New Phytologist

Research

Genomic imprints of unparalleled growth

Omer Murik[1,2], Or Geffen[3], Yoram Shotland[4], Noe Fernandez-Pozo[5], Kristian Karsten Ullrich[5,6], Dirk Walther[7], Stefan Andreas Rensing[5,8] and Haim Treves[3,9]

[1]Department of Plant and Environmental Sciences, Hebrew University of Jerusalem, 91904, Jerusalem, Israel; [2]Medical Genetics Institute, Shaare Zedek Medical Center, 93722, Jerusalem, Israel; [3]School of Plant Sciences and Food Security, Tel-Aviv University, 39040, Tel-Aviv, Israel; [4]Chemical Engineering, Shamoon College of Engineering, 84100, Beer-Sheva, Israel; [5]Plant Cell Biology, Department of Biology, University of Marburg, 35037, Marburg, Germany; [6]Max-Planck Institute for Evolutionary Biology, 24306, Plön, Germany; [7]Max-Planck Institute for Molecular Plant Physiology, 14476, Potsdam, Germany; [8]Center for Biological Signaling Studies (BIOSS), University of Freiburg, 79098, Freiburg, Germany; [9]Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau, 67663, Kaiserslautern, Germany

Authors for correspondence:
Haim Treves
Email: h.treves@rptu.de

Omer Murik
Email: omerm@szmc.org.il

Received: *29 August 2023*
Accepted: *31 October 2023*

New Phytologist (2023)
doi: 10.1111/nph.19444

Key words: algal evolution, *Chlorophyta*, codon bias, growth rate, introns.

Summary

- *Chlorella ohadii* was isolated from desert biological soil crusts, one of the harshest habitats on Earth, and is emerging as an exciting new green model for studying growth, photosynthesis and metabolism under a wide range of conditions.
- Here, we compared the genome of *C. ohadii*, the fastest growing alga on record, to that of other green algae, to reveal the genomic imprints empowering its unparalleled growth rate and resistance to various stressors, including extreme illumination. This included the genome of its close relative, but slower growing and photodamage sensitive, *C. sorokiniana* UTEX 1663.
- A larger number of ribosome-encoding genes, high intron abundance, increased codon bias and unique genes potentially involved in metabolic flexibility and resistance to photodamage are all consistent with the faster growth of *C. ohadii*. Some of these characteristics highlight general trends in *Chlorophyta* and *Chlorella* spp. evolution, and others open new broad avenues for mechanistic exploration of their relationship with growth.
- This work entails a unique case study for the genomic adaptations and costs of exceptionally fast growth and sheds light on the genomic signatures of fast growth in photosynthetic cells. It also provides an important resource for future studies leveraging the unique properties of *C. ohadii* for photosynthesis and stress response research alongside their utilization for synthetic biology and biotechnology aims.

Introduction

Organisms inhabiting desert biological soil crust (BSC), such as the green alga *Chlorella ohadii* (Treves *et al.*, 2013), cope with extreme daily and seasonal fluctuations in the ambient conditions. These include extreme temperatures, frequent hydration (by early morning dew) followed by desiccation, large changes in the osmotic potential and very high irradiation of both UV and visible light (Raanan *et al.*, 2016; Weber *et al.*, 2016; Oren *et al.*, 2019).

It is commonly accepted that adaptation to extreme environments is usually accompanied by reduced performance under optimal conditions (referred to here, as conditions in which cultures exhibit maximal growth rate; Brock, 1978; Kushner, 1978). However, apparently, *C. ohadii* does *not* appear to obey this rule. In addition to its ability to cope with ambient conditions that would pose extreme stress for many other organisms, when grown under optimal laboratory settings, *C. ohadii* shows the fastest growth rate ever reported for a photosynthetic eukaryote. Following physiological studies of *C. ohadii* optimal growth temperature (35°C) and illumination (3000 μmol photons m^{-2} s^{-1}), photoautotrophic generation times of 2 h or less have been recorded (Ananyev *et al.*, 2017; Treves *et al.*, 2017), close to that attained by unicellular heterotrophic eukaryotes (Bowler *et al.*, 2010; Flynn *et al.*, 2010; Raven *et al.*, 2013; Flynn & Raven, 2017; Nalley *et al.*, 2018). This raises the obvious questions of what sets the ceiling on the maximal growth of photosynthetic or what distinguishes the fast-growing *C. ohadii* from much slower *Chlorella* sp. and other algae. Despite intensive research, little is known about the cellular features that rate-limit growth, even under what we may consider optimal conditions (Banse, 1976; Beardall *et al.*, 2009; Flynn *et al.*, 2010; Fanesi *et al.*, 2014; Singh & Singh, 2015; Flynn & Raven, 2017).

Two models are leading the discussions in the field. Elser and colleagues suggested that the ribosome content, strongly affected by the availability of phosphate, sets the maximal growth rate (Andersen *et al.*, 2004; Weider *et al.*, 2005). Flynn *et al.* (2010) examined this hypothesis, specifically referring to phytoplankton, and concluded that for the most part, this growth hypothesis is not applicable there. Burnap and others (Burnap, 2015 and references therein) raised the 'proteomic constraint' hypothesis that

suggests that cellular space limitations in conjunction with cell surface-to-volume ratios restrict the maximal growth rate of autotrophic microbes and that the overall content of cellular proteome, and diffusion within the cells, rate-limits growth. Flynn & Raven (2017) concluded that the kinetic properties of ribulose bisphosphate carboxylase/oxygenase (RuBisCO) measured *in vitro* cannot account for carbon production in fast-growing photosynthetic organisms, supporting the notion that the question is still wide open and more research is essential. Our recent study implicated metabolic shifts, indicated by changing from fully photoautotrophic growth to photoheterotrophic under otherwise constant ambient conditions, as an important factor affecting the growth rate of *C. ohadii* (Treves *et al.*, 2017, 2020). The data also implicated polyamines in the metabolic shifts (Treves *et al.*, 2017), but the mechanism involved is yet unknown.

Elegant studies by Pakrasi and colleagues (Mueller *et al.*, 2017; Ungerer *et al.*, 2018a,b) on the fast-growing cyanobacterium *Synechococcus elongatus* UTEX 2973, and genome comparison with that of a slower growing relative, *Synechococcus elongatus* PCC 7942, identified specific alleles that distinguish between the two strains in three genes, atpA, ppnK and rpaA. Replacement of these genes in the strain PCC 7942 with those from UTEX 2973 significantly enhanced the growth rate of the former (Ungerer *et al.*, 2018b). However, the situation in eukaryotes with over 10-fold larger genome, larger cell size, multicompartmental structure and metabolite fluxes between them is far more complex (De Vries & Rensing, 2020).

Photosynthetic organisms undergo photodamage, a process that significantly lowers photosynthetic productivity, particularly when exposed to excess illumination (EIL), that is higher than required to saturate the CO_2 fixation rate (see Aro *et al.*, 1993; Campbell *et al.*, 1996; Keren & Krieger-Liszkay, 2011; Ohad *et al.*, 2011; Cardona *et al.*, 2012; Krupnik *et al.*, 2013; Croce *et al.*, 2018; and references therein). The emerging view from recent studies is that antenna events leading to nonphotochemical quenching (NPQ) play a major role in EIL dissipation in plants and algae (see Baroli *et al.*, 2004; Finazzi *et al.*, 2006; Bailey & Grossman, 2008; Li *et al.*, 2009; Gorbunov *et al.*, 2011; Jahns & Holzwarth, 2012; Erickson *et al.*, 2015; Petroutsos *et al.*, 2016; Foyer *et al.*, 2017; Gollan *et al.*, 2017; Ruban, 2017; Wittkopp *et al.*, 2017; Croce *et al.*, 2018; and references therein). However, we concluded that the 'classical' NPQ plays a minor role in EIL dissipation by *C. ohadii* (Treves *et al.*, 2016). Previous physiological and 'omics studies of *C. ohadii* under EIL, have uncovered a remarkably robust PSII and protective carotenoids accumulation, alongside higher PSI electron transfer rates, increased metabolic capacity supporting rapid poising of redox status, rapid post-translational redox regulation of protein kinases and reactive-oxygen-species and heat-shock responses and novel thylakoid remodeling as factors contributing to this extraordinary resilience (Ananyev *et al.*, 2017; Treves *et al.*, 2017, 2020; Caspy *et al.*, 2021; Levin *et al.*, 2021).

In the past decade, since the publication of the *Chlorella variabilis* draft whole genome (Blanc *et al.*, 2010), the first to be published from the genus *Chlorella*, 16 other *Chlorella* genomes have been published. These genomes have helped researchers to gain new insights into the evolution of the genus, adaptation to environmental cues, lipid biosynthesis and other metabolic processes (Blanc *et al.*, 2010; Juneja *et al.*, 2016; Zuñiga *et al.*, 2016; Arriola *et al.*, 2018; Guarnieri *et al.*, 2018; Cecchin *et al.*, 2019; Petrushin Ivan *et al.*, 2020).

Likely, the ability of *C. ohadii* to withstand the harsh BSC habitat and show unparalleled growth rate under favorable conditions must be imprinted in its genome. Here, we initiate studies to uncover the unique genome characteristics that enable these capabilities. Of the 16 *Chlorella* genomes published, six are of *C. sorokiniana* isolates, mostly isolated from wastewaters or warm surface waters and sequenced due to the growing biotechnological interest in algal productivity (De-Bashan *et al.*, 2008; Arriola *et al.*, 2018; Hovde *et al.*, 2018; Wu *et al.*, 2019). We used the fact that *C. ohadii* is relatively close to *C. sorokiniana* UTEX 1663, isolated from warm surface freshwater (Starr & Zeikus, 1993; Running *et al.*, 1994; > 92% amino acid sequence identity in 5216 single-copy orthologues, 100% chloroplast genome sequence identity, see Treves *et al.*, 2013), and the detailed phenotypic data related to growth and photosynthesis of the two species to conduct a powerful comparative genomics analysis aiming to find the imprints of these phenotypes in the *C. ohadii* genome. Our analyses revealed which functional aspects associated with *C. ohadii* stress response were under stronger selective pressure and pointed to several general features of its genome which were likely optimized to support its remarkable growth and performance.

Materials and Methods

Growth conditions, DNA and RNA extraction and sequencing

For nucleic acid extraction, cultures of *Chlorella ohadii* and *Chlorella sorokiniana* were grown in Erlenmeyer flasks, on a shaker, in Tris–acetate-phosphate buffer (TAP) or acetate-free (TP) media at 30°C, 100 rpm and continuous cool-white illumination of 100 μmol photons $m^{-2}\,s^{-1}$ as in Treves *et al.* (2016). The cultures were axenic, as validated with light microscopy and LB plating/incubation before harvesting, for DNA extraction, or before the first sampling point in each replica, for RNA extraction. Contamination control was performed using a light microscope, Eclipse E200 (Nikon, Melville, NY, USA).

DNA extraction was performed using a modified phenol–chloroform method (Sambrook & Russell, 2006) with the addition of a preliminary bead-beating step (≤ 106 μm glass beads (Sigma), 2×30 s, 6 m s^{-1}). To obtain high-molecular DNA fragments required for SMRT sequencing, we used cut tips and avoided vigorous shaking at any step downstream to the cell breakage. Genomic DNA was sequenced using Roche-454 (9×), Illumina mate-pair (48×, Illumina Next-Seq 500) and paired-end (42×) reads (HUJI Genome Center) and 4 SMRT cells (35×, PacBio RS II; BaseClear BV, Leiden, the Netherlands).

For RNA extraction, cells were harvested from 20 ml cultures by centrifugation (5 min, 2000 ***g***, 4°C). Cell pellets were treated with Tri-Reagent (Molecular Research Center, Cincinnati, OH, USA) according to the provided protocol and combined with bead-beating

(see above) in the Tri-Reagent solution. The RNA pellets were dried at 55°C, dissolved in 20–40 µl DNase/RNase/Protease free water (Sigma) and then stored at −80°C. DNase (Turbo DNA-*free*; Ambion) treatment was carried out twice on the RNA samples according to the manufacturer's instructions. RNA libraries were prepared using ScriptSeq Complete Kit (Plant Leaf; Epicenter, San Diego, CA, USA) according to the manufacturer's protocol. Libraries were sequenced on HiSeq 2500 (Illumina) 100 PE run at the Technion Genome Center, Haifa, Israel.

Genome assembly and annotation

Raw sequencing reads were filtered using TRIMMOMATIC with default settings (Bolger *et al.*, 2014). Quality-filtered reads were assembled using PacificBiosciences assembly tool to yield a 57 Mbp genome with 486 scaffolds for *C. ohadii* (N50 = 179.2 kbp, Table 1) and polished using the short-read data with PILON (v.1.17; Walker *et al.*, 2014). The *C. sorokiniana UTEX* 1663 sequencing data were assembled using spades resulting in a 57 Mbp genome with 6933 scaffolds. The qualities of the assemblies were assessed with QUAST (Gurevich *et al.*, 2013). Gene models for the genome assemblies were predicted using AUGUSTUS (v.3.2; Stanke *et al.*, 2006). For *C. ohadii*, RNA-seq data were used to train AUGUSTUS. The quality of the gene predictions was assessed with BUSCO (Seppey *et al.*, 2019; v 5.1.2) using the Chlorophyta dataset (odb 10). The resulting 11 047 genes were functionally annotated using BLAST2GO (Gotz *et al.*, 2008) and EGGNOG-MAPPER (Huerta-Cepas *et al.*, 2017). GO-slim annotation was performed using BLAST2GO. Enrichment of Gene Ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) terms for specific gene sets was analyzed using CLUSTERPROFILER (Yu *et al.*, 2012). Prediction of subcellular localization was performed using CHLOROP-1.1 (Emanuelsson *et al.*, 1999), TARGETP-2.0 (Emanuelsson *et al.*, 2007) and LOCALIZER (Sperschneider *et al.*, 2017). Mitochondrial and plastid genome sequences were filtered out from the genome assembly and were annotated independently using GESEQ (Tillich *et al.*, 2017), and their maps were created using OGDRAW (Greiner *et al.*, 2019).

RNA-seq analysis

Between 16 and 24 million, 100 bp, paired-end reads per sample were quality checked using FASTQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and adapter trimming or low-quality filtering were done using Trimmomatic (Bolger *et al.*, 2014). Filtered reads were mapped to the *C. ohadii* genome assembly with TOPHAT2 (Kim *et al.*, 2013). Counting and normalization of mapped reads and analysis of differential expression were done using the cufflinks suite (Trapnell *et al.*, 2012).

Phylogenetic analysis

Genome and protein sequences of representative cyanobacteria, algae and plants were obtained from PHYTOZOME (https://phytozome.jgi.doe.gov) or REFSEQ. For orthologs detection, each *C. ohadii* protein was analyzed with BLASTP (Altschul

Table 1 *Ohadii* and other green algae genomes' general statistics.

Assembly	*Ohadii*	*Chlamydomonas reinhardtii*	*Coccomyxa subellipsoidea C-169*	*Chromochloris zofingiensis*	*Chlorella sorokiniana UTEX1602*	*Chlorella sorokiniana UTEX1663*	*Chlorella variabilis*	*Auxenochlorella protothecoides*
Number of nucleotides (bases)	57 091 696	111 100 715	49 193 098	60 363 192	59 566 223	57 197 880	46 159 512	22 929 133
Number of scaffolds	486	54	47	220	159	6933	414	406
Largest scaffold (bases)	1007 468	9730 733	4035 500	4698 708	4192 086	103 718	3119 887	1144 347
N50 (bases)	328 978 ($n = 53$)	7783 580 ($n = 9$)	1959 569 ($n = 7$)	3274 445 ($n = 8$)	2592 956 ($n = 9$)	15 198 ($n = 1130$)	1469 606 ($n = 12$)	285 534 ($n = 24$)
G + C content, %	63.80	64.08	52.92	50.51	63.85	63.61	67.14	63
Total number of genes	11 047	17 741	9629 (9851)	15 369	10 384	12 044	9780	7039
Gene density	83.6%	81.1%	76.5%	63.54%	83%	73.45%	62.02%	87.9%
Average no. of introns per transcript	12.01	8.61	8.06	4.94	11.54	9.19	8.3	5.72
Accession	This study	GCA_000002595.3	GCF_000258705.1	Roth *et al.* (2017)	GCA_002245835.2	This study	GCF_000147415.1	Gao *et al.* (2014)

Genome characteristics of *Ohadii* and seven other green algae. For comparison, gene model prediction was performed using the same parameters on all genomes, without using hints from RNA-seq data. This may result in different sets of predicted genes as found in public databases for the same genomes.

et al., 1990) against all proteins of the other species, with an *e*-value cutoff of 1E-5, identity of 30% and bitscore of 50. d*N*/d*S* values for each orthologs pair were determined by codeml (Yang, 1997) based on coding sequence alignments. Orthologous groups were analyzed using ORTHOFINDER, v.2.5.4 (Emms & Kelly, 2019). For whole genome phylogenetic analysis, we first used BUSCO (Seppey *et al.*, 2019; v 5.1.2) to filter complete single-copy conserved genes for each of 16 *Chlorella* or other chlorophyte species from the corresponding genome sequences. Amino acid sequences of 495 single-copy BUSCO proteins shared by all 16 species were concatenated and aligned with MAFFT website phylogenetic tool. Best fitting substitution model for phylogenetic analysis was determined with MODELTEST-NG v.0.1.7 (Darriba *et al.*, 2020). Phylogenetic trees were constructed with either RAXML-NG v.1.1 (Kozlov *et al.*, 2019) using LG + G8 + F model and 500 bootstrap replicates or MRBAYES v.3.2 (Ronquist *et al.*, 2012) using the mixed amino acid rate matrices preset. 500 000 generations were computed; the standard deviation of split frequencies converged at 10 000. Both trees showed essentially the same topology. Genes were considered as horizontally transferred from bacteria if their blast hits were only to bacterial sequences or the blast score was higher by 20% or more for a bacterial target over the best algal target and were closer to the bacterial branch in a phylogenetic tree using FASTTREE (v.2.2.10; Price *et al.*, 2010).

rDNA gene copy number

For copy number estimation of rDNA genes in different species, raw genomic DNA sequencing reads were mapped to the genome assembly of each species. The copy number was estimated by dividing the average coverage within the ribosomal genes region by the overall average coverage. For *C. ohadii*, rDNA gene copy number was also measured. DNA was extracted as above. Quantitative PCR (qPCR) was performed on a Rotor-Gene 6000 system (Corbett Research, Cambridge, UK) using PowerUp™ SYBR® Green Master Mix (Thermo). rDNA copy number was calculated as fold change over the single-copy gene PsbQ according to the $\Delta\Delta C_t$ method (Livak & Schmittgen, 2001). Primers used in this assay:

rDNA_F: GCTGGAATTACCGCGGCT;
rDNA_R: CGG CTACCACATCCAAGGAA.
PsbQ_F: CGACATCATCTACGAGGCCC;
PsbQ_R: TTCTCAGCCTCCTTGATGCG.

Codon usage bias analysis

Codon usage bias was estimated by computing for every amino acid residue type, aa, the codon usage entropy (CUE) as

$$\mathrm{CUE}_{\mathrm{aa}} = -\sum_{i=1}^{N_{\mathrm{aa}}} \frac{f_i}{T} \log\left(\frac{f_i}{T}\right) \qquad \text{Eqn 1}$$

where N_{aa} is the number of codons encoding amino acid residue type aa, f_i denotes the observed count of codon i, $T = \sum_i^{N_{\mathrm{aa}}} f_i$, is the total number of codons counted for aa, and applying the natural logarithm.

CUE is maximal if all possible codons for a given amino acid residue type are used at the same relative frequency (even usage of codons), while it is minimized when usage is uneven and zero when only one codon of all possible codons is used. Thus, low values of CUE signify high codon usage bias. Codon usage table and codon adaptation index (CAI) were calculated using cusp and cai from EMBOSS tools (v.6.6.0).

Results

General characteristics of the genome

We sequenced and annotated the genome of *C. ohadii* at high depth, combining long- and short-read sequencing technologies at 90× and 35× depth, respectively (accession no. PRJNA573576). The assembly of all sequencing data resulted in a 57.1Mbp draft genome organized in 486 scaffolds with an N50 of 328 977 (Table 1). We also assembled full circular mitochondrial and plastid genomes (52 and 109 kbp, respectively). For comparison, we sequenced the slower growing and far more sensitive to EIL, *C. sorokiniana* UTEX 1663 (accession no. PRJNA573576, 57.2 Mb; Treves *et al.*, 2016). The assemblies were screened for bacterial and vector contaminations, using KRAKEN2 (Wood *et al.*, 2019) and UniVec (https://www.ncbi.nlm.nih.gov/tools/vecscreen/), respectively. No contig was found to be contaminant of nonalgal origin. Gene modeling using AUGUSTUS (Stanke *et al.*, 2008) trained on RNA-seq data from *C. ohadii* (Treves *et al.*, 2020) predicted 11 047 genes, with a BUSCO completeness of 97%. The vast majority (96.7%, FPKM > 3) of the predicted genes were supported by RNA-seq reads (Supporting Information Table S1). Over 98.6% of the predicted genes have a complete CDS with both start and stop codons (Table S2). Table S2 also provides the number/fraction of genes recognized by BLAST, Eggnog or InterProScan and those annotated by Gene Ontology (GO) or KEGG. The annotations of the various *C. ohadii* genes are provided in Table S3. We compared the genome and annotation characteristics to publicly available information on the genomes of several other green algae: *Chlamydomonas reinhardtii* (Merchant *et al.*, 2007), *Coccomyxa subellipsoidea* C-169 (Peng *et al.*, 2016), *Chromochloris zofingiensis* (previously named *Chlorella zofingiensis*; Roth *et al.*, 2017) and *C. sorokiniana* UTEX 1602 (Arriola *et al.*, 2018; Table 1). Among the *Chlorella* spp., the genomes of the fast-growing *C. ohadii* and *C. sorokiniana* UTEX 1602 are relatively densely packed; genes occupy over 83% of the nuclear genomic DNA sequences, but only 63.5% and 62.1% in the slower growing organisms *C. zofingiensis* and *C. variabilis*, respectively (see below). Noticeably, the different gene densities coincide with an increased coverage assigned for introns in *C. ohadii* (see below).

Phylogenetic analysis of the genome

Given the harsh conditions in its natural habitat on one hand and its immense growth capabilities on the other, it is likely that *C. ohadii* underwent many changes during its evolution. Phylogenetic analysis (Figs 1a, S1A) confirmed that *C. ohadii* is indeed a

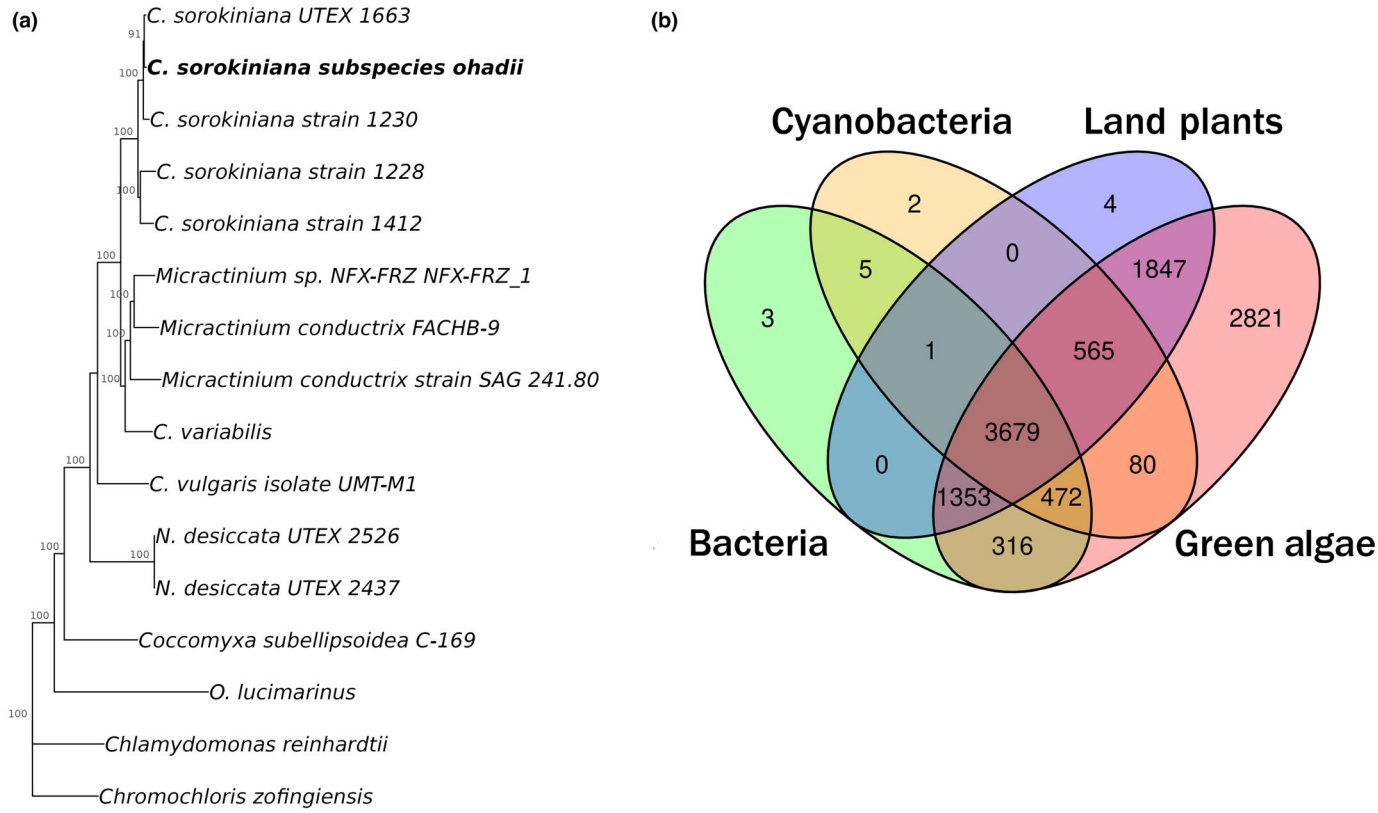

Fig. 1 *Ohadii* genome phylogenetic analysis. (a) Maximum likelihood phylogenetic tree based on 495 single-copy Busco genes present in all species. Confidence values, resulting from 500 bootstrap replicates, are shown at the nodes. The tree is rooted using *C. reinhardtii* as outgroup. (b) Venn diagram of *Ohadii* genes with homologs in bacteria, cyanobacteria, plants and green algae, not including *C. sorokiniana* UTEX1663. The number on the bottom right indicates the number of *Ohadii* genes with no detected homolog in any of these taxonomic groups. For genome accessions used, see Supporting Information Table S12.

Chlorella sp. and indicated that it belongs to a clade of *C. sorokiniana*. We therefore refer from this point to *C. ohadii* as *Chlorella sorokiniana* subspecies *ohadii* (hereafter *Ohadii*, for convenience). Genome-wide gene comparison was performed between the 11 407 *Ohadii* predicted proteins and those of the other taxonomic groups (Fig. 1b) and other *Chlorella* species (Fig. S1B). This further confirmed that *Ohadii* is closely related to *C. sorokiniana* sharing far more genes with strains of this species than with other *Chlorella* species and was more related in terms of gene content to the strain UTEX1663 (Fig. S1B). Given the documented relative heat- and light-tolerance and fast growth of *C. sorokiniana* strains (Kessler, 1972), and their growth history in warm surface or wastewater, this similarity may point to these habitats as a potential source for desert BSC colonization by *Ohadii*.

To discern the functions, where sequence alterations occurred during its evolution that thereby distinguish *Ohadii* from other *Chlorella* spp., we analyzed the genetic variations in coding genes as nonsynonymous (K_a) to synonymous (K_s) variations between each pair of single-copy orthologues. The largest values of $K_a : K_s$ in *Ohadii* were observed in proteins engaged in photosynthesis, energy and carbohydrate metabolism, DNA metabolism, sensing and responding to environmental cues and stress, both on average and compared with other *Chlorella* spp. (Fig. 2, the comparisons with specific *Chlorella* spp. are shown in Fig. S2). These findings are in line with the functions where *Ohadii* demonstrates improved performance – photosynthesis, growth rate and ability to withstand harsh conditions.

The gene content analysis (Fig. 1b) showed that *Ohadii* possesses genes present in bacteria, cyanobacteria and land plant, which were undetected in other green algae. Of particular interest are six genes found only in *Ohadii* and bacteria, but not in any other photosynthetic eukaryote from the green lineage (Viridiplantae) as yet (Table S4; Fig. S3). Another 11 genes show significantly higher similarity to bacterial genes than to any other eukaryotic gene. Those, presumably acquired by lateral gene transfer, have been recognized so far in *Ohadii* only. A subset of five genes show higher similarity to nonphotosynthetic eukaryotes than to bacteria. While these genes might have been acquired vertically by *Ohadii* and lost in all other green photosynthetic eukaryotes, it is more likely that they were acquired through a separate HGT event. Some of these genes are discussed with respect to their specific functions and possible role in *Ohadii's* growth in Notes S1 and S2. Although an increase in DNA mobility through transposable elements could potentially explain much of the *Ohadii* genome architecture, the content of mobile and transposable elements in the genome showed no significant difference from that observed in other algae (Table S5).

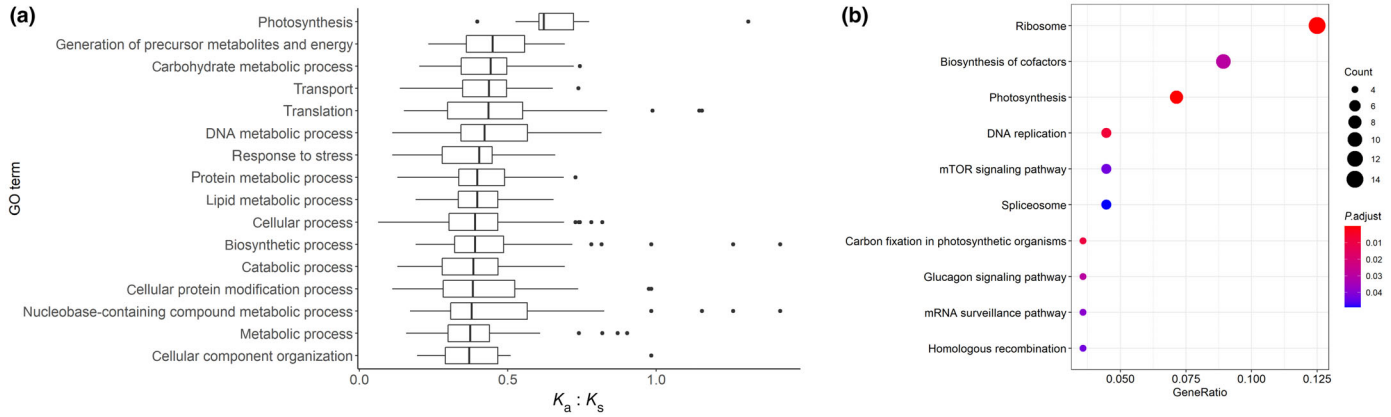

Fig. 2 Rate of selection pressure on genes as estimated by the ratio of nonsynonymous to synonymous ($K_a : K_s$) changes between *Ohadii* genes and their orthologs in four related *Chlorella* species. (a) Boxplot representation of $K_a : K_s$ values according to the biological process Gene Ontology (GO) of all *Ohadii* proteins with detectable orthologues in other *Chlorella* species (*A. pyrenoidosa*, *C. variabilis*, *C. sorokiniana* and *C. subellipsoidosa*). Error bars mark 95% confidence intervals. (b) Enriched KEGG pathways among proteins with predicted positive selected proteins ($K_a : K_s > 1$).

Abundant ribosomal DNA copies may support *Ohadii* rapid growth

As indicated in the Introduction section, ribosome counts were proposed as an important parameter affecting growth rate. Bacterial growth rate is strongly influenced by the copy number of ribosome-encoding genes and the fraction of functioning ribosomes. Mutants of *Escherichia coli* K-12 with reduced number of rRNA operons, or partial inhibition of ribosome function by various drugs, showed a significantly lower maximum growth rate than the control (Levin *et al.*, 2017). To the best of our knowledge, manipulation of ribosomal gene copy numbers was not yet accomplished in eukaryotes, but modifications of their transcription or function by various means severely inhibit growth (Cheng *et al.*, 2019). This correlation between growth rate and number of functional ribosomes is consistent with the proposed impact of the ribosome abundance per cell on the growth rate, though likely not across temperature ranges (Hessen *et al.*, 2008; Moody *et al.*, 2017). In order to estimate the copy number of ribosomal genes, we calculated the ratio of the coverage of DNA sequencing reads on the 18S ribosomal gene region and the average coverage of reads on the genome excluding ribosomal regions. This analysis showed a much larger number of rDNA encoding genes in *Ohadii* (54.5) than in several other slower growing *Chlorella* spp. like *C. sorokiniana*, *C. variabilis* and *Micractinium conductrix* (Fig. S4). We also performed qPCR on genomic DNA, targeting the 18S rDNA gene and the single-copy gene PsbQ (see the Materials and Methods section), revealing 55-fold change between the abundances of these two genomic regions and validating the *in silico* estimation. The lowest rDNA copy number was observed in the temperate soil alga *C. zofingiensis* (10) and the photobiont *C. variabilis* (7), where the generation times are 40–50-fold (Roth *et al.*, 2019) and 5–10-fold (Cheng *et al.*, 2015; Kodama & Fujishima, 2015) longer than in *Ohadii*, respectively. The ribosome quota per cell in *Chlamydomonas* (133) is larger, but this may be related to its cell volume, *c.* 50-fold larger than *Ohadii*. Raven and colleagues (Raven *et al.*, 2019) proposed that larger cells may need to produce more rRNA than smaller ones as the concentration of rRNA likely limits the abundance of ribosomes and thereby protein synthesis and growth. Noticeably, among the 30 genes, whose transcript abundance is the highest in *Ohadii* (Table S1), 21 encode ribosomal genes, two code for elongation factors and two code for RNA processing.

To further examine the role of this potential elevated translation capacity in *Ohadii* response to EIL, we explored gene expression patterns of ribosomal protein-encoding genes during a shift from LL to EIL. Strikingly, we observed significant upregulation of 16 and 5 large and small plastid ribosome subunits encoding genes, respectively. This massive response coincided with respective upregulation of 10 and 7 large and small cytosolic ribosome subunits encoding genes, but interestingly, with no response or even mild downregulation in expressed mitochondrial ribosome subunits encoding genes (Table S6). This confirms the importance of a robust and dynamic protein synthesis machinery for *Ohadii* response to abiotic changes, especially in the unusual diurnal cycle prevailing in the desert BSC (Treves *et al.*, 2020).

Intron abundance correlates with growth rate

The majority of the genes (98.3%), identified in *Ohadii*, contain multiple exons (Tables 1, S2, S6). The average number of *I*ntrons *P*er *G*ene (hereafter IPG, Table S7) is largest in *Ohadii* among all sequenced chlorophytes analyzed here (Table 1), including other *Chlorella* sp. concurring with the large genomic space occupied by introns (24.6% vs 16.2% and 1.7% in *C. reinhardtii* and *C. variabilis*, respectively) and an opposite trend for intergenic space (17.5% vs 34.6% and 38.3%). Interestingly, the largest abundance of introns is observed in genes that *Ohadii* shares with plants and bacteria (Fig. 3a), whereas the IPG values are lowest in the orphan genes. The largest number of introns was observed in *Ohadii*'s gene g2009 encoding one of its five type I polyketide synthases. Among the 193 genes lacking introns, 35 encode histone components, most others are either not recognized by BLAST

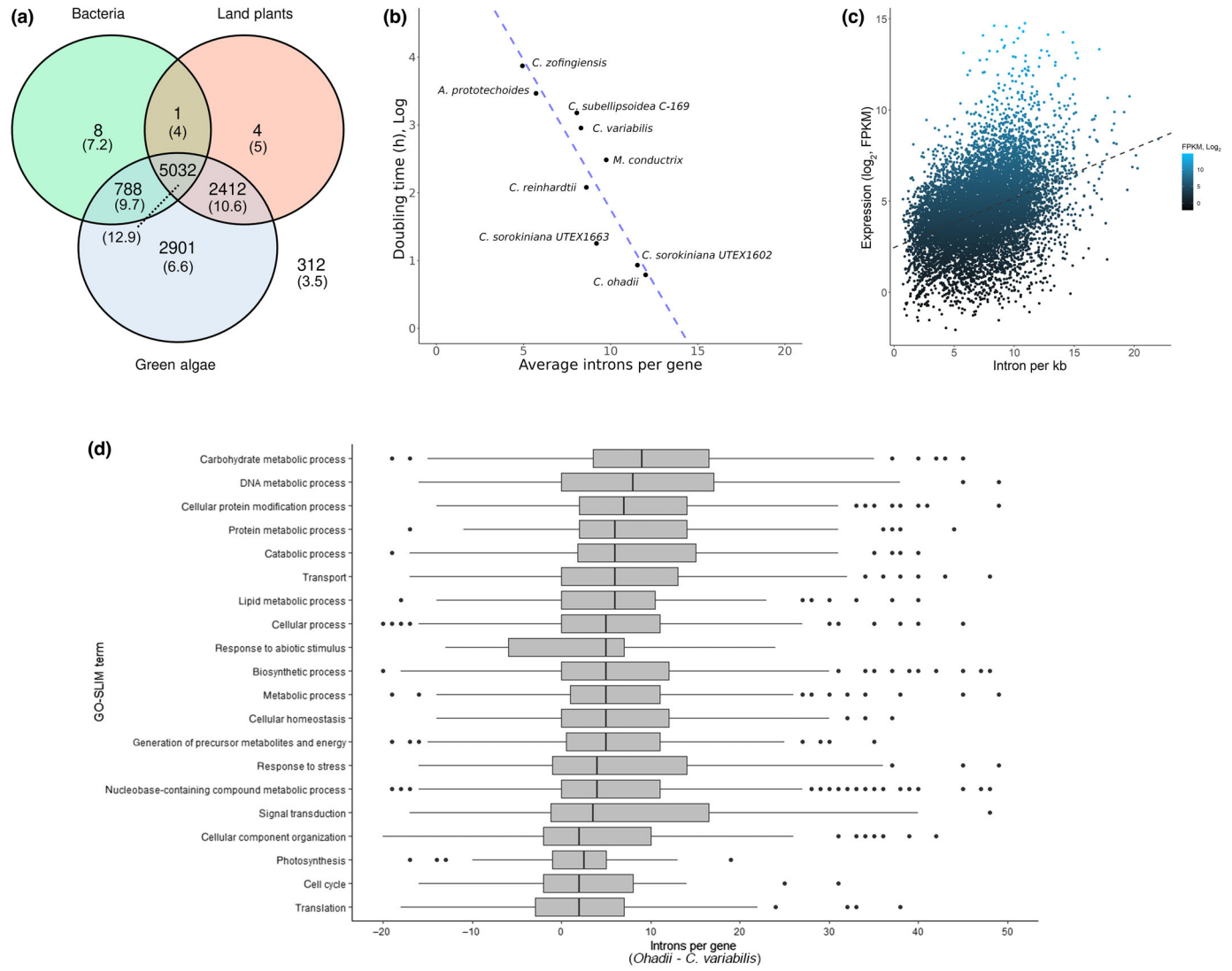

Fig. 3 Expansion of introns in *Ohadii*'s genes. (a) The average number of introns, in parenthesis, of each group of *Ohadii* genes based on the presence of orthologues in bacteria, plants and other green algae. The number in the bottom right indicates the number of *Ohadii* genes with no detectable orthologue in any of these lineages. (b) Increasing growth rate, represented by doubling time, with the increase in the average number of introns per gene in *Chlorella* spp. and *C. reinhardtii*. Maximal growth rates under species-specific optimal conditions are presented. Growth data for *Ohadii*, *C. sorokiniana* UTEX1602, *C. sorokiniana* UTEX1663, *C. variabilis* and *C. reinhardtii* were obtained in this study or extracted from (Treves *et al*., 2013, 2017). Optimal growth data for *C. subellipsoidea*, *A. protothecoides* and *C. zofingiensis* were extracted from Liu *et al*. (2022), Yu *et al*. (2019), Pikoli *et al*. (2019), Arriola *et al*. (2018) and Chowdhary *et al*. (2022), respectively. Blue dashed line represents a linear regression trendline. (c) Significant increase in transcript abundance, expressed as $\log_2$-FPKM values of autotrophically grown *Ohadii* cells, with the increase in intron density (number per kbp transcript). Each dot represents a single gene. *P*-value = 1.8×10^{-241}, Pearson correlation. Dashed line represents a linear regression trendline. (d) Boxplot representing the difference in the number of introns per gene between *Ohadii* genes and their orthologue in *C. variabilis*, based on different GO terms. Error bars mark 95% confidence intervals.

or are of unknown function (Table S7). Noticeably, our analysis identified at least 120 highly transcribed (FPKM > 500) intron-localized short (< 40 bp) noncoding sequences in at least one of the growth conditions examined (Treves *et al*., 2020), but their role awaits further experimental verification. The exceptionally high IPG and genome occupation suggested they may play regulatory roles and affect various cell functions, including growth rate (Baier *et al*., 2020).

Fig. 3(b) shows a strong positive correlation ($R^2 = 0.83$, *P*-value = 0.0006893) between the growth rate in various green algae (presented as generation time) and average IPG. As an example, in the slowest growing alga examined here, *C. zofingiensis* (Roth *et al*., 2017), isolated from temperate forest soils (Dönz, 1934), the growth rate is *c.* 40–50-fold slower and the IPG is about one third of that of *Ohadii*, isolated from the harsh desert crusts environment, where the limited daily time interval enabling photoautotrophic activity favors fast growth. This is consistent with the suggestion that the presence of introns accelerates the expression of genes without serving as a binding site for transcription regulators, a phenomenon termed 'intron-mediated enhancement' (Shaul, 2017; Lasin *et al*., 2020; Back & Walther, 2021; and references therein). Furthermore, a recent

study showed that in addition to the first intron, those located downstream also affect the expression of rbcS (the nuclear gene encoding the small subunit of RuBisCO) in *C. reinhardtii* (Baier *et al.*, 2018). Supporting this notion, our analysis found a correlation between intron density per kbp and mRNA abundance (Table S1) in cultures grown under autotrophic ($R = 0.39$, P-value $= 1.8 \times 10^{-241}$, Pearson correlation, Fig. 3c) or mixotrophic ($r = 0.25$, P-value $= 2.5 \times 10^{-168}$, Fig. S5) conditions (Treves *et al.*, 2020). This indicates that introns may have cumulative effect on transcript abundance of various *Ohadii* genes; however, the mechanisms involved should be further explored (see the Discussion section).

To assess the predictive value of intron density compared with growth parameters like medium and light, thereby gauging their relevance with regard to determining gene expression levels, we performed a three-way ANOVA with interaction effects (Table S8; Fig. S6). Intron density (number of introns per kb mRNA) proved to be most informative (largest F-value), with higher densities associated with higher gene expression levels, followed by 'medium', with both parameters also showing a significant interaction effect. By contrast, the factor 'light' was not observed to be significantly related to gene expression levels. Of note, when defining intron density as the number of introns per kb genomic sequence, 'medium' was found to be most important ($F = 1000.1$), followed by 'intron density' ($F = 258.1$) and 'light' ($F = 2.4$). Thus, intron length seems to also be important with regard to the influence on gene expression levels, with longer introns found associated with increased gene expression levels (The Pearson correlation coefficient of log(IDmRNA : IDgenomic) and log(FPKM + 0.1) was determined as +0.3 ($P < 2.2 \times 10^{-16}$, ID-intron density). The larger the ratio IDmRNA/IDgenomic, the longer the intron sequences).

In view of the large IPG in genes shared with core plant genes (Fig. 3a), we examined intron abundance in various gene ontologies (GO) compared with that observed in *C. variabilis* (Fig. 3d). The latter was chosen since the growth rate and resistance to abiotic stress are far less robust than in *Ohadii*. In the GO-slim criteria presented here, encompassing a large variety of biological activities, the IPG values are higher in *Ohadii*, most pronounced are those involved in carbohydrate, DNA, protein and lipid metabolisms, catabolism, response to abiotic stimulus and to stress and others, which was also found to be the case when comparing *Ohadii* to the significantly slower growing *C. subellipsoidosa* (Fig. S7A). Essentially, these are the same GO criteria, where the $K_a : K_s$ ratio is high (Figs 2, S2), and are GO functions where *Ohadii* excels, enabling its unparalleled performance growth and stress tolerance performance. To complement this analysis, we have also examined relative intron abundance among different GO categories between *Ohadii* with *C. sorokiniana* 1663 (Fig. S7B), representing an intermediate case in terms of growth rate and light resistance (Treves *et al.*, 2016). Median IPG values of the latter were found to reside between those of *Ohadii* and *C. variabilis* (Fig. S7C). However, while ΔIPG between *Ohadii* with *C. sorokiniana* 1663 were expectedly smaller, several of the top ΔIPG scoring GO criteria were similar (Fig. S7B). Of particular interest are the specific genes involved. Many of the IPG-rich genes are upregulated under, for instance, EIL and are affected by the trophic conditions (Table S1). But apart from the dependence of transcript abundance on the IPG and the strong correlation between average IPG and growth rate (Fig. 3b), at this time we cannot provide clear evidence for the dependence of growth or stress resilience on IPG abundance.

Ohadii genes exhibit stronger codon bias

To examine the potential role of codon usage and possible optimization therein, the Codon Adaptation Index (CAI) was calculated for each gene of *Ohadii* genome and related to its transcription under normal mixotrophic growth (Fig. 4a). The two parameters were significant, albeit moderately correlated ($r = 0.1$, P-value $= 2.05 \times 10^{-24}$), in agreement with the widely accepted view that the higher the codon adaptation, the faster the rate of transcription (Ullrich *et al.*, 2015; Hiss *et al.*, 2017). The large variation in FPKM values around the high density region of the plots at the high CAI range (Fig. 4a) suggests, as expected, that other parameters also affect the transcription rate. Based on the codon counts, we computed codon usage entropies (CUE, Eqn 1) for all 18 amino acid residue types encoded by more than one codon (all 20 biogenic amino acids except methionine and tryptophan) to estimate codon usage bias in several *Chlorella* spp. and *C. reinhardtii*. Relative to *Ohadii*, all other species exhibit increased CUE-values, albeit significance was established for *C. sorokiniana* and *C. variabilis* only, but not for *C. reinhardtii* (Fig. 4b). Thus, of the four considered species, *Ohadii* is characterized by the smallest CUE-values and thus exhibits high codon usage bias relative to the other three species. Taken together, these data show that *Ohadii* genes possess more adapted codon usage than other algae, supporting the notion that it can translate, and hence grow, faster.

Gene families

An analysis of gene families within the *Ohadii* genome using OrthoFinder revealed that 4231 genes belong to 1638 orthogroups consisting of two to 17 paralogs (Table S9). Thirty-seven of these orthogroups are unique to *Ohadii* and another 206 are shared only with *C. sorokiniana* UTEX1663. Another subset of 170 orthogroups has significantly higher number of gene copies in *Ohadii* and *C. sorokiniana* than in other algal genomes (t-test, Benjamini–Hochberg adjusted for FDR, P-value < 0.05, Table S9). From *Ohadii*'s unique/enriched orthogroups, several may support the phenotypes associated with fast metabolism, stress response and acclimation to EIL, such as Flavin monoamine oxidases (with only one or two copies found in other algal genomes compared with 10 copies in *Ohadii*, six of which upregulated under EIL), involved in several mitochondrial functions, including polyamine metabolism; Diacylglycerol acyltransferase, a key enzyme in triglyceride synthesis; and low CO_2 inducible protein, which may be involved in the response to limiting CO_2 and EIL and in which all members were upregulated in response to EIL (Table S1). Also pronounced are the 14 genes encoding prohibitin in *Ohadii* compared with a single gene in *C.*

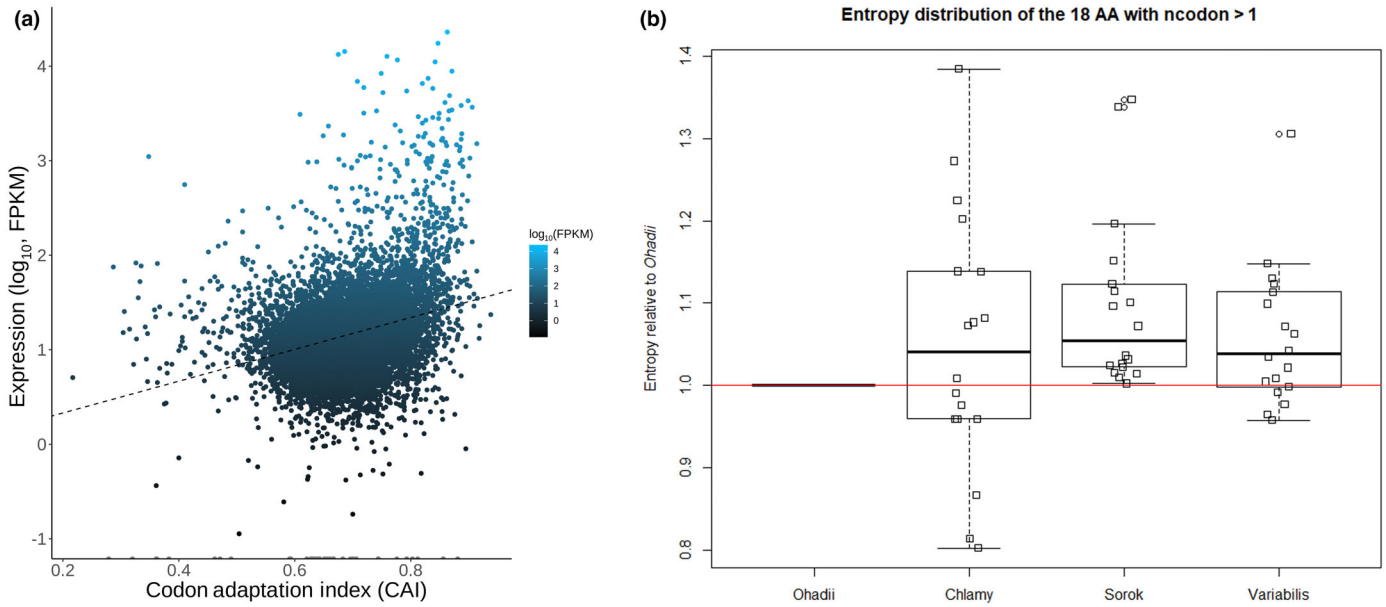

Fig. 4 Codon usage and entropy in *Ohadii* genome. (a) CAI vs FPKM in mixotrophically grown *Ohadii* cultures. Codon adaptation index for each transcript was compared with its FPKM value (log_{10}). *P*-value = 2.05×10^{-24}, Pearson's correlation. (b) Codon usage entropy distribution of algal species relative to *Ohadii*. Boxplots capture the overall distribution over all amino acid residues with individual amino acid residue types (one-letter code) added with some horizontal jitter for better visual separation. Considered are the 18 amino acid residue types encoded by more than one codon with their associated value of codon usage entropy (CUE, Eqn 1) expressed relative to its value determined in *Ohadii*. Natural log of CUE-value ratios was taken to render ratios symmetric to ratio 1 (zero in log space) signifying no difference and denoted by the blue horizontal line. The associated *t*-test *P*-values of deviation from zero across all 18 amino acid residue types are as follows: *C. reinhardtii*: 0.27 (0.27), *C. sorokiniana*: 0.000768 (0.0023), *C. variabilis*: 0.00861 (0.013) with values in parenthesis corresponding to the Benjamini–Hochberg-adjusted values (three species/tests).

reinhardtii and six to seven in *C. sorokiniana*. Prohibitins, encoded by PHB genes in various eukaryotes, have been implicated in numerous functions (Tatsuta *et al.*, 2004; Van Aken *et al.*, 2016; Chen *et al.*, 2019), many of them related to mitochondrial homeostasis and response to stress (Hernando-Rodríguez & Artal-Sanz, 2018; Signorile *et al.*, 2019). Likely, this may be due to the super-complex formed by PHBs in the inner mitochondrial membrane (Tatsuta *et al.*, 2004). Transcript analyses showed that seven of the *Ohadii*'s PHB genes were upregulated when *Ohadii* cells were exposed to EIL as revealed by RNA-seq (Tables S1, S9B). Among their many functions, PHBs regulate the activity of mitochondrial alternative oxidase (AOX) in *Arabidopsis* (Van Aken *et al.*, 2016). Thus, prohibitins may be involved in the 'burning' of excess redox in EIL-exposed *Ohadii* cells and in the metabolic shift from photoautotrophic (net O_2 evolution and alkalization due to CO_2 removal) to photoheterotrophic (net O_2 uptake and acidification due to CO_2 evolution (Treves *et al.*, 2017)) and may therefore play a role both in *Ohadii* fast growth and its resistance to EIL. Smaller but significant expansion was also observed in orthogroups associated with starch synthesis and degradation, like chloroplast beta-amylase and granule-bound starch synthase (*t*-test, Benjamini–Hochberg adjusted for FDR, *P*-value = 0.043 and 0.018, respectively, Table S9), with two copies of the latter being upregulated under EIL. Starch as a spillover metabolite may increase up to 60-fold in EIL-grown *Ohadii* cultures, and its metabolism has been shown to play a major role in these cells' metabolic response to excess light (Treves *et al.*, 2020).

Orphan genes

Orphan genes are genes found uniquely in a single species without any detectable homolog in any other known lineage (Tautz & Domazet-Lošo, 2011). We considered a gene as an orphan if it had no significant BLAST hit in the NCBI's NR database or any hit in the *C. sorokiniana* UTEX1663 genome published here. Accordingly, the *Ohadii* genome encodes 179 orphan genes to date, in addition to another 133 genes only found in *Ohadii* and *C. sorokiniana* UTEX1663 (Fig. S1B; Table S10). Naturally, as most annotation tools are using inheritance of annotation based on homology, there are limited experimental data assisting the analysis of those genes. However, with the use of expression profiles under various conditions, domain identification, subcellular targeting and short stretches of homology to known proteins, we further characterized several orphan genes. Below, we briefly present a few cases representing various physiological aspects of *Ohadii*'s biology.

As indicated below, nuclear-encoded genes are likely involved in the photodamage resistance of *Ohadii*. Six orphan genes g788, g3410, g3651, g6364, g6906 and g10379 are strongly upregulated in *Ohadii* cells exposed to EIL (Table S1) and (Treves *et al.*, 2020) and possess a putative chloroplast directing transit peptide suggesting chloroplastic location of the mature protein (see the Materials and Methods section).

One example is g2474, which is strongly upregulated in *Ohadii* cells exposed to EIL in the presence or absence of acetate in their growth media (Table S1). A blast analysis revealed a

similarity to a *C. sorokiniana* UTEX 1602 gene annotated as signal transduction histidine kinase (sequence ID: PRW56131.1). However, all other BLAST hits did not support this annotation. On the contrary, protein domain analysis using InterProScan identified a strong carbohydrate/starch binding domain in its N terminus, suggesting that it may sense the carbohydrate status of the cells under various light intensities, possibly playing a role in the metabolic flexibility enabling a fast shift from photoautotrophic to photoheterotrophic metabolism in *Ohadii* cultures (see below). In this respect, we also found that C status of *Ohadii* cells can affect growth through cell cycle regulation, including upregulation of chloroplast division negative regulator genes under C limitation (see Notes S3).

Another example for orphan genes, which may be associated with *Ohadii* growth, is the case of the two dual-specificity tyrosine-regulated kinases (DYRK) in *Ohadii* g9828 and g3579. In g9828, the 270 amino acid residues (AA) at the N terminus and the 160 amino acids on the C terminus, encode parts of a DYRK protein. The central 750 AA region of g9828 shows no significant similarity to other *Chlorella* spp., but InterProScan identified an Aurora-type serine/threonine kinase motif, missing in g3579. Aurora-type proteins were implicated in the regulation of chromosomal alignment and segregation during cell division (Borisa & Bhatt, 2017), suggesting they may be involved in growth regulation in a fast-growing organism such as *Ohadii*, or may aid homologous recombination and DNA repair.

Transcription factors and regulators

Transcription-associated proteins (TAPs) comprise transcription factors (TFs, acting in sequence-specific manner, typically by binding to *cis*-regulatory elements) and transcriptional regulators (TRs, acting on chromatin or via protein–protein interaction). We classified all *Ohadii* TAPs into 122 families and subfamilies based on a domain-based rule set (Lang *et al.*, 2010; Wilhelmsson *et al.*, 2017). We compared this genome-wide classification with eight other Chlorophyta and six Prasinophyta, as well as five land plants, five Rhodophyta, the streptophyte alga *Klebsormidium* and the glaucophyte *Cyanophora* (Table S11, Sheet 1). All proteins, in which a domain was found, are listed in Table S11, Sheet 2, along with their family assignment.

The *Ohadii* genome encodes 398 proteins (Table S11) likely involved in transcriptional regulation, which is the largest number among the seven studied *Chlorella* spp. genomes and higher than the average (338 ± 89) for green algae. Total number of TAPs for *Ohadii* is found between those reported from most algae and Zygnematophyceae algae, which have been recently found to be the closest relatives of land plants (Feng *et al.*, 2023). The fraction of such proteins encoded by the genome is also comparatively high (3.47% vs 3.0%, the average for green algae).

Ohadii lacks eight families of transcription factors (TFs) and five transcriptional regulators (TR) that are present in at least one other green alga. Like most other published *Chlorella* genomes, it lacks MADS TF and Sigma70-like TR family members (Table S11) that are otherwise generally encoded by green algal genomes. *Ohadii* shares with several other *Chlorella* genomes an unusually high number of SBP TFs (34–37) compared with an average of 20 in green algae.

In terms of gene families that are potentially expanded, *Ohadii* encodes 23 C3H (Cys3His zinc finger) TFs, the highest number observed in green algae (average 14.8). Similarly, there are four HD (homeodomain) TFs while the average is 1.3 (0–2) and 11 Jumonji_other TRs (average 7.4). For all the above TAPs, numbers for *Ohadii* fall within the range of Zygnematophyceae algae (Feng *et al.*, 2023). The Jumonji_other TR represent proteins with a JmjC domain that are not part of the PKDM7 subfamily. They might act as protein hydroxylases that catalyze histone modification (Trewick *et al.*, 2005). There are 15 RWP-RK (RKD) TFs, higher than the average in green algae (8.6), only superseded by *C. reinhardtii* (16) and rivalled by *C. sorokiniana* (14). These TFs are involved in processes such as mating type determination and microbial symbiosis signaling (Hernández-Reyes *et al.*, 2022).

The set of differentially expressed genes (DEGs) of three comparisons (mixotrophically grown cells as control vs 15/120-min EIL treatment – TAPHL15 and TAPHL120, respectively, and photoautotrophically grown cells following 120-min EIL treatment – TPHL120) was scanned for presence of TAPs (Table S11, Sheets 3–5; plus denotes upregulation under treatment as compared to control, minus denotes downregulation under treatment). Interestingly, in genes differentially regulated under high and low light intensities, there are C3H and SBP TFs under the up- and downregulated genes and Jumonji_other TRs among the downregulated genes (Table S11). Hence, three out of five families that harbor unusually many members might be instrumental for the regulatory changes required by different light regimes. We have also examined transcriptional response of *Ohadii* to the above EIL treatments for known light-responsive TAPs from other green algae. Here, orthologs of the conserved transcriptional factor *CONSTANS* (Table S11) and the subunit *CONSTITUTIVE PHOTOMORPHOGENIC 1* (COP1) of its associated E3 ubiquitin ligase complex (Table S1) were upregulated under TAPHL15 and TPHL120. *CONSTANS*-like genes are key TF in the light response of both the model *C. reinhardtii* (Gabilly *et al.*, 2019) and *Mesotaenium endlicherianum*, a member of the Zygnematophyceae algae (Dadras *et al.*, 2023). In the latter, the light-responsive *GOLDEN2-LIKE 1* (GLK1) TF, acts as a major regulatory hub of gene expression, but despite being conserved among Zygnematophyceae and land plants, no GLK1 homologs could be found in *C. reinhardtii* (Riaño-Pachón *et al.*, 2008) or *Ohadii* (this work). We have nevertheless identified a member of the G2-like family (to which GLK1 belongs) that was upregulated in *Ohadii* under TAPHL15 and TPHL120 (Table S11) and may play a similar role in these algae. Additional work focusing on *Ohadii* gene expression regulatory network under EIL or other conditions may help reveal both novel and conserved transcriptional response hubs in this extremophile alga's response to various abiotic stresses.

Discussion

Accumulating mechanistic studies on nonmodel organisms from extreme environments (Ohad *et al.*, 2010; Garcia-Pichel

et al., 2013; Couradeau *et al.*, 2016; Guida & Garcia-Pichel, 2016; Treves *et al.*, 2016; Koyama *et al.*, 2018; Kalra *et al.*, 2020) support the notion that unique capabilities are more likely to manifest in organisms facing adversity. While this can be anticipated for stress response and resistance properties, we did not expect to isolate a remarkably fast-growing alga (when provided with optimal conditions) from the harsh conditions of the Negev desert. Additional physiological and omics studies of *Ohadii* (Treves *et al.*, 2016, 2020) point to the unusual diurnal cycle in the BSC as a major driving force for its increased photosynthetic and growth performance. Specifically, the limited daily time interval at which sufficient humidity and illumination coincide to allow photoautotrophic activity (Ohad *et al.*, 2010) was likely the driver of *Ohadii*'s high-basal photosynthetic and metabolic capacity ('high-light ready cell' (Treves *et al.*, 2020)), and its relatively higher investment of photosynthate into growth (Treves *et al.*, 2022), possibly aimed at achieving a positive biomass balance during a narrow 'window of opportunity'. Our functional evolutionary analysis of the *Ohadii* genome (Fig. 2) demonstrates that processes serving to mitigate these conditions were under stronger selective pressure than others, reflecting *Ohadii* growth history.

To keep pace with these rates of biomass accumulation, the cellular machinery controlling upstream gene expression must be tightly regulated and tuned, as summarized in Fig. 5. Indeed, the *Ohadii* genome encodes more proteins involved in transcriptional regulation than any other studied *Chlorella* spp. (Table S11), including the highest number among all green algae of C3H zinc

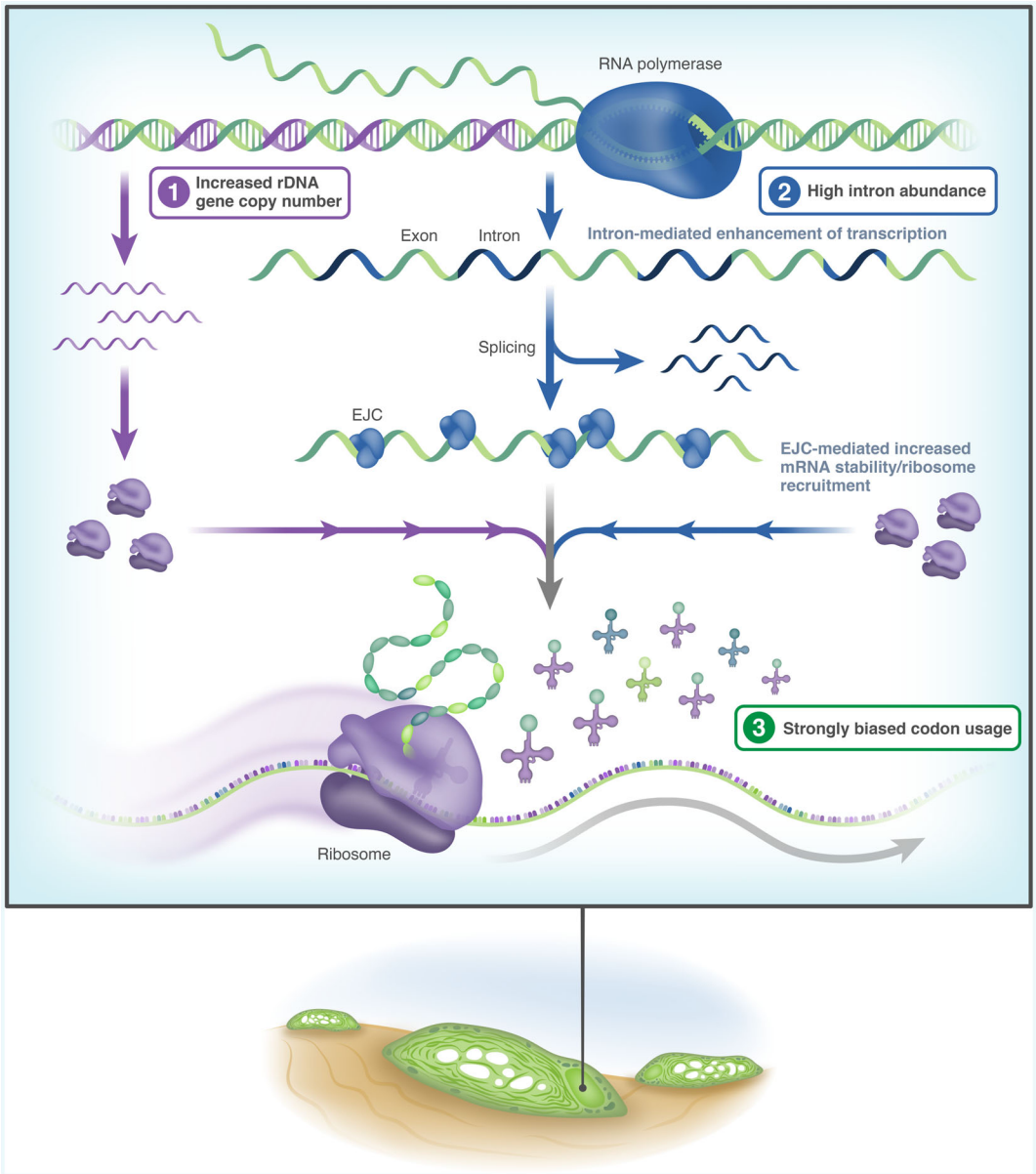

Fig. 5 Schematic illustration of *Ohadii* genome-imprinted optimizations of transcription and translation machineries. The *Ohadii* genome features several adaptations supporting its fast growth rate, including higher (1) rDNA gene copy number (2) intron abundance and (3) stronger codon bias compared with other green algae.

finger TF, which play important roles in the regulation of plant growth, developmental processes and environmental responses (reviewed in Liu *et al.*, 2020). We also observed a much higher copy number of rDNA encoding genes compared with several other *Chlorella* spp. and, when cell volume is taken into account, to *C. reinhardtii* (Figs 5, S4). This coincides with ribosomal genes occupying the top positions of most highly expressed genes (Table S1). In this respect, we have also observed a relatively large variance between *Ohadii* variants and their homologs of plastid and mitochondrial genes encoding organellar RNA polymerase subunits and ribosomal proteins (see Notes S4, S5). Beyond capacity, another optimizing factor for the function of translation is the relatively higher codon usage bias in *Ohadii* compared with other *Chlorella* spp. and *C. reinhardtii* (Fig. 4b), which, as demonstrated by the correlation between codon bias and expression (Fig. 4a), likely contributes to the concerted operation of this organism transcription and translation machineries (Fig. 5).

While most of these features represent a somewhat expected optimization within a previously described pattern, the relationship between intron abundance, expression and growth rate, together with the large intron- and small intergenic-allocated genomic space, is perhaps the most extraordinary aspect of the *Ohadii* genome (Fig. 5). This is further highlighted by the lack of evidence for increased splicing variants in *Ohadii* gene expression data despite higher intron abundance, underlining our gap of mechanistic understanding of the correlation between intron abundance and growth rate of the studied *Chlorella* spp. (Fig. 3b), or transcript abundance in *Ohadii* (Figs 3c, S4).

Intron abundance and specifically intron density had been reported to positively correlate with mRNA abundance both in human and in plants (Wang *et al.*, 2007; Duan *et al.*, 2013). Introns were also implicated in the increased efficiency of mRNA translation or export to the cytosol through the function of exon junction complex (EJC) proteins (reviewed by Shaul, 2017). In the absence of a stable transformation system that will allow differential expression analysis of intron-content on the same genes, we could not conclude whether introns control transcription or the rate of mRNA decay, or both in the transcription/splicing machinery of an *Ohadii* cell or in other *Chlorella* spp. Finally, recent work performed in *Saccharomyces cerevisiae* revealed an additional role for introns in highly expressed genes. Bonnet and colleagues (Bonnet *et al.*, 2017) showed that, through a mechanism conserved from yeast to human, introns are involved in the protection against DNA damage accumulation by preventing R-loop formation. Recently, seminal studies by Weigel and colleagues (Monroe *et al.*, 2022) found mutation rates to be higher in genes lacking introns and lower in genes with more and longer introns, emphasizing their role in mutation bias evolution. As the need to maintain genomic stability/integrity increases with growth rate (i.e. through more doubling events and higher gene expression), this can act as a selective force favoring elevated IPG numbers and intron length along the studied *Chlorella* spp. growth rate continuum. In any case, the mere functional similarity between high $K_a : K_s$ (Fig. 2) and IPG (Fig. 3d) groups, alongside the implication of IPG as a major determinant of gene expression (Figs 3c, S5, S7; Table S8), strongly suggests that the abundance of introns and their particular location in specific genes supports *Ohadii's* unique stress response and growth properties (Fig. 5).

Finally, both orphan and potentially horizontally transferred genes in *Ohadii* genome are likely to contribute to its growth performance and resilience, based on their expression patterns and annotation (Tables S4, S10). A more thorough account of specific genes is provided in Notes S2, S6–S8 and includes expressed bacterial-originated genes playing a role in polyamine metabolism, which has been shown to regulate *Ohadii* growth (Treves *et al.*, 2017), in CO_2 concentration mechanism operation and in cellular redox control, which is a major factor in this alga resistance to EIL (Treves *et al.*, 2020). However, validation of these predicted roles must await the development of a genetic system for this remarkable green alga.

Acknowledgements

This research was supported by Israeli Science Foundation grant 1697/22 (supporting HT and OG) and by the Junta de Andalucia Emergia program EMERGIA20_00286 (supporting NFP). We thank the Alon fellowship for the financial support of HT. We thank Prof. Aaron Kaplan from HUJI for constructive discussions regarding *Ohadii* physiology interaction with its genome. We thank the Technion Genome Center (Haifa, Israel) for their support during the transcriptome sequencing and analysis.

Competing interests

None declared.

Author contributions

HT, OM and SAR designed the research. HT, OG and OM performed research. HT, OM, NF-P, YS, KKU and DW analyzed data. HT, OM, DW, NF-P and SAR wrote the article. All authors contributed to editing the article.

ORCID

Noe Fernandez-Pozo https://orcid.org/0000-0002-6489-5566
Omer Murik https://orcid.org/0000-0002-3093-6980
Stefan Andreas Rensing https://orcid.org/0000-0002-0225-873X
Yoram Shotland https://orcid.org/0000-0002-0403-6356
Haim Treves https://orcid.org/0000-0002-3431-6965
Kristian Karsten Ullrich https://orcid.org/0000-0003-4308-9626
Dirk Walther https://orcid.org/0000-0002-5755-9265

Data availability

The genomes of *Ohadii* and *C. sorokiniana* have deposited in NCBI, accessions JADXDR000000000 and JACXBM000000000,

respectively. Raw sequencing data are available in the Sequence Read Archive (SRA) under BioProject PRJNA573576. The *Ohadii* Genome DB was implemented using easyGDB (Fernandez-Pozo & Bombarely, 2022). *Ohadii* annotations and sequences, together with bioinformatics tools such as Blast, genome browser, annotation search, gene expression atlas and sequence and annotation downloading for a list of genes, are available at https://cohadiigenomedb.url.com. The alignment used to construct the phylogenetic trees is available at doi: 10.6084/m9.figshare.23634423.v1. The authentic *C. sorokiniana* subspecies *ohadii* is deposited at Culture Collection of Algae at University Göttingen (SAG, Göttingen, Germany) under strain number SAG 2657 and is available by request from SAG (*epsag@uni-goettingen.de*) for noncommercial academic research.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410.

Ananyev G, Gates C, Kaplan A, Dismukes GC. 2017. Photosystem II-cyclic electron flow powers exceptional photoprotection and record growth in the microalga *Chlorella ohadii*. *Biochimica et Biophysica Acta-Bioenergetics* **1858**: 873–883.

Andersen T, Elser JJ, Hessen DO. 2004. Stoichiometry and population dynamics. *Ecology Letters* 7: 884–900.

Aro E-M, Virgin I, Andersson B. 1993. Photoinhibition of photosystem II. Inactivation, protein damage and turnover. *Biochimica et Biophysica Acta* **1143**: 113–134.

Arriola MB, Velmurugan N, Zhang Y, Plunkett MH, Hondzo H, Barney BM. 2018. Genome sequences of *Chlorella sorokiniana* UTEX 1602 and *Micractinium conductrix* SAG 241.80: implications to maltose excretion by a green alga. *The Plant Journal* **93**: 566–586.

Back G, Walther D. 2021. Identification of *cis*-regulatory motifs in first introns and the prediction of intron-mediated enhancement of gene expression in *Arabidopsis thaliana*. *BMC Genomics* **22**: 390.

Baier T, Jacobebbinghaus N, Einhaus A, Lauersen KJ, Kruse O. 2020. Introns mediate post-transcriptional enhancement of nuclear gene expression in the green microalga *Chlamydomonas reinhardtii*. *PLoS Genetics* **16**: e1008944.

Baier T, Wichmann J, Kruse O, Lauersen KJ. 2018. Intron-containing algal transgenes mediate efficient recombinant gene expression in the green microalga *Chlamydomonas reinhardtii*. *Nucleic Acids Research* **46**: 6909–6919.

Bailey S, Grossman A. 2008. Photoprotection in Cyanobacteria: regulation of light harvesting. *Photochemistry and Photobiology* **84**: 1410–1420.

Banse K. 1976. Rates of growth, respiration and photosynthesis of unicellular algae as related to cell size – a review. *Journal of Phycology* **12**: 135–140.

Baroli I, Gutman BL, Ledford HK, Shin JW, Chin BL, Havaux M, Niyogi KK. 2004. Photo-oxidative stress in a xanthophyll-deficient mutant of *Chlamydomonas*. *The Journal of Biological Chemistry* **279**: 6337–6344.

Beardall J, Allen D, Bragg J, Finkel ZV, Flynn KJ, Quigg A, Rees T, Alwyn V, Richardson A, Raven JA. 2009. Allometry and stoichiometry of unicellular, colonial and multicellular phytoplankton. *New Phytologist* **181**: 295–309.

Blanc G, Duncan G, Agarkova I, Borodovsky M, Gurnon J, Kuo A, Lindquist E, Lucas S, Pangilinan J, Polle J *et al.* 2010. The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. *Plant Cell* **22**: 2943–2955.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Bonnet A, Grosso AR, Elkaoutari A, Coleno E, Presle A, Sridhara SC, Janbon G, Géli V, De Almeida SF, Palancade B. 2017. Introns protect eukaryotic genomes from transcription-associated genetic instability. *Molecular Cell* **67**: 608–621.

Borisa AC, Bhatt HG. 2017. A comprehensive review on Aurora kinase: small molecule inhibitors and clinical trial studies. *European Journal of Medicinal Chemistry* **140**: 1–19.

Bowler C, DE Martino A, Falciatore A. 2010. Diatom cell division in an environmental context. *Current Opinion in Plant Biology* **13**: 623–630.

Brock TD. 1978. *Thermophilic microorganisms and life at high temperatures*. New York, NY, USA: Springer-Verlag.

Burnap RL. 2015. Systems and photosystems: cellular limits of autotrophic productivity in Cyanobacteria. *Frontiers in Bioengineering and Biotechnology* **3**: 1.

Campbell D, Clarke AK, Gustafsson P, Oquist G. 1996. D1 exchange and the Photosystem II repair cycle in the cyanobacterium Synechococcus. *Plant Science* **115**: 183–190.

Cardona T, Sedoud A, Cox N, Rutherford AW. 2012. Charge separation in Photosystem II: a comparative and evolutionary overview. *Biochimica et Biophysica Acta (BBA) – Bioenergetics* **1817**: 26–43.

Caspy I, Neumann E, Fadeeva M, Liveanu V, Savitsky A, Frank A, Kalisman YL, Shkolnisky Y, Murik O, Treves H *et al.* 2021. Cryo-EM photosystem I structure reveals adaptation mechanisms to extreme high light in *Chlorella ohadii*. *Nature Plants* 7: 1314–1322.

Cecchin M, Marcolungo L, Rossato M, Girolomoni L, Cosentino E, Cuine S, Li-Beisson Y, Delledonne M, Ballottari M. 2019. Chlorella vulgaris genome assembly and annotation reveals the molecular basis for metabolic acclimation to high light conditions. *The Plant Journal* **100**: 1289–1305.

Chen K, Guo T, Li X-M, Yang Y-B, Dong N-Q, Shi C-L, Ye W-W, Shan J-X, Lin H-X. 2019. NAL8 encodes a prohibitin that contributes to leaf and spikelet development by regulating mitochondria and chloroplasts stability in rice. *BMC Plant Biology* **19**: 395.

Cheng Y-S, Labavitch J, Vandergheynst JS. 2015. Organic and inorganic nitrogen impact *Chlorella variabilis* productivity and host quality for viral production and cell lysis. *Applied Biochemistry and Biotechnology* **176**: 467–479.

Cheng Z, Mugler CF, Keskin A, Hodapp S, Chan LY, Weis K, Mertins P, Regev A, Jovanovic M, Brar GA. 2019. Small and large ribosomal subunit deficiencies lead to distinct gene expression signatures that reflect cellular growth rate. *Molecular Cell* **73**: 36–47.

Chowdhary AK, Kishi M, Toda T. 2022. Enhanced growth of *Chromochloris zofingiensis* through the transition of nutritional modes. *Algal Research* **65**: 102723.

Couradeau E, Karaoz U, Lim HC, Nunes DA Rocha U, Northen T, Brodie E, Garcia-Pichel F. 2016. Bacteria increase arid-land soil surface temperature through the production of sunscreens. *Nature Communications* 7: 10373.

Croce R, Van Grondelle R, Van Amerongen H, Van Stokkum I. 2018. *Light harvesting in photosynthesis*. Boca Raton, FL, USA: CRC Press.

Dadras A, Fürst-Jansen JMR, Darienko T, Krone D, Scholz P, Sun S, Herrfurth C, Rieseberg TP, Irisarri I, Steinkamp R *et al.* 2023. Environmental gradients reveal stress hubs pre-dating plant terrestrialization. *Nature Plants* **9**: 1419–1438.

Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. 2020. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Molecular Biology and Evolution* **37**: 291–294.

De Vries J, Rensing SA. 2020. Gene gains paved the path to land. *Nature Plants* **6**: 7–8.

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. *Bioresource Technology* **99**: 4980–4989.

Dönz O. 1934. *Chlorella zofingiensis*, eine neue Bodenalge. *Berichte der Deutschen Botanischen Gesellschaft* **43**: 127–131.

Duan J, Shi J, Ge X, Dölken L, Moy W, He D, Shi S, Sanders AR, Ross J, Gejman PV. 2013. Genome-wide survey of interindividual differences of RNA stability in human lymphoblastoid cell lines. *Scientific Reports* **3**: 1318.

Emanuelsson O, Brunak S, Von Heijne G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* **2**: 953–971.

Emanuelsson O, Nielsen H, Von Heijne G. 1999. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Science* **8**: 978–984.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**: 238.

Erickson E, Wakao S, Niyogi KK. 2015. Light stress and photoprotection in *Chlamydomonas reinhardtii*. *The Plant Journal* **82**: 449–465.

Fanesi A, Raven JA, Giordano M. 2014. Growth rate affects the responses of the green alga *Tetraselmis suecica* to external perturbations. *Plant, Cell & Environment* **37**: 512–519.

Feng X, Zheng J, Irisarri I, Yu H, Zheng B, Ali Z, De Vries S, Keller J, Fürst-Jansen JMR, Dadras A *et al.* 2023. Chromosome-level genomes of multicellular algal sisters to land plants illuminate signaling network evolution. *bioRxiv*. doi: 10.1101/2023.01.31.526407.

Fernandez-Pozo N, Bombarely A. 2022. EasyGDB: a low-maintenance and highly customizable system to develop genomics portals. *Bioinformatics* **38**: 4048–4050.

Finazzi G, Johnson GN, DALL'osto L, Zito F, Bonente G, Bassi R, Wollman FA. 2006. Nonphotochemical quenching of chlorophyll fluorescence in *Chlamydomonas reinhardtii*. *Biochemistry* **45**: 1490–1498.

Flynn KJ, Raven JA. 2017. What is the limit for photoautotrophic plankton growth rates? *Journal of Plankton Research* **39**: 13–22.

Flynn KJ, Raven JA, Rees TAV, Finkel Z, Quigg A, Beardall J. 2010. Is the growth rate hypothesis applicable to microalgae? *Journal of Phycology* **46**: 1–12.

Foyer CH, Ruban AV, Noctor G. 2017. Viewing oxidative stress through the lens of oxidative signalling rather than damage. *Biochemical Journal* **474**: 877–883.

Gabilly ST, Baker CR, Wakao S, Crisanto T, Guan K, Bi K, Guiet E, Guadagno CR, Niyogi KK. 2019. Regulation of photoprotection gene expression in Chlamydomonas by a putative E3 ubiquitin ligase complex and a homolog of CONSTANS. *Proceedings of the National Academy of Sciences, USA* **116**: 17556–17562.

Gao C, Wang Y, Shen Y, Dong Y, Xi H, Junbiao D, Qingyu W. 2014. Oil accumulation mechanisms of the oleaginous microalga *Chlorella protothecoides* revealed through its genome, transcriptomes, and proteomes. *BMC Genomics* **15**: 582.

Garcia-Pichel F, Loza V, Marusenko Y, Mateo P, Potrafka Ruth M. 2013. Temperature drives the continental-scale distribution of key microbes in topsoil communities. *Science* **340**: 1574–1577.

Gollan PJ, Lima-Melo Y, Tiwari A, Tikkanen M, Aro E-M. 2017. Interaction between photosynthetic electron transport and chloroplast sinks triggers protection and signalling important for plant productivity. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **372**: 20160390.

Gorbunov MY, Kuzminov FI, Fadeev VV, Kim JD, Falkowski PG. 2011. A kinetic model of non-photochemical quenching in cyanobacteria. *Biochimica et Biophysica Acta (BBA) – Bioenergetics* **1807**: 1591–1599.

Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2Go suite. *Nucleic Acids Research* **36**: 3420–3435.

Greiner S, Lehwark P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) v.1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* **47**: W59–W64.

Guarnieri MT, Levering J, Henard CA, Boore JL, Betenbaugh MJ, Zengler K, Knoshaug EP. 2018. Genome sequence of the oleaginous green alga, *Chlorella vulgaris* UTEX 395. *Frontiers in Bioengineering and Biotechnology* **6**: 37.

Guida BS, Garcia-Pichel F. 2016. Extreme cellular adaptations and cell differentiation required by a cyanobacterium for carbonate excavation. *Proceedings of the National Academy of Sciences, USA* **113**: 5712–5717.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. Quast: quality assessment tool for genome assemblies. *Bioinformatics* **29**: 1072–1075.

Hernández-Reyes C, Lichtenberg E, Keller J, Delaux PM, Ott T, Schenk ST. 2022. NIN-like proteins: interesting players in rhizobia-induced nitrate signaling response during interaction with non-legume host *Arabidopsis thaliana*. *Molecular Plant–Microbe Interactions* **35**: 230–243.

Hernando-Rodríguez B, Artal-Sanz M. 2018. Mitochondrial quality control mechanisms and the PHB (Prohibitin) complex. *Cell* **7**: 238.

Hessen DO, Ventura M, Elser JJ. 2008. Do phosphorus requirements for RNA limit genome size in crustacean zooplankton? *Genome* **51**: 685–691.

Hiss M, Schneider L, Grosche C, Barth MA, Neu C, Symeonidi A, Ullrich KK, Perroud P-F, Schallenberg-Rüdinger M, Rensing SA. 2017. Combination of the endogenous lhcsr1 promoter and codon usage optimization boosts protein expression in the moss *Physcomitrella patens*. *Frontiers in Plant Science* **8**: 1842.

Hovde BT, Hanschen ER, Steadman Tyler CR, Lo C-C, Kunde Y, Davenport K, Daligault H, Msanne J, Canny S, Eyun S-I *et al.* 2018. Genomic characterization reveals significant divergence within *Chlorella sorokiniana* (Chlorellales, Trebouxiophyceae). *Algal Research* **35**: 449–461.

Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Von Mering C, Bork P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Molecular Biology and Evolution* **34**: 2115–2122.

Jahns P, Holzwarth AR. 2012. The role of the xanthophyll cycle and of lutein in photoprotection of photosystem II. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1817**: 182–193.

Juneja A, Chaplen FWR, Murthy GS. 2016. Genome scale metabolic reconstruction of *Chlorella variabilis* for exploring its metabolic potential for biofuels. *Bioresource Technology* **213**: 103–110.

Kalra I, Wang X, Cvetkovska M, Jeong J, Mchargue W, Zhang R, Hüner N, Yuan JS, Morgan-Kiss R. 2020. *Chlamydomonas* sp. UWO 241 exhibits high cyclic electron flow and rewired metabolism under high salinity. *Plant Physiology* **183**: 588–601.

Keren N, Krieger-Liszkay A. 2011. Photoinhibition: molecular mechanisms and physiological significance. *Physiologia Plantarum* **142**: 1–5.

Kessler E. 1972. Physiologische und biochemische Beiträge zur Taxonomie der Gattung Chlorella. *Archiv für Mikrobiologie* **87**: 243–248.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**: R36.

Kodama Y, Fujishima M. 2015. Differences in infectivity between endosymbiotic *Chlorella variabilis* cultivated outside host *Paramecium bursaria* for 50 years and those immediately isolated from host cells after one year of reendosymbiosis. *Biology Open* **5**: 55–61.

Koyama A, Harlow B, Kuske CR, Belnap J, Evans RD. 2018. Plant and microbial biomarkers suggest mechanisms of soil organic carbon accumulation in a Mojave Desert ecosystem under elevated CO_2. *Soil Biology and Biochemistry* **120**: 48–57.

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**: 4453–4455.

Krupnik T, Kotabová E, Van Bezouwen LS, Mazur R, Garstka M, Nixon PJ, Barber J, Kaňa R, Boekema EJ, Kargul J. 2013. A reaction centre-dependent photoprotection mechanism in a highly robust photosystem II from an extremophilic red alga *Cyanidioschyzon merolae*. *The Journal of Biological Chemistry* **288**: 23529–23542.

Kushner DJ. 1978. *Microbial life in extreme environment*. London, UK: Academic Press.

Lang D, Weiche B, Timmerhaus G, Richardt S, Riano-Pachon DM, Correa LG, Reski R, Mueller-Roeber B, Rensing SA. 2010. Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity. *Genome Biology and Evolution* **2**: 488–503.

Lasin P, Weise A, Reinders A, Ward JM. 2020. Arabidopsis sucrose transporter AtSuc1 introns act as strong enhancers of expression. *Plant and Cell Physiology* **61**: 1054–1063.

Levin BR, Mccall IC, Perrot V, Weiss H, Ovesepian A, Baquero F. 2017. A numbers game: ribosome densities, bacterial growth, and antibiotic-mediated stasis and death. *mBio* **8**: e02253-16.

Levin G, Kulikovsky S, Liveanu V, Eichenbaum B, Meir A, Isaacson T, Tadmor Y, Adir N, Schuster G. 2021. The desert green algae *Chlorella ohadii* thrives at excessively high light intensities by exceptionally enhancing the mechanisms that protect photosynthesis from photoinhibition. *The Plant Journal* **106**: 1260–1277.

Li ZR, Wakao S, Fischer BB, Niyogi KK. 2009. Sensing and responding to excess light. *Annual Review of Plant Biology* **60**: 239–260.

Liu C, Xu X, Kan J, Cheng ZM, Chang Y, Lin J, Li H. 2020. Genome-wide analysis of the C3H zinc finger family reveals its functions in salt stress responses of *Pyrus betulaefolia*. *PeerJ* **8**: e9328.

Liu Y, Wei D, Chen W. 2022. Oleaginous microalga *Coccomyxa subellipsoidea* as a highly effective cell factory for CO(2) fixation and high-protein biomass production by optimal supply of inorganic carbon and nitrogen. *Frontiers in Bioengineering and Biotechnology* **10**: 921024.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* **25**: 402–408.

Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, Terry A, Salamov A, Fritz-Laylin LK, Marechal-Drouard L *et al.* 2007. The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science* **318**: 245–250.

Monroe JG, Srikant T, Carbonell-Bejerano P, Becker C, Lensink M, Exposito-Alonso M, Klein M, Hildebrandt J, Neumann M, Kliebenstein D *et al.* 2022. Mutation bias reflects natural selection in *Arabidopsis thaliana*. *Nature* **602**: 101–105.

Moody EK, Rugenski AT, Sabo JL, Turner BL, Elser JJ. 2017. Does the growth rate hypothesis apply across temperatures? Variation in the growth rate and body phosphorus of neotropical benthic grazers. *Frontiers in Environmental Science* **5**.

Mueller TJ, Ungerer JL, Pakrasi HB, Maranas CD. 2017. Identifying the metabolic differences of a fast-growth phenotype in Synechococcus UTEX 2973. *Scientific Reports* 7: 41569.

Nalley JO, O'donnell DR, Litchman E. 2018. Temperature effects on growth rates and fatty acid content in freshwater algae and cyanobacteria. *Algal Research* **35**: 500–507.

Ohad I, Berg A, Berkowicz SM, Kaplan A, Keren N. 2011. Photoinactivation of photosystem II: is there more than one way to skin a cat? *Physiologia Plantarum* **142**: 79–86.

Ohad I, Raanan H, Keren N, Tchernov D, Kaplan A. 2010. Light-induced changes within Photosystem II protects Microcoleus sp. in biological desert sand crusts against excess light. *PLoS ONE* **5**: e11000.

Oren N, Raanan H, Kedem I, Turjeman A, Bronstein M, Kaplan A, Murik O. 2019. Desert cyanobacteria prepare in advance for dehydration and rewetting: the role of light and temperature sensing. *Molecular Ecology* **28**: 2305–2320.

Peng H, Wei D, Chen G, Chen F. 2016. Transcriptome analysis reveals global regulation in response to CO_2 supplementation in oleaginous microalga Coccomyxa subellipsoidea C-169. *Biotechnology for Biofuels* **9**: 151.

Petroutsos D, Tokutsu R, Maruyama S, Flori S, Greiner A, Magneschi L, Cusant L, Kottke T, Mittag M, Hegemann P *et al.* 2016. A blue-light photoreceptor mediates the feedback regulation of photosynthesis. *Nature* **537**: 563–566.

Petrushin Ivan S, Belikov Sergei I, Belykh Olga I, Tikhonova I, Chernogor Lubov I, Stajich Jason E. 2020. Draft genome sequence of the green microalga Chlorella sp. strain BAC9706, isolated from Lake Baikal, Russia. *Microbiology Resource Announcements* **9**: e00966-20.

Pikoli MR, Sari AF, Solihat NA, Permana AH. 2019. Characteristics of tropical freshwater microalgae *Micractinium conductrix*, *Monoraphidium* sp. and *Choricystis parasitica*, and their potency as biodiesel feedstock. *Heliyon* **5**: e02922.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**: e9490.

Raanan H, Oren N, Treves H, Keren N, Ohad I, Berkowicz SM, Hagemann M, Koch M, Shotland Y, Kaplan A. 2016. Towards clarifying what distinguishes cyanobacteria able to resurrect after desiccation from those that cannot: the photosynthetic aspect. *Biochimica et Biophysica Acta (BBA) – Bioenergetics* **1857**: 715–722.

Raven JA, Beardall J, Larkum AWD, Sa'Nchez-Baracaldo P. 2013. Interactions of photosynthesis with genome size and function. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **368**: 20120264.

Raven JA, Knight CA, Beardall JJPIP. 2019. Genome and cell size variation across algal taxa. *Perspectives in Phycology* **6**: 59–80.

Riaño-Pachón DM, Corrêa LG, Trejos-Espinosa R, Mueller-Roeber B. 2008. Green transcription factors: a chlamydomonas overview. *Genetics* **179**: 31–39.

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**: 539–542.

Roth MS, Cokus SJ, Gallaher SD, Walter A, Lopez D, Erickson E, Endelman B, Westcott D, Larabell CA, Merchant SS *et al.* 2017. Chromosome-level genome assembly and transcriptome of the green alga *Chromochloris zofingiensis* illuminates astaxanthin production. *Proceedings of the National Academy of Sciences, USA* **114**: E4296–E4305.

Roth MS, Gallaher SD, Westcott DJ, Iwai M, Louie KB, Mueller M, Walter A, Foflonker F, Bowen BP, Ataii NN *et al.* 2019. Regulation of oxygenic photosynthesis during trophic transitions in the green alga *Chromochloris zofingiensis*. *Plant Cell* **31**: 579–601.

Ruban AV. 2017. Quantifying the efficiency of photoprotection. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**: 20160393.

Running JA, Huss RJ, Olson PT. 1994. Heterotrophic production of ascorbic acid by microalgae. *Journal of Applied Phycology* **6**: 99–104.

Sambrook J, Russell DW. 2006. Purification of nucleic acids by extraction with phenol : chloroform. *Cold Spring Harbor Protocols* **2006**: pdb.prot4455.

Seppey M, Manni M, Zdobnov EM. 2019. Busco: assessing genome assembly and annotation completeness. *Methods in Molecular Biology* **1962**: 227–245.

Shaul O. 2017. How introns enhance gene expression. *The International Journal of Biochemistry & Cell Biology* **91**: 145–155.

Signorile A, Sgaramella G, Bellomo F, De Rasmo D. 2019. Prohibitins: a critical role in mitochondrial functions and implication in diseases. *Cell* **8**: 71.

Singh SP, Singh P. 2015. Effect of temperature and light on the growth of algae species: a review. *Renewable and Sustainable Energy Reviews* **50**: 431–444.

Sperschneider J, Catanzariti A-M, Deboer K, Petre B, Gardiner DM, Singh KB, Dodds PN, Taylor JM. 2017. LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. *Scientific Reports* **7**: 44598.

Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* **24**: 637–644.

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. Augustus: *ab initio* prediction of alternative transcripts. *Nucleic Acids Research* **34**: W435–W439.

Starr RC, Zeikus JA. 1993. UTEX—the culture collection of algae at the University of Texas at Austin 1993 list of cultures. *Journal of Phycology* **29**: 1–106.

Tatsuta T, Model K, Langer T. 2004. Formation of membrane-bound ring complexes by prohibitins in mitochondria. *Molecular Biology of the Cell* **16**: 248–259.

Tautz D, Domazet-Lošo T. 2011. The evolutionary origin of orphan genes. *Nature Reviews. Genetics* **12**: 692–702.

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* **45**: W6–W11.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* 7: 562–578.

Treves H, Küken A, Arrivault S, Ishihara H, Hoppe I, Erban A, Höhne M, Moraes TA, Kopka J, Szymanski J *et al.* 2022. Carbon flux through photosynthesis and central carbon metabolism show distinct patterns between algae, C3 and C4 plants. *Nature Plants* **8**: 78–91.

Treves H, Murik O, Kedem I, Eisenstadt D, Meir S, Rogachev I, Szymanski J, Keren N, Orf I, Tiburcio AF *et al.* 2017. Metabolic flexibility underpins growth capabilities of the fastest growing alga. *Current Biology* **27**: 2559–2567.

Treves H, Raanan H, Finkel MO, Berkowicz SM, Keren N, Kaplan A. 2013. A newly isolated *Chlorella* sp. from desert sand crusts exhibits a unique resistance to excess light intensity. *FEMS Microbiology Ecology* **86**: 373–380.

Treves H, Raanan H, Kedem I, Murik O, Keren N, Zer H, Berkowicz SM, Giordano M, Norici A, Shotland Y *et al.* 2016. The mechanisms whereby the green alga *Chlorella ohadii*, isolated from desert soil crust, exhibits unparalleled photodamage resistance. *New Phytologist* **210**: 1229–1243.

Treves H, Siemiatkowska B, Luzarowska U, Murik O, Fernandez-Pozo N, Moraes TA, Erban A, Armbruster U, Brotman Y, Kopka J *et al.* 2020. Multi-omics reveals mechanisms of total resistance to extreme illumination of a desert alga. *Nature Plants* **6**: 1031–1043.

Trewick SC, Mclaughlin PJ, Allshire RC. 2005. Methylation: lost in hydroxylation? *EMBO Reports* **6**: 315–320.

Ullrich KK, Hiss M, Rensing SA. 2015. Means to optimize protein expression in transgenic plants. *Current Opinion in Biotechnology* **32**: 61–67.

Ungerer J, Lin P-C, Chen H-Y, Pakrasi Himadri B, Giovannoni Stephen J, Burnap R, Sherman L. 2018a. Adjustments to photosystem stoichiometry and electron transfer proteins are key to the remarkably fast growth of the cyanobacterium *Synechococcus elongatus* UTEX 2973. *mBio* **9**: e02327-17.

Ungerer J, Wendt KE, Hendry JI, Maranas CD, Pakrasi HB. 2018b. Comparative genomics reveals the molecular determinants of rapid growth of the cyanobacterium *Synechococcus elongatus* UTEX 2973. *Proceedings of the National Academy of Science USA* **115**: E11761–E11770.

Van Aken O, Ford E, Lister R, Huang S, Millar AH. 2016. Retrograde signalling caused by heritable mitochondrial dysfunction is partially mediated by ANAC017 and improves plant performance. *The Plant Journal* **88**: 542–558.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK *et al.* 2014. PILON: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**: e112963.

Wang H-F, Feng L, Niu D-K. 2007. Relationship between mRNA stability and intron presence. *Biochemical and Biophysical Research Communications* **354**: 203–208.

Weber B, Bowker M, Zhang Y, Belnap J. 2016. Natural recovery of biological soil crusts after disturbance. In: *Biological soil crusts: an organizing principle in drylands*. Basel, Switzerland: Springer.

Weider LJ, Elser JJ, Crease TJ, Mateos M, Cotner JB, Markow TA. 2005. The functional significance of ribosomal (r)DNA variation: impacts on the evolutionary ecology of organisms. *Annual Review of Ecology, Evolution, and Systematics* **36**: 219–242.

Wilhelmsson PKI, Muhlich C, Ullrich KK, Rensing SA. 2017. Comprehensive genome-wide classification reveals that many plant-specific transcription factors evolved in Streptophyte algae. *Genome Biology and Evolution* **9**: 3384–3397.

Wittkopp TM, Schmollinger S, Saroussi S, Hu W, Zhang W, Fan Q, Gallaher SD, Leonard MT, Soubeyrand E, Basset GJ *et al.* 2017. Bilin-dependent photoacclimation in *Chlamydomonas reinhardtii*. *Plant Cell* **29**: 2711–2726.

Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with KRAKEN 2. *Genome Biology* **20**: 257.

Wu M, Zhang H, Sun W, Li Y, Hu Q, Zhou H, Han D. 2019. Metabolic plasticity of the starchless mutant of *Chlorella sorokiniana* and mechanisms underlying its enhanced lipid production revealed by comparative metabolomics analysis. *Algal Research* **42**: 101587.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* **13**: 555–556.

Yu G, Wang LG, Han Y, He QY. 2012. CLUSTERPROFILER: an R package for comparing biological themes among gene clusters. *Omics* **16**: 284–287.

Yu H, Kim J, Lee C. 2019. Nutrient removal and microalgal biomass production from different anaerobic digestion effluents with Chlorella species. *Scientific Reports* **9**: 6123.

Zuñiga C, Li C-T, Huelsman T, Levering J, Zielinski DC, Mcconnell BO, Long CP, Knoshaug EP, Guarnieri MT, Antoniewicz MR *et al.* 2016. Genome-scale metabolic model for the green alga *Chlorella vulgaris* UTEX 395 accurately predicts phenotypes under autotrophic, heterotrophic, and mixotrophic growth conditions. *Plant Physiology* **172**: 589–602.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 *Ohadii* genome phylogenetic analysis.

Fig. S2 Calculated K_a : K_s values for different GO biological process terms in several *Chlorella* spp.

Fig. S3 Phylogenetic trees of *Ohadii* genes potentially horizontally transferred from bacteria.

Fig. S4 *In silico* estimation of rDNA copy number in the genomes of several algal species.

Fig. S5 Transcript abundance, of mixotrophically grown *Ohadii* cells, as a function of intron density.

Fig. S6 Boxplot of gene expression values vs all parameter combinations.

Fig. S7 Comparative intron abundance analysis of *Chlorella* spp.

Fig. S8 Effect of C level on *Ohadii* vegetative cell cycle.

Fig. S9 *Ohadii* organellar genomes.

Notes S1 Orphan genes.

Notes S2 Metabolic flexibility and polyamines.

Notes S3 Light and C regulation of cell cycle in *Ohadii* cultures.

Notes S4 Comparative analysis of chloroplast- and mitochondrial-encoded genes.

Notes S5 Mitochondrial alternative oxidase.

Notes S6 Carbonic anhydrases and bestrophin-like proteins involved in the CO_2 concentrating mechanism.

Notes S7 Ascorbate glutathione cycle.

Notes S8 Enrichment in glutamines and cysteines.

Table S1 Expression profiles of *Ohadii* genes under various conditions.

Table S2 Gene annotation statistics and QC parameters.

Table S3 BLAST, eggnog, GO, KEGG and KOG annotations for all 11 407 *Ohadii* genes.

Table S4 Genes putatively acquired from bacteria through horizontal gene transfer.

Table S5 Output of the RepeatMasker tool for the assembly of the *Ohadii* genome.

Table S6 Ribosomal proteins associated gene expression under EIL.

Table S7 Introns per gene in *Ohadii* and *C. variabilis* genomes.

Table S8 Three-way ANOVA analysis with interaction effects of gene expression levels.

Table S9 Orthologues groups in several algal genomes. Prohibitin genes and EIL expression.

Table S10 List of orphan genes found in *Ohadii* genome.

Table S11 Transcription-associated proteins in *Ohadii* genome.

Table S12 Genome accessions used in this study.

Table S13 Cell cycle-associated gene expression.

Table S14 Glutamine content per protein in *Ohadii* and orthologous proteins in other algae.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.