomic characterization reveals significant divergence within *Chlorella sorokiniana* (Chlorellales, Trebouxiophyceae). *Algal Res.* **35**: 449-461.
18. Sweiss MA. 2017. Microalgae for wastewater treatment and biomass production from bioprospecting to biotechnology. Ph.D., University of Bath, Bath, UK.
19. Andersen RA. 2005. *Algal Culturing Techniques*, pp. 437. Elsevier Academic Press, London, UK.
20. Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19**: 11-15.
21. Timmins M, Thomas-Hall SR, Darling A, Zhang E, Hankamer B, Marx UC, *et al.* 2009. Phylogenetic and molecular analysis of hydrogen-producing green algae. *J. Exp. Bot.* **60**: 1691-1702.
22. Famà P, Wysor B, Kooistra WHCF, Zuccarello GC. 2002. Molecular phylogeny of the genus *Caulerpa* (Caulerpales, Chlorophyta) inferred from chloroplast *tufA* gene. *J. Phycol.* **38**: 1040-1050.
23. Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schäffer AA. 2008. Database indexing for production MegaBLAST searches. *Bioinformatics* **24**: 1757-1764.
24. Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **7**: 203-214.
25. Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**: 512-526.
26. Tamura K, Stecher G, Kumar S. 2021. MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **38**: 3022-3027.
27. Okonechnikov K, Golosova O, Fursov M. 2012. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28**: 1166-1167.
28. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* **35**: W71-W74.
29. Dice LR. 1945. Measures of the amount of ecologic association between species. *Ecology* **26**: 297-302.
30. Leliaert F, Verbruggen H, Vanormelingen P, Steen F, López-Bautista JM, Zuccarello GC, *et al.* 2014. DNA-based species delimitation in algae. *Eur. J. Phycol.* **49**: 179-196.
31. Liu J, Chen F. 2016. Biology and industrial applications of *Chlorella*: advances and prospects, pp.1-35. In Posten C, Chen SF, (eds.), *Microalgae Biotechnology*, Springer International Publishing Switzerland.
32. Dörr R, Huss VAR. 1990. Characterization of nuclear DNA in 12 species of *Chlorella* (chlorococcales, chlorophyta) by DNA reassociation. *BioSystems* **24**: 145-155.
33. Krasovec M, Sanchez-Brosseau S, Grimsley N, Piganeau G. 2018. Spontaneous mutation rate as a source of diversity for improving desirable traits in cultured microalgae. *Algal Res.* **35**: 85-90.
34. Wu T, Li L, Jiang X, Yang Y, Song Y, Chen L, *et al.* 2019. Sequencing and comparative analysis of three *Chlorella* genomes provide insights into strain-specific adaptation to wastewater. *Sci. Rep.* **9**: 9514.
35. Nelson DR, Hazzouri KM, Lauersen KJ, Jaiswal A, Chaiboonchoe A, Mystikou A, *et al.* 2021. Large-scale genome sequencing reveals the driving forces of viruses in microalgal evolution. *Cell Host Microbe* **29**: 250-266.e8.
36. Sjöqvist CO, Kremp A. 2016. Genetic diversity affects ecological performance and stress response of marine diatom populations. *ISME J.* **10**: 2755-2766.
37. Varshney P, Beardall J, Bhattacharya S, Wangikar PP. 2018. Isolation and biochemical characterisation of two thermophilic green algal species- *Asterarcys quadricellulare* and *Chlorella sorokiniana*, which are tolerant to high levels of carbon dioxide and nitric oxide. *Algal Res.* **30**: 28-37.
38. Dasgupta CN, Singh VK, Nayaka S, Kishore S, Lavania S. 2020. Molecular phylogeny of a commercially important thermophilic microalga *Chlorella sorokiniana* LWG002615 and associated bacterium *Aquimonas* sp. NBRI01 isolated from Jeori thermal spring, Shimla, India. *Nucleus* **63**: 203-210.
39. Choi YY, Hong ME, Chang WS, Sim SJ. 2019. Autotrophic biodiesel production from the thermotolerant microalga *Chlorella sorokiniana* by enhancing the carbon availability with temperature adjustment. *Biotechnol. Bioprocess Eng.* **24**: 223-231.
40. de-Bashan LE, Trejo A, Huss VAR, Hernandez JP, Bashan Y. 2008. *Chlorella sorokiniana* UTEX 2805, a heat and intense, sunlight-tolerant microalga with potential for removing ammonium from wastewater. *Bioresour. Technol.* **99**: 4980-4989.
41. Nair A, Chakraborty S. 2020. Synergistic effects between autotrophy and heterotrophy in optimization of mixotrophic cultivation of *Chlorella sorokiniana* in bubble-column photobioreactors. *Algal Res.* **46**: 101799.
42. Kumar K, Banerjee D, Das D. 2014. Carbon dioxide sequestration from industrial flue gas by *Chlorella sorokiniana*. *Bioresour. Technol.* **152**: 225-233.
43. Kobayashi N, Noel EA, Barnes A, Watson A, Rosenberg JN, Erickson G, *et al.* 2013. Characterization of three *Chlorella sorokiniana* strains in anaerobic digested effluent from cattle manure. *Bioresour. Technol.* **150**: 377-386.
44. Shen S. 2008. Genetic diversity analysis with ISSR PCR on green algae *Chlorella vulgaris* and *Chlorella pyrenoidosa*. *Chinese J. Oceanol. Limnol.* **26**: 380-384.
45. Mostafa N, Omar H, Tan SG, Napis S. 2011. Studies on the genetic variation of the green unicellular alga *Haematococcus pluvialis* (Chlorophyceae) obtained from different geographical locations using ISSR and RAPD molecular marker. *Molecules* **16**: 2599-2608.
46. Varela-Álvarez E, Balau AC, Paulino C, Bercibar E, Pearson GA, Serrão EA. 2018. Isolation and characterization of microsatellite markers for the red alga *Porphyra umbilicalis*. *Plant Genet. Resour. Charact. Util.* **16**: 390-393.
47. Zhao J, Jiang P, Liu Z, Wang J, Cui Y, Qin S. 2011. Genetic variation of *Ulva* (*Enteromorpha*) *prolifera* (Ulvales, Chlorophyta)—the causative species of the green tides in the Yellow Sea, China. *J. Appl. Phycol.* **23**: 227-233.
48. Nagai S, Lian C, Yamaguchi S, Hamaguchi M, Matsuyama Y, Itakura S, *et al.* 2007. Microsatellite markers reveal population genetic structure of the toxic dinoflagellate *Alexandrium tamarense* (Dinophyceae) in Japanese coastal waters. *J. Phycol.* **43**: 43-54.
49. Nagai S, Lian C, Hamaguchi M, Matsuyama Y, Itakura S, Hogetsu T. 2004. Development of microsatellite markers in the toxic dinoflagellate *Alexandrium tamarense* (Dinophyceae). *Mol. Ecol. Notes* **4**: 83-85.
50. Huang H jia, Gan C qiao, Xiao S wen, Zou C, Balamurugan S, Li H ye, *et al.* 2020. Genetic diversity of *Prorocentrum donghaiense* population during bloom in the East China Sea revealed by microsatellite. *J. Appl. Phycol.* **32**: 1851-1862.
51. Xue S, Xuecheng Z, Yunxiang M, Zhenghong S, Song Q. 2006. Identification of phase and sex-related ISSR markers of red alga *Gracilaria lemaneiformis*. *J. Ocean Univ. China* **5**: 82-84.
52. Diaz Martinez S, Boedeker C, Zuccarello GC. 2020. Microsatellite design for species delimitation and insights into ploidy for the Lake Baikal *Cladophoraceae* species flock. *Phycologia* **59**: 355-364.
53. Mauger S, Baud A, Le Corguillé G, Tanguy G, Legeay E, Creis E, *et al.* 2023. Genetic resources of macroalgae: development of an efficient method using microsatellite markers in non-model organisms. *Algal Res.* **75**: 103251.
54. Ren Q, Wang Y chu, Lin Y, Zhen Z, Cui Y, Qin S. 2021. The extremely large chloroplast genome of the green alga *Haematococcus pluvialis*: genome structure, and comparative analysis. *Algal Res.* **56**: 102308.
55. Sun X, Zhang X, Mao Y, Liu J, Sui Z, Qin S, *et al.* 2003. ISSR analysis of marine red algae *Gracilaria* (Rhodophyceae). *High Technol. Lett.* **13**: 89-93.
56. Wang X, Zhao F, Hu Z, Critchley AT, Morrell SL, Duan D. 2008. Inter-simple sequence repeat (ISSR) analysis of genetic variation of *Chondrus crispus* populations from North Atlantic. *Aquat. Bot.* **88**: 154-159.