# Targeted genetic manipulation and yeast-like evolutionary genomics in the green alga *Auxenochlorella*

Rory J. Craig[1,2], Marco A. Dueñas[3*], Dimitrios J. Camacho[4*], Sean D. Gallaher[1], Maria Clara Avendaño-Monsalve[1], Yang-Tsung Lin[1], Crysten E. Blaby-Haas[5,6], Jeffrey L. Moseley[1], Sabeeha S. Merchant[1,3,4,7]

[1] California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, CA 94720, USA
[2] School of BioSciences, University of Melbourne, Parkville, VIC 3010, Australia
[3] Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA
[4] Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA
[5] Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[6] US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[7] Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

Correspondence: Rory J. Craig, rory.craig@unimelb.edu.au; Sabeeha S. Merchant, sabeeha@berkeley.edu

*Authors contributed equally

Keywords: permuted tRNA, lncRNA, chlorophyll aerobic cyclase, CHL27, reverse genetics, gene targeting, oleaginous, triacylglyceride, GreenCut, epigenetics, synthetic biology

Short title: *Auxenochlorella* diploid genomics & genetic tools

One-sentence summary: *Auxenochlorella*, green algae shaped by evolutionary forces acting on vegetative diploids, are amenable to discovery research and bioengineering via efficient site-specific homologous recombination

## ABSTRACT

*Auxenochlorella* spp. are diploid oleaginous green algae whose streamlined genomes can be readily manipulated by homologous recombination, making them highly amenable to discovery research and bioengineering. Vegetatively diploid organisms experience specific evolutionary phenomena, including allodiploid hybridization, mitotic recombination, loss-of-heterozygosity and aneuploidy; however, studies of these forces have largely focused on yeasts. Here, we present a telomere-to-telomere phased diploid genome assembly of *Auxenochlorella* UTEX 250-A (haploid length 22 Mb) and introduce a genetic toolkit for site-specific manipulation of the nuclear genome in multiple strains, featuring several selectable markers, inducible promoters, and fluorescent reporters for protein localization. UTEX 250-A is an allodiploid hybrid of *Auxenochlorella protothecoides* and *Auxenochlorella symbiontica*, two species differentiated by extensive chromosomal rearrangements. UTEX 250-A haplotypes are a mosaic of each parental species following mitotic recombination, and two chromosomes are trisomic. Loss-of-heterozygosity events are pervasive across *Auxenochlorella* and can evolve rapidly in the laboratory. High-quality structural annotation yielded ~7,500 genes per haplotype. *Auxenochlorella* have experienced gene family loss and reduction, including core photosynthesis genes, and exhibit periodic adenine and cytosine methylation at promoters and gene bodies, respectively. Approximately 10% of genes, especially those involved in DNA repair and sex, overlap antisense long noncoding RNAs, which may participate in a regulatory mechanism. We demonstrate the utility of *Auxenochlorella* for fundamental research by knockout of a chlorophyll biosynthesis enzyme, and confirm one trisomy by allele-specific transformation. These results demonstrate the generality of several evolutionary forces associated with vegetative diploidy and provide a foundation for use of *Auxenochlorella* as a reference organism.

## INTRODUCTION

Green algae present excellent opportunities for discovery in plant biology, especially concerning fundamental molecular and cellular processes. The primary algal model *Chlamydomonas reinhardtii* has been central to research in photosynthesis and chloroplast biogenesis, in addition to functions and processes that are broadly relevant to eukaryotic cells, such as cilia / ciliogenesis and the cell cycle (Sasso et al. 2018; Salome and Merchant 2019). As photosynthetic organisms that can utilize atmospheric carbon dioxide in the production of high value bioproducts, including specialty chemicals and nutraceuticals, green algae also hold great potential in biotechnology (Goold et al. 2024). The ability to edit, add or remove genes is essential to both of these endeavors. Despite the high growth rates and experimental tractability of several algae, the tools to genetically manipulate algal genomes have typically lagged behind those available for many bacteria and yeasts (Sproles et al. 2021). In *C. reinhardtii*, transgenes are typically integrated at untargeted locations via non-homologous end-joining, although recent advances in ribonucleoprotein-mediated approaches have enabled site-specific manipulation (Nievergelt 2025). Non-targeted transgene integration and CRISPR-Cas9 gene editing are also possible in species from industrially relevant trebouxiophyte genera such as *Chlorella* and *Picochlorum* (Yang et al. 2016; Lin and Ng 2020; Krishnan et al. 2025). However, efficient site-specific transformation by homologous recombination, as is commonplace in the budding yeast *Saccharomyces cerevisiae* (Orr-Weaver et al. 1981), is incredibly rare in green algae, having been reported sparingly in species like the ecologically relevant prasinophyte *Ostreococcus tauri* (Lozano et al. 2014).

*Auxenochlorella* is a genus of trebouxiophyte green algae in which homologous recombination is efficient. Site-specific transformation of the nuclear genome in *Auxenochlorella protothecoides*, as well its close relatives from the genus *Prototheca*, has been reported in patent filings (Franklin et al. 2013; Moseley et al. 2024). While *Prototheca* are obligate heterotrophs that have lost the ability to photosynthesize, *Auxenochlorella* can grow robustly as a phototroph, mixotroph or heterotroph, undergoing a metabolic switch that degrades the photosynthetic apparatus when provided with an organic carbon source (Matsuka et al. 1966). Therefore, it is facile to study photosynthetic gene function through reverse genetics, making *Auxenochlorella* an attractive

model in plant biology. Both *Auxenochlorella* and *Prototheca* are oleaginous, and *Auxenochlorella* oil and biomass have generally recognized as safe (GRAS) status, favoring their use for engineering of specific lipids for bioproducts and food applications, in addition to biofuels (Franklin et al. 2013; Brooks et al. 2024; Goold et al. 2024). We recently demonstrated the utility of *Auxenochlorella* for discovery research in a study requiring introduction and quantitation of gene expression from several synthetic gene constructs to deduce the mechanism underlying the translation of bicistronic genes (Dueñas et al. 2025a), which are widespread in green algae (Gallaher et al. 2021).

Like budding yeast, *Auxenochlorella* are vegetatively diploid. Microbial organisms that undergo periods of asexual reproduction as diploids are associated with several molecular and evolutionary phenomena. These processes are best studied in *S. cerevisiae* and its close relatives, which undergo a facultatively sexual life cycle featuring long periods of asexual division as a diploid (Tsai et al. 2008; Fischer et al. 2021). Although inter-tetrad outcrossing is relatively rare in yeast, one interesting facet of this life cycle is occasional alloploid hybridization. Mating of haploid gametes from divergent lineages produces an allodiploid hybrid that inherits one set of chromosomes from each parent, whereas allotetraploid hybrids can arise via conjugation of divergent diploid cells or whole-genome duplication of an allodiploid individual (Sipiczki 2008; Gabaldón 2020). Although allodiploid hybrids may be incapable of meiosis due to incompatibilities between the parental chromosomes, they can persist via asexual division (Gabaldón 2020). Indeed, hybrids may exhibit increased fitness or unique phenotypes relative to their parents (i.e., heterosis), and hybrid *Saccharomyces* strains have successfully adapted to anthropomorphic environments such as alcoholic ferments (Sipiczki 2008; Peris et al. 2018; Gallone et al. 2019; Langdon et al. 2019) and olive brine (Pontes et al. 2019). Hybridization has also been linked to the evolution of pathogenicity in taxa including *Candida* yeasts (Pryszcz et al. 2015; Schroder et al. 2016) and *Aspergillus* filamentous fungi (Steenwyk et al. 2020). Importantly, allodiploid hybridization is not necessarily an evolutionary dead end, since the correct pairing of homologous chromosomes and sexual compatibility can be restored via whole-genome duplication (Charron et al. 2019). Exemplifying this process, an ancient hybridization event is thought to have preceded the whole-genome duplication in the lineage leading to *Saccharomyces* (Marcet-Houben and Gabaldón 2015). Allopolyploidy (hybridization and

128 genome duplication) has also long been recognized as a major force in plant evolution and
129 speciation (Soltis and Soltis 2009).
130
131 Although recombination is primarily associated with meiotic cell division, various mitotic
132 recombination processes can occur during asexual growth. In *S. cerevisiae*, mitotic
133 recombination occurs randomly due to double-strand breaks at a rate that is orders of magnitude
134 lower than meiotic recombination (Mancera et al. 2008; Sui et al. 2020). Nonetheless, over
135 lengthy periods of asexuality, mitotic recombination can drastically homogenize genomes via
136 loss-of-heterozygosity (LOH) events, where an allele of one chromosome is replaced by the
137 allele carried by its homologous chromosome (Dutta and Schacherer 2025). LOH can occur at
138 short regions along chromosomes of generally less than 10 kb via gene conversion (i.e.,
139 interstitial LOH), or at larger regions that typically extend from an internal location to the
140 telomere (terminal LOH), which are the products of mitotic crossovers or the break-induced
141 replication DNA repair pathway (Sui et al. 2020; Dutta et al. 2021). While LOH can negatively
142 impact fitness (e.g., by unmasking recessive deleterious mutations), it is also an important
143 mechanism by which allelic incompatibilities are purged in hybrids (Smukowski Heil et al. 2017;
144 Lancaster et al. 2019). Furthermore, reciprocal mitotic crossovers result in the shuffling of
145 haplotypes between the parental sub-genomes of hybrids, producing chromosomes that are a
146 mosaic of the ancestral parents (Sipiczki 2008). Mitotic recombination has also been
147 characterized in diploid oomycetes (Dale et al. 2019) and diatoms (Bulankova et al. 2021),
148 suggesting that it is an important evolutionary process across the tree of life.
149
150 Another phenomenon associated with asexual reproduction is the emergence of aneuploidy via
151 the loss or gain of copies of specific chromosomes. Unbalanced chromosome numbers have been
152 observed in many eukaryotes and are prevalent in several fungi (Vande Zande et al. 2023). As
153 with hybridization, aneuploidy and its subsequent effects (e.g., gene expression changes) are
154 associated with adaptation to new environments and pathogenicity, although negative fitness
155 effects are also common (Santaguida and Amon 2015). For example, whole or partial
156 aneuploidies are common among clinical isolates of *Candida albicans* (Hirakawa et al. 2015),
157 and an aneuploidy affecting a specific chromosome arm causes antifungal drug resistance
158 (Selmecki et al. 2006). Under laboratory evolution of *C. albicans*, chromosomal duplications

159 have also been observed as transient adaptations to abrupt heat stress (Yona et al. 2012).
160 Transient aneuploidy can result in LOH at the scale of entire chromosomes via duplication of
161 one chromosome copy and loss of the other (Yuen et al. 2007; Andersen et al. 2008).
162
163 These drivers of evolution have received scarce attention in chlorophyte algae. Although
164 facultatively sexual, *C. reinhardtii* (class Chlorophyceae) grows asexually as a haploid, with the
165 diploid stage usually limited to a dormant zygospore that obligately divides by meiosis (Harris
166 2001). Similarly, asexual reproduction in *Chlorella* (Trebouxiophyceae) is haploid, with a
167 possible cryptic uncharacterized sexual cycle (Blanc et al. 2010; Fučiková et al. 2015). Ancient
168 mating type loci have also been identified in haploid *Ostreococcus* species (Mamelliophyceae),
169 demonstrating the presence of haplontic life cycles at the deepest evolutionary branch of the
170 Chlorophyta (Benites et al. 2021). Despite this, several vegetatively diploid species have recently
171 been reported (Figure 1A). Diploid genome assemblies have been assembled for trebouxiophyte
172 algae from the related genera *Nannochloris* (Sanders et al. 2022) and *Picochlorum* (Foflonker et
173 al. 2018; Becker et al. 2020; Barten et al. 2022; da Roza et al. 2024), and for the lichen-forming
174 *Trebouxia lynniae* (Gazquez et al. 2024). *Chloropicon primus* (Chloropicophyceae) is trisomic
175 (three copies) for a single chromosome and is otherwise diploid (Lemieux et al. 2019). In the
176 Chlorophyceae, diploids from the genera *Haematococcus* (Volvocales) and *Tetradesmus*
177 (Sphaeropleales) have been reported (Calhoun et al. 2021; Marcolungo et al. 2024). Most
178 notably, Biondi et al. (2024) described a diploid *Tetradesmus obliquus* strain with extensive
179 heterozygosity but one entirely homozygous chromosome, patterns potentially consistent with
180 hybridization followed by chromosome-scale LOH. These results suggest that the presence of
181 diplontic life cycles may be underestimated in Chlorophyta, although the prevalence of
182 hybridization, LOH and aneuploidy is presently unknown.
183
184 Here, we present a fully phased, telomere-to-telomere diploid genome assembly for the
185 *Auxenochlorella* strain UTEX 250-A. We find an unusual genome architecture where six of the
186 twelve chromosome pairs are highly rearranged and non-homologous. We demonstrate that
187 UTEX 250-A is an allodiploid hybrid of *Auxenochlorella protothecoides* and *Auxenochlorella*
188 *symbiontica*, which each differ in their respective karyotypes. Despite the high heterozygosity
189 introduced by hybridization, several local genomic regions and three entire chromosomes are

homozygous following extensive LOH, while two other chromosomes are trisomic. We introduce a highly curated structural annotation of the genome, reporting a streamlined nuclear genome architecture and gene content. Finally, we present a toolkit for genetic manipulation, featuring multiple selectable markers and fluorescent protein expression for intracellular localization studies. We demonstrate application of these tools for reverse genetics via targeted disruption of a gene encoding a chlorophyll biosynthesis enzyme in UTEX 250-A and strains of both parental species, and experimentally verify one UTEX 250-A trisomy by allele-specific targeting of a reporter construct. These resources demonstrate that a green algal genome can be shaped by forces experienced by diplontic taxa like yeasts, and provides a platform for the use of *Auxenochlorella* as a facile reference organism in plant biology and bioengineering.

## RESULTS

### The genus *Auxenochlorella* and existing genomic data

First isolated in the 1890s as "*Chlorella*" *protothecoides* (Krüger 1894b; Krüger 1894a), *Auxenochlorella* was later established as an independent monotypic genus based on physiological, morphological and genetic data (Kalina and Punčochárová 1987; Huss et al. 1999). Indeed, *Auxenochlorella* and true *Chlorella* species likely diverged more than 550 million years ago (Del Cortona et al. 2020) and form distinct lineages within the taxonomically diverse order Chlorellales (Figure 1A). The closest relatives of *Auxenochlorella* have all lost photosynthesis; *Prototheca* is a polyphyletic assemblage of heterotrophic free-living and opportunistically parasitic algae, and *Helicosporidium* are obligate parasites of arthropods (Jagielski et al. 2019). Collectively, these algae form the "AHP lineage" (Ueno et al. 2005), and phylogenetic analyses support multiple independent transitions to heterotrophy, with *Auxenochlorella* enduring as the sole photosynthetic lineage (Figueroa-Martinez et al. 2015; Suzuki et al. 2018; Guo et al. 2022).

Several isolates classified as *A. protothecoides* are maintained in culture centers, including the original strain isolated by W. Krüger from the sap of a poplar tree (UTEX 25 at the University of Texas Culture Collection of Algae). *Auxenochlorella* have also been isolated from fresh and salt water, soil, and as symbionts of freshwater hydrozoans and sponges (Figure 1B). Based on ribosomal DNA (rDNA) analysis, Darienko and Pröschold (2015) described a second species, *A.*

*symbiontica*, with the type strain CCAP 211/61 isolated from *Hydra viridis*. We produced an 18S rDNA phylogeny including some additional isolates (Dataset S1), which supports an association between *A. protothecoides* and both tree sap and freshwater habitats, whereas most *A. symbiontica* strains were isolated from marine environments (Figure 1B). However, the marine strain UTEX 2341 groups with *A. protothecoides* and freshwater symbionts have been isolated for both species, suggesting that both *A. protothecoides* and *A. symbiontica* can live in multiple and overlapping environments.

Gao et al. (2014) produced the first draft assembly of *A. protothecoides* "0710" (=UTEX 25, see below), while additional draft assemblies of UTEX 25 (Vogler et al. 2018) and UTEX 2341 are available at NCBI (Dataset S2). Several draft assemblies of *Prototheca* and *Helicosporidium* have been produced (Pombert et al. 2014; Severgnini et al. 2018; Suzuki et al. 2018; Bakuła et al. 2021; Guo et al. 2022). Although these assemblies are all collapsed to haploids, Guo et al. (2022) reported heterozygosity in two *Prototheca wickerhamii* genomes. Every available AHP lineage genome is highly streamlined relative to the ~40-60 Mb genomes of *Chlorella* species; for example, the *A. protothecoides* 0710 assembly has a haploid length of 22.9 Mb, with ~7,000 predicted genes and low repeat content (Gao et al. 2014) (Table 1). It is noteworthy that *Picochlorum* and *Nannochloris* species also possess diploid streamlined genomes (Table 1), and phylogenetic placement of these species as a sister lineage of *Chlorella* (including *Micractinium*) received low bootstrap support in our analysis (Figure 1A). Whether the transition to a diplontic lifecycle occurred independently in the AHP lineage and *Picochlorum*/*Nannochloris*, or whether these lineages may in fact form sister groups, is an outstanding question.

| Genome | | Size (Mb) | Genes | TE genes | lncRNA genes | % GC | % Repeats | Reference |
|---|---|---|---|---|---|---|---|---|
| *Auxenochlorella* UTEX250-A | Haplotype A | 22.0 | 7,509 | 86 | 547 | 63.9 | 6.22 | This study |
| | Haplotype B | 22.0 | 7,515 | 121 | 574 | 63.9 | 6.21 | This study |
| *Auxenochlorella protothecoides* 0710 | | 22.9 | 6,988 | 26 | N/A | 63.6 | 5.58 | Gao, et al. 2014 |
| *Prototheca cutis* JCM15793 | | 20.0 | 5,600 | 5 | N/A | 61.3 | 5.84 | Suzuki et al. (2018) Kwon et al. (2023) |
| *Chlorella vulgaris* 211/11P | | 40.2 | 10,341 | 357 | N/A | 61.8 | 13.1 | Cecchin et al. (2019) |
| *Chlorella sorokiniana* UTEX 1602 | | 59.4 | 9,489 | 37 | N/A | 64.1 | 12.9 | Arriola et al. (2018) |
| *Micractinium conductrix* SAG 241.80 | | 60.8 | 9,027 | 190 | N/A | 67.3 | 21.5 | Arriola et al. (2018) |
| *Nannochloris desiccata* UTEX 2526 | | 21.6 | 9,123 | 187 | N/A | 45.0 | 8.12 | Sanders et al. (2022) |
| *Picochlorum* sp. BPE23 | Haplotype A | 14.8 | 7,216 | 11 | N/A | 46.3 | 9.72 | Barten et al. (2022) |
| *Coccomyxa subellipsoidea* v3 | | 48.9 | 10,836 | 57 | N/A | 52.9 | 6.04 | Blanc, et al. (2012) |
| *Chlamydomonas reinhardtii* v6.1 | | 114.0 | 16,801 | 810 | N/A | 64.1 | 26.4 | Craig et al. (2023) |

**Table 1.** Summary statistics of representative trebouxiophyte genomes with *C. reinhardtii* as an outgroup. See methods for determination of transposable element (TE) genes included in annotations; note that these numbers are not representative of the true number of TE genes in each genome due to differences in annotation methodology. lncRNA = long noncoding RNA. N/A = not applicable.

## Telomere-to-telomere, phased diploid assembly of *Auxenochlorella* UTEX 250-A

To serve as a foundation for discovery research and bioengineering, we sought to produce a reference-quality genome assembly of *Auxenochlorella*. We selected the strain UTEX 250 based on its robust performance in high cell density fermentations (Brooks et al. 2024). Metadata on UTEX 250 are sparse; the strain was isolated before 1950 in the Netherlands by A. J. Kluyver, possibly from freshwater. Although described as *A. protothecoides*, UTEX 250 groups with *A. symbiontica* based on 18S rDNA (Darienko and Pröschold 2015) (Figure 1B). Since our line of the strain has been independently maintained for several years after acquisition from UTEX, we performed single colony purification prior to genome sequencing, and we designate this sub-clone "UTEX 250-A".

We produced high-coverage Pacific Biosciences (PacBio) HiFi and linked-read OmniC datasets for UTEX 250-A, from which we produced a genome assembly via automated and manual assembly (see Methods). This effort yielded a telomere-to-telomere diploid genome assembly comprising two haplotypes, each featuring 12 gapless nuclear chromosomes spanning ~22.0 Mb (i.e., collectively 24 nuclear chromosomes spanning 44.0 Mb) (Table 1). Heterozygosity between the two haplotypes is 2.71% (i.e., homologous chromosomal regions differ by ~27 single nucleotide polymorphisms per 1 kb; heterozygosity calculations exclude entirely

homozygous regions, see below). This extensive inter-haplotype variation made it possible to fully phase the assembled chromosomes using the accurate PacBio reads, so that each haplotype corresponds to a physical chromosome in UTEX 250-A. We arbitrarily label these haplotypes "A" and "B". As presented below, only six of the twelve chromosome pairs exhibit one-to-one homology, and the remaining six are highly rearranged between the two haplotypes.

All chromosomes terminate in the telomeric repeat (TTAGGG)n, except for one chromosome arm per haplotype that terminates in a truncated rDNA array. This telomeric motif differs from the $(TTTAGGG)_n$ repeat found in the trebouxiophyte genera *Chlorella* and *Coccomyxa*, which is likely ancestral to green algae, suggesting an independent transition to $(TTAGGG)_n$ repeats as observed in some other green algae (Fulnečková et al. 2012). The 5S rDNA genes, which generally form tandem arrays independent of the major rDNA arrays, are present as 16 (haplotype A) or 17 (haplotype B) standalone genes dispersed throughout the genome. Similar non-tandem 5S rDNA genes have been observed in species including the yeasts *Schizosaccharomyces pombe* and *Yarrowia lipolytica* (Torres-Machorro et al. 2010). An initial search for tRNA genes using tRNAscan-SE (Chan et al. 2021) yielded an incomplete set, and we subsequently detected permuted tRNA genes in which the 5' and 3' halves of the tRNA gene are organized in reverse (Figure S1). This unusual structure was previously reported in the genomes of the red alga *Cyanidioschyzon merolae* (Soma et al. 2007) and prasinophyte green algae (Maruyama et al. 2010).

As in *A. protothecoides* 0710, repeat content is low, with interspersed and tandem repeats spanning 3.4% and 2.8% of the genome, respectively (Table 1). Among the interspersed repeats, we identified potentially active families of *Metaviridae* long terminal repeat (LTR) retrotransposons, and *hAT* and *EnSpm-Plavaka* cut-and-paste DNA transposons, with many insertions segregating as polymorphisms between the two haplotypes (Figure S2, Dataset S3). The most abundant repeat is an uncharacterized element that encodes a Fanzor protein, which are RNA-guided endonucleases carried by diverse transposons in eukaryotes and their giant viruses (Saito et al. 2023). Phylogenetically diverse Fanzors are sporadically found in green algal genomes including *C. reinhardtii* (Bao and Jurka 2013) and *Chloropicon primus* (Yoon et al. 2023), and presumably arise via lateral gene transfer.

307 Green algal centromeres identified thus far are transposon-enriched regions of tens to hundreds
308 of kilobases in species of *Chlorella*, *Coccomyxa* and *Chlamydomonas* (Blanc et al. 2012; Craig
309 et al. 2021; Wang et al. 2024). In line with the paucity of repeats, we found no obvious repeat-
310 rich centromere candidates in UTEX 250-A. We searched for short regional centromeres,
311 nonrepetitive AT-rich regions of 1-5 kb as found in *C. merolae* (Kanesaki et al. 2015) and the
312 diatom *Phaeodactylum tricornutum* (Diner et al. 2017). On 9 of the 12 chromosomes, we
313 identified single AT-rich (mean 48.0% GC, relative to 63.9% GC genome-wide) intergenic
314 regions spanning 2.1 – 7.6 kb (Figure S3, Dataset S4). Although ChIP-sequencing of the
315 centromeric histone variant will be required to confirm centromere locations, considered
316 alongside the low repeat content and non-tandem 5S rDNA genes, these candidates support a
317 highly streamlined architecture of the UTEX 250-A genome.
318
319 We produced complete circular genome assemblies for the plastome (84.6 kb, 30.8% GC; Figure
320 S4) and mitogenome (54.0 kb, 29.0% GC; Figure S5). The UTEX 250-A plastome is broadly
321 consistent with existing assemblies from *A. protothecoides* UTEX 25 (Yan et al. 2015; Park et al.
322 2022). Briefly, the plastome features 78 protein-coding genes and lacks internal repeats, introns,
323 and trans-spliced genes, resulting in a more compact architecture relative to most trebouxiophyte
324 plastomes. The mitogenome features 37 protein-coding genes, including two LAGLIDADG
325 homing endonucleases encoded by introns in *cox1*, as also observed in the *P. wickerhamii*
326 mitogenome (Wolff et al. 1993).
327
328 ***Auxenochlorella* UTEX 250 is a putative allodiploid hybrid**
329 Considering the high heterozygosity and rearranged nature of half of the UTEX 250-A diploid
330 chromosomes, we speculated that UTEX 250 may be an allodiploid hybrid between two distinct
331 parental species that possess different karyotypes. Given the close evolutionary relationship
332 between *A. protothecoides* and *A. symbiontica* (Figure 1B), we explored this hypothesis via
333 genome sequencing of strains of both species. We assembled gapless chromosome-level
334 assemblies for *A. protothecoides* strains UTEX 25 and UTEX 2341, and *A. symbiontica* CCAP
335 211/61, from PacBio HiFi data (supplemented by Oxford Nanopore Technologies, ONT,
336 sequencing for CCAP 211/61; see Methods). The three assemblies were broadly consistent with
337 our UTEX 250-A assembly; each nuclear genome comprises 12 diploid chromosomes, with

338 haploid genome sizes ranging from 21.9 Mb to 23.0 Mb (Dataset S5). However, unlike UTEX
339 250-A, we observed no rearrangements between the homologous chromosome pairs in each
340 genome, implying a typical diploid karyotype in these strains.
341
342 We first considered the six homologous UTEX 250-A chromosomes. Each of these six
343 chromosomes corresponds to a pair of homologous chromosomes in UTEX 25, UTEX 2341, and
344 CCAP 211/61, suggesting that they are ancestral to *Auxenochlorella*. In UTEX 250-A, three of
345 these chromosomes (1, 2 and 6) are mostly heterozygous, whereas the remaining chromosomes
346 (4, 9 and 10) are entirely homozygous (Figure 2A). We presume that the homozygous
347 chromosomes were originally heterozygous, with homozygosity arising following chromosome-
348 scale LOH events (see below). As introduced, heterozygosity in UTEX 250-A is ~2.7%,
349 although many local genomic regions are homozygous (e.g., the ~73 kb homozygous region at
350 the left arm terminus of chromosome 1, Figure 2A). In *A. protothecoides*, heterozygosity is
351 substantially lower at 0.59% (UTEX 25) and 0.62% (UTEX 2341) in heterozygous regions, with
352 both genomes also exhibiting extensive regional homozygosity. Average heterozygosity in
353 CCAP 211/61 is 1.35% in heterozygous regions, implying that *A. symbiontica* harbors greater
354 genetic diversity than *A. protothecoides*.
355
356 When comparing the A and B haplotypes of the UTEX 250-A chromosomes to the other
357 *Auxenochlorella* strains, we observed substantial inter-haplotype variation in genetic divergence
358 (Figure 2A). For example, on chromosome 2, haplotype A differs on average by 0.57% and
359 0.49% relative to the homologous chromosomes of UTEX 25 and UTEX 2341, respectively.
360 Meanwhile, haplotype B differs by 2.69% (UTEX 25) and 2.66% (UTEX 2341). Therefore, for
361 chromosome 2, the divergence between haplotype A and the *A. protothecoides* genomes is
362 comparable to heterozygosity between chromosome pairs in *A. protothecoides*, whereas the
363 divergence between haplotype B and *A. protothecoides* is comparable to heterozygosity in
364 UTEX 250-A. The reciprocal pattern is observed via comparison to *A. symbiontica*; divergence
365 between haplotype B and CCAP 211/61 (1.64%) is comparable to heterozygosity in CCAP
366 211/61, whereas divergence between haplotype A and CCAP 211/61 (2.84%) is comparable to
367 UTEX 250-A heterozygosity. These results are consistent with haplotype A of chromosome 2
368 originating from an *A. protothecoides*-like ancestor, and haplotype B originating from an *A.*

369 *symbiontica*-like ancestor. Extending this logic, it is possible to use genetic divergence to assign
370 a putative parent of origin to all genomic regions in UTEX 250-A (Figure 2A, B).
371
372 In contrast to chromosome 2, we observed intra-haplotype switches in assigned parentage on the
373 remaining five homologous chromosome pairs (Figure 2A). On chromosome 1, a single switch
374 corresponds to the 73 kb homozygous region at the left arm terminus, and can be attributed to a
375 LOH event in which the ancestral haplotype A copy of this region was replaced by haplotype B,
376 resulting in both resembling *A. protothecoides*. Conversely, the switches on chromosome 6 are
377 reciprocal, resulting in a ~360 kb region of *A. protothecoides*-like sequence on an otherwise *A.*
378 *symbiontica*-like chromosome for haplotype A (and vice versa on haplotype B). This mosaicism
379 could be explained by mitotic crossover following hybridization, i.e., each haplotype originally
380 corresponded to one parent (as for chromosome 2), but has subsequently undergone reciprocal
381 exchange. The homozygous chromosomes 4, 9 and 10 also exhibit switches in assigned
382 parentage that are consistent with either LOH or crossover events occurring in a heterozygous
383 state that preceded complete LOH. Note that the A and B labels are arbitrary, and the genomic
384 mosaicism makes it impossible to correspond physical haplotypes to parents of origin in UTEX
385 250-A.
386
387 We next tested whether the UTEX 250-A rearranged chromosomes correspond to the ancestral
388 karyotypes of *A. protothecoides* and *A. symbiontica*. The extent of the rearrangements among
389 these chromosomes can only be explained by multiple translocations, one fusion, and one fission
390 (Figure 2B). Remarkably, eight of the twelve chromosome copies correspond exactly to a
391 homologous chromosome pair in either *A. protothecoides* (both UTEX 25 and UTEX 2341,
392 which have an identical karyotype) or *A. symbiontica* (CCAP 211/61). Specifically, UTEX 250-
393 A chromosomes B05 (i.e., chromosome 5, haplotype B), A08 and A11 correspond to
394 homologous chromosome pairs in *A. protothecoides*, whereas A05, A07, B08, B11 and A12
395 correspond to homologous chromosome pairs in *A. symbiontica* (Figure 2C). Therefore, most of
396 the rearrangements in UTEX 250-A can be explained by direct inheritance from parental species
397 with different karyotypes. The remaining chromosomal rearrangements can be explained by
398 events that occurred following hybridization. The UTEX 250-A chromosomes B03 and B12
399 correspond to a pair of *A. protothecoides* chromosomes, suggesting fission of the ancestral *A.*

400 *protothecoides* chromosome 3 and subsequent fusion of one of the resulting chromosomal
401 fragments (which is predicted to have lacked a centromere) to the ancestral *A. protothecoides*
402 chromosome 12 (Figure 2C). Interestingly, LTR elements are present at both the *de novo*
403 terminus of B03 and the fusion site of B12, suggesting that retrotransposons may have invaded
404 (and potentially stabilized) the unprotected chromosomal ends following fission. Finally, the
405 UTEX 250-A chromosomes A03 and B07 are consistent with a reciprocal translocation between
406 the ancestral *A. protothecoides* chromosome 7 and *A. symbiontica* chromosome 3 (Figure 2C).
407 These ancestral chromosomes feature a common locus of ~6 kb that corresponds to two genes
408 that are part of the largest *Auxenochlorella*-specific gene family (see below), suggesting that the
409 translocation occurred via ectopic recombination between these nonhomologous repeats. It is
410 also noteworthy that switches in the assigned parent of origin are frequently observed on the
411 rearranged chromosomes (Figure 2B, C), implying that mitotic recombination has occurred
412 between homologous regions despite the rearrangements.
413
414 Finally, we constructed a phylogeny based on genomic regions that i) are present on UTEX 250-
415 A chromosomes that have not obviously experienced mitotic crossovers, and ii) are not affected
416 by LOH in any of the four strains. As expected, this analysis clearly divided the UTEX 250-A
417 genome into regions that clustered with either UTEX 25 and UTEX 2341 (*A. protothecoides*-
418 like) or CCAP 211/61 (*A. symbiontica*-like), with the longer branch lengths for *A. symbiontica*
419 reflecting the higher genetic diversity in the species (Figure 2D). We also constructed trees for
420 the plastome (Figure 2E) and mitogenome (Figure 2F), which demonstrated putative inheritance
421 from *A. symbiontica* in both cases.
422
423 Overall, *A. protothecoides* and *A. symbiontica* represent authentic species that are ~2.7%
424 divergent at the sequence-level and are distinguished at the chromosome-level by multiple
425 rearrangements. Assuming that the sequenced strains are representative, genetic diversity may be
426 more than two-fold higher in *A. symbiontica* than *A. protothecoides*. UTEX 250 is a putative
427 allodiploid hybrid that presumably originated via the fusion of haploid cells of *A. symbiontica*
428 and *A. protothecoides*. The ancestrally inherited haplotypes have subsequently been mixed via
429 nonreciprocal and reciprocal mitotic exchange, and four ancestrally inherited chromosomes have
430 been involved in post-hybridization rearrangements. Given the efficiency of homologous

431 recombination in *Auxenochlorella* (see below), it is possible that the rapid emergence of
432 chromosomal differences between the closely related *A. protothecoides* and *A. symbiontica* may
433 have been driven by ectopic recombination events. Indeed, assuming a sexual life cycle for
434 *Auxenochlorella*, UTEX 250-A is presumably incapable of meiosis due to the non-homologous
435 nature of its chromosomes (Garagna et al. 2014), and the karyotypic variation between the two
436 species may present a barrier to gene flow.
437
438 **Pervasive loss-of-heterozygosity and evolution in the laboratory**
439 LOH events are associated with several mechanisms and can affect local genomic regions
440 (interstitial LOH), large regions that extend to a chromosome terminus (terminal LOH), or entire
441 chromosomes. We observed all three LOH categories in UTEX 250-A, resulting in at least
442 35.5% of the genome being homozygous (conservatively assuming a minimum interstitial LOH
443 length of 1 kb). The three entirely homozygous chromosomes, 4, 9 and 10, collectively span
444 28.2% of the genome. Terminal LOH events affect seven chromosomal termini (e.g., the left arm
445 terminus of chromosome A01, Figure 2A), ranging from ~12 kb to ~271 kb (mean 111 kb) and
446 collectively spanning 3.5% of the genome. We observed 278 putative interstitial LOH events, 36
447 of which exceed 5 kb in length, reaching a maximum of 37.1 kb. Notably, LOH is pervasive in
448 all of the sequenced *Auxenochlorella* genomes; 23.3% of the *A. symbiontica* CCAP 211/61
449 genome is homozygous, whereas levels of homozygosity reach 67.2% (UTEX 25) and 80.3%
450 (UTEX 2341) in *A. protothecoides*.
451
452 Since all the studied *Auxenochlorella* strains have been in culture for decades, we questioned
453 whether LOH could have accumulated in the laboratory. UTEX 25 is one of the oldest microbes
454 in culture, having been isolated in the late 19$^{th}$ century (Krüger 1894b; Krüger 1894a). CCAP
455 211/61 was isolated in 1982, and UTEX 2341 prior to 1984 (Seto et al. 1984). Conveniently,
456 many of these strains are maintained at different culture centers, including UTEX and CCAP
457 (Culture Collection of Algae and Protozoa). Although records are uncertain, a sister lineage of
458 UTEX 250 has likely been independently maintained for more than 70 years as CCAP 211/7C
459 (Figure 3A). To test whether LOH has occurred during culture, we obtained CCAP 211/7C and
460 performed high-coverage ONT sequencing.
461

462 Remarkably, LOH events affect only 19.7% of the CCAP 211/7C genome. Although
463 chromosome 10 is entirely homozygous in both UTEX 250-A and CCAP 211/7C, chromosomes
464 4 and 9 are heterozygous in CCAP 211/7C (Figure 3B, S6). Thus, two of the chromosome-scale
465 LOH events must have occurred during culture at UTEX or in our laboratory. Similarly, of the
466 seven UTEX 250-A terminal LOH events, only five are present in CCAP 211/7C (Figure S6).
467 For example, both terminal LOH events on chromosome A07 and the right arm terminal LOH
468 event on A08 are shared, whereas the event on the left arm terminus of A08 is unique to UTEX
469 250-A (Figure 3B). The same pattern is observed for interstitial LOH, for example, the largest
470 interstitial event in the genome (37.1 kb at ~1.66 Mb on chromosome A08) is shared, whereas
471 the second largest (29.2 kb, ~0.69 Mb on A07) is unique to UTEX 250-A (Figure 3B).
472 Interestingly, almost all of the strain-specific LOH events are unique to UTEX 250-A,
473 suggesting that the frequency of LOH during culture at CCAP is substantially lower.
474
475 These results support the rapid accumulation of LOH events under at least some laboratory
476 conditions, as has been observed experimentally in yeast (Sui et al. 2020; Dutta et al. 2021).
477 Contrastingly, the UTEX 250-A chromosomal rearrangements that likely followed hybridization
478 are also present in CCAP 211/7C (Figure S6), suggesting that these events occurred prior to
479 sampling of the original strain (or at least before the UTEX and CCAP lineages were
480 established). Although it is possible that some of the LOH events in UTEX 250-A and CCAP
481 211/7C were fixed by positive selection to overcome post-hybridization incompatibles, there is
482 no obvious directionality with respect to parental species: approximately equal numbers of bases
483 within regions affected by interstitial and terminal LOH events can be assigned to *A.*
484 *protothecoides* and *A. symbiontica*. Furthermore, LOH is even more prominent in the non-hybrid
485 *A. protothecoides* strains, suggesting that it is a widespread phenomenon in *Auxenochlorella*. It is
486 possible that the different frequencies of LOH events in CCAP 211/7C and UTEX 250-A could
487 be attributed to differences in culture conditions (e.g., liquid vs agar cultures that would affect
488 clonal population sizes). We recommend cryopreservation and avoiding bottlenecks in culture
489 sizes to avoid the fixation of LOH events by drift.
490
491
492

493 **Aneuploidy is widespread in *Auxenochlorella***

494 While assembling the UTEX250-A genome, we noticed that two genomic regions exhibited
495 unexpectedly high read coverage. Specifically, when mapping the PacBio reads against the
496 diploid assembly, the entirety of chromosome B03 and a 932 kb fragment of B05 exhibit median
497 read coverages of 528x and 519x, respectively, approximately twice the 272x coverage of the
498 other individual chromosomes, suggesting that these regions are trisomic when considering both
499 A and B haplotypes (Figure 4A). The putatively duplicated part of B05 features the right arm
500 terminus and associated telomere, and we identified reads featuring telomeric repeats that map
501 internally to B05 precisely at the location where coverage doubles (Figure S7). Conceivably,
502 B05 may have duplicated and then undergone fission, with the fragment retaining the centromere
503 acquiring a de novo telomere at its left terminus, yielding a partly aneuploid chromosome. We
504 compared read coverage between UTEX 250-A and CCAP 211/7C to explore whether the
505 observed aneuploidies could have evolved during laboratory culture. Surprisingly, B05 exhibits
506 normal coverage in the CCAP 211/7C ONT read dataset, whereas B03 has three times the
507 expected coverage (163x relative to 53x genome-wide, Figure 4B), suggesting tetrasomy for this
508 chromosome (three copies of B03 and one copy of the homologous region on A03; Figure S6).
509 Thus, the B05 duplication appears to have occurred during culture leading to UTEX 250-A (or
510 alternatively CCAP 211/7C has reverted to disomy), whereas aneuploidy of B03 may have
511 existed prior to sampling, with subsequent copy number change during culture.
512
513 We next checked for possible aneuploidies in *A. protothecoides* UTEX 25 and UTEX 2341, and
514 *A. symbiontica* CCAP 211/61, by mapping PacBio reads for each strain against one
515 representative haplotype from their respective genome assembly. For UTEX 2341 and CCAP
516 211/61, coverage estimates were consistent across all chromosomes, suggesting euploidy (Figure
517 S8). However, median read coverage for chromosome 1 (76x) of UTEX 25 was approximately
518 50% higher than median coverage across the rest of the genome (54x), implying trisomy (Figure
519 4C). We also noted that the existing *A. protothecoides* "0710" genome assembly exhibited
520 almost no variation relative to our UTEX 25 assembly, indicating that they are replicates of the
521 same strain. Surprisingly, when mapping the "0710" Illumina genomic reads to our UTEX 25
522 assembly we found no increase in coverage for chromosome 1, whereas coverage for
523 chromosomes 7 and 10 (131 and 134x, respectively) were ~50% higher than the median of 89x

524 across the rest of the genome (Figure 4C). Thus, it appears that the evolution of aneuploidy is
525 dynamic, occurring on short timescales of laboratory culture, at least in the UTEX 250 and
526 UTEX 25 genetic backgrounds. These results are in line with chromosome duplication events
527 observed in several haploid green algal species during mutation accumulation experiments
528 (Krasovec et al. 2023; Lopez-Cortegano et al. 2023). Transient aneuploidy may explain the rapid
529 emergence of whole-chromosome LOH, suggesting that it is a prominent evolutionary force in
530 *Auxenochlorella*. It is unclear whether any of the observed aneuploidies confer a selective
531 advantage; at least one additional copy of B03 has been maintained over decades of culture in
532 UTEX 250 and CCAP 211/7C, potentially indicating that it could be beneficial. There are 149
533 genes on B03, and analyses focusing on the effects of dosage on gene and protein expression will
534 be required to address this.
535
536 **Gene family reduction and loss contribute to the streamlined *Auxenochlorella* genome**
537 To facilitate the development of *Auxenochlorella* as a model system, we aimed to produce a
538 reference quality structural annotation for the UTEX 250-A genome. We sequenced PacBio
539 cDNA (IsoSeq) libraries from UTEX 250-A grown either autotrophically or with glucose, and
540 multiple Illumina RNAseq libraries from different growth conditions. Utilizing these datasets as
541 evidence, we annotated gene models independently for the A and B haplotypes using a
542 combination of gene predictors and extensive manual curation (see Methods). This yielded 7,509
543 and 7,515 protein-coding genes on the A and B haplotypes, respectively, with an additional 86
544 and 121 genes carried by transposable elements included in the annotation (Table 1). More than
545 two thirds of the gene models are derived from multiple full-length IsoSeq reads, enabling high
546 confidence identification of transcription start sites (TSS), terminators, and alternative splice
547 variants. Approximately 43% of genes are annotated with at least one alternative transcript,
548 totaling 13,125 and 13,345 transcripts on the A and B haplotypes, respectively. Considering the
549 A haplotype, the majority (56%) of alternative isoforms differ only in their UTR sequences (e.g.,
550 due to alternative TSSs or terminators) relative to another isoform of the same gene. For
551 alternative isoforms that alter the predicted protein sequence, 31% feature a retained intron and
552 18% an exon with an alternate 5' or 3' splice junction. The remaining 51% were classified as
553 complex isoforms, many of which are predicted to originate from internal promoters that produce
554 truncated proteins (see below). Alternative transcripts that skip exons are rare (<1%). Finally,
555 612 gene models were corrected based on manual curation, including 43 bicistronic loci analyzed

556 by Dueñas et al. (2025a), which are frequently misannotated by automated approaches (Gallaher
557 et al. 2021). Speaking to the compactness of the UTEX 250-A genome, 74% of genes encode a
558 protein with a Pfam domain (Dataset S6), a metric that is typically 25-50% in algal genomes
559 (Blaby-Haas and Merchant 2019). Via orthology to *C. reinhardtii* proteins, one-third of genes
560 were also assigned formal gene symbols (Dataset S7), which generally correspond to at least
561 some level of functional curation (Blaby et al. 2014; Craig et al. 2023).
562
563 Via synteny analysis, we were able to match ~98% of genes as allelic pairs between the A and B
564 haplotypes. Genes were given a unique ID featuring the haplotype and chromosome (e.g., A11,
565 B08) and a five-digit identifier that is shared between allelic pairs, e.g., despite being located on
566 rearranged chromosomes, UTEX250_A11.30955 and UTEX250_B08.30955 refer to alleles of
567 the same gene. As expected, heterozygosity between the A and B alleles at four-fold degenerate
568 sites in coding sequence is substantially higher (3.44%) than at zero-fold degenerate sites
569 (0.84%), although this nonetheless corresponds to more than 35,000 allelic variants that result in
570 amino acid differences. Heterozygosity is highest in intergenic (3.51%) and intronic (4.02%)
571 sites (Dataset S8). Only ~15% of the haplotype-specific genes encode a predicted protein with a
572 Pfam domain (Dataset S6), and many of these genes may be evolutionarily young or
573 misannotated long noncoding RNA (lncRNA) genes that carry a spurious open reading frame
574 (ORF). Some haplotype-specific genes are the result of structural variation between the A and B
575 haplotypes, including a few copy number and presence/absence variants.
576
577 Although we identified over 500 additional genes in our UTEX 250-A genome relative to the
578 existing *A. protothecoides* 0710 assembly, the *Auxenochlorella* gene number remains lower than
579 that of most trebouxiophyte species, and substantially lower than that of chlorophyceaen species
580 such as *C. reinhardtii* (Table 1). We performed an orthology analysis of UTEX 250-A proteins
581 against the best quality chlorophyte annotations, arbitrarily using the A haplotype genes
582 supplemented with genes unique to the B haplotype. Figure 1A presents a phylogeny based on
583 669 single-copy orthologs identified by this analysis, and Figure 5A shows the intersecting sets
584 of orthogroups among Chlorellales species, with *C. subellipsoidea* and *C. reinhardtii* as
585 outgroups. Ignoring species-specific orthogroups, the largest set corresponds to gene families
586 present in all species (2,974 orthogroups), many of which are presumably essential core-

587 Chlorophyte gene families (Dataset S9). The second largest set features gene families
588 specifically absent in the non-photosynthetic *P. cutis* (520 orthogroups), most of which are
589 expected to have functions related to photosynthesis (Suzuki et al. 2018). With respect to
590 reduced gene content in *Auxenochlorella*, 297 orthogroups are absent across all Chlorellales
591 species, and 201 orthogroups are specifically absent in *Auxenochlorella* UTEX 250-A and *P.*
592 *cutis* but present in all other species.
593
594 For more detailed analyses, we compared the *Auxenochlorella* UTEX 250-A annotation to the
595 high-quality gene models of *Chlorella vulgaris* (Cecchin et al. 2019) and *C. reinhardtii* (Craig et
596 al. 2023), excluding any genes carried by transposable elements. Considering gene family loss,
597 there are 782 orthogroups that are absent from UTEX 250-A but present in both other species,
598 featuring 1,025 *C. vulgaris* genes and 1,212 *C. reinhardtii* genes. Considering gene family size,
599 1,341 orthogroups are present in all three species and have more than one gene in at least one
600 species, featuring 2,228 UTEX 250-A genes, 2,915 *C. vulgaris* genes, and 3,272 *C. reinhardtii*
601 genes. Approximately 56% of these orthogroups feature a single UTEX 250-A gene, relative to
602 35% and 29% for *C. vulgaris* and *C. reinhardtii*, respectively (Figure 5B). UTEX 250-A genes
603 also contain fewer and shorter introns (mean 5.1 introns per gene, mean length 179 bp) relative
604 to *C. vulgaris* (means 7.3 and 206 bp) and *C. reinhardtii* (means 7.8 and 272 bp) (Figure 5C, D).
605 Thus, the streamlined *Auxenochlorella* UTEX 250-A genome can be explained by both complete
606 loss or lower complexity of gene families, alongside lower repeat content (Table 1) and the
607 presence of fewer and shorter introns.
608
609 Many of the lost genes can be linked to known metabolic pathways. Two distinguishing features
610 of *Auxenochlorella* are an absolute requirement for thiamine and the inability to grow on nitrate
611 as a N source (Kalina and Punčochárová 1987). As expected, the nitrate reductase gene *NIA1* and
612 *THIC1*, encoding 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate synthase necessary
613 for thiamine biosynthesis, are among the genes present in *C. vulgaris* and *C. reinhardtii* but
614 absent in UTEX 250-A. All other components of the thiamine biosynthesis pathway are present
615 in UTEX 250-A, enabling thiamine production via salvage of breakdown products. Similarly, as
616 reported by Gao et al. (2014), *Auxenochlorella* cannot utilize urea as an N source and genes
617 responsible for urea assimilation (*DUR1*, *DUR2*) and transport (*DUR3*) have been lost. Several

618 other transporters present in at least one copy in both *C. vulgaris* and *C. reinhardtii* are also
619 absent in UTEX 250-A, including the borate transporter *BOR1* and the *SLT*-family
620 sodium/sulfate co-transporters. Other transporter gene families feature a single copy in UTEX
621 250-A, including the *CTR*-type copper ion transporters (three genes in each of *C. vulgaris* and *C.*
622 *reinhardtii*) and the *PTB*-family sodium/phosphate symporters (four genes in *C. vulgaris*, at least
623 nine in *C. reinhardtii*).
624
625 The reduction in gene complexity can also be seen by BUSCO (Benchmarking Universal Single
626 Copy Orthologs) (Manni et al. 2021) analysis against the UTEX 250-A predicted proteins. Of the
627 1,519 chlorophyte BUSCO genes, 1,488 (98%) were identified as complete, with only nine
628 (0.6%) complete and duplicated. One protein was called as fragmented, although it was manually
629 confirmed to be a correct gene model. Of the 30 missing BUSCOs, eight could be manually
630 recovered by cross-checking against the orthogroup analysis. The remaining 22 genes were also
631 undetected when searching BUSCOs against the UTEX 250-A genome, suggesting they are
632 biologically absent and not misannotated (Dataset S10). Most notably, nine of the 22 missing
633 BUSCOs are part of GreenCut2, a set of almost 600 genes conserved among select
634 photosynthetic eukaryotes and absent from non-photosynthetic species (Karpowicz et al. 2011).
635 A specific analysis of genes involved in photosynthesis and sexual reproduction is presented
636 below.
637
638 We also analyzed a small number of gene families that have specifically expanded in UTEX
639 250-A. Import of glucose and galactose in *Chlorella kessleri* is facilitated by three genes
640 encoding H+/Hexose symporters (*HUP1, HUP2* and *HUP3*) (Stadler et al. 1995), and three
641 paralogous genes that are most similar to *C. kessleri HUP2* were identified in the *A.*
642 *protothecoides* 0710 genome (Gao et al. 2014). We identified five *HUP* genes on both
643 haplotypes of the UTEX 250-A genome, including a tandem array of three genes on
644 chromosome A12/B12 that is disrupted by contig breaks in the 0710 assembly that resulted in
645 misannotation. For orthogroups featuring genes from UTEX 250-A, *C. vulgaris* and *C.*
646 *reinhardtii* (Figure 5B), we identified 13 cases involving UTEX 250-A specific expansion via
647 multiple gene duplications (Dataset S11). A tandem array of four genes on A07/B12 encode
648 predicted ribonuclease (RNAse) T2 proteins, which have diverse functions including roles in

649 stress response and pathogen defense (MacIntosh 2011), although the function of the single
650 RNAse T2 gene from *C. reinhardtii* is unknown. A tandem array of six (A01) or five (B01)
651 genes encode homologs of *C. reinhardtii* ferrireductase *FRE1*, which reduces Fe(III) to Fe(II) for
652 Fe assimilation (Allen et al. 2007). Four of the expanded orthogroups are associated with
653 glycosylation. Three of these correspond to gene families (*HPAT*, *RRA* and xyloglucanase-113)
654 that are specifically involved in arabinose *O*-glycosylation of the Hyp-rich glycoproteins
655 (HRGPs) that form the vegetative and zygote cell walls in *C. reinhardtii* (Joo et al. 2017). With
656 respect to proteins potentially involved in cell wall formation, it is also notable that UTEX 250-A
657 carries multiple genes encoding polyketide synthases, as reported previously for *A.*
658 *protothecoides* 0710 (He et al. 2016; Heimerl et al. 2018). These include homologs of *PKS1*
659 from *C. reinhardtii*, encoding a type I polyketide synthase involved in formation of the zygote
660 cell wall (Heimerl et al. 2018), and *LAP5* from *Arabidopsis thaliana*, encoding a type III
661 polyketide synthase involved in the formation of sporopollenin (Kim et al. 2010; He et al. 2016).
662 Further functional characterization of these expanded gene families may provide insights into the
663 unique biology of *Auxenochlorella* species.
664
665 Of the 146 and 655 genes that are uniquely present in the AHP lineage or only in UTEX 250-A,
666 respectively, only 10% encode predicted proteins with an identifiable domain. Most notably, 61
667 haplotype A genes are present in five AHP or *Auxenochlorella*-specific orthogroups that are
668 associated with pentapeptide repeat domains, and a further 39 genes are present in a single
669 similarly associated orthogroup that otherwise features only one gene in each of the Chlorellales
670 species (including *P. cutis*). Genes encoding pentapeptide repeat proteins are abundant in many
671 bacterial genomes, particularly those of some cyanobacteria, although their functions remain
672 elusive (Vetting et al. 2006). Two highly conserved GreenCut2 proteins feature a pentapeptide
673 repeat domain of unknown function, CPLD59 and CPLD44, and are localized to the thylakoid
674 lumen (Kieselbach et al. 1998; Schubert et al. 2002). Some mycobacterial pentapeptide repeat
675 proteins adopt a fold that mimics double-stranded DNA and can compete with DNA in DNA-
676 protein interactions (Feng et al. 2021). Although their functions are unclear, these 100 genes are
677 dispersed throughout the UTEX 250-A genome, and their repetitive nature can apparently
678 mediate genomic rearrangements via ectopic recombination (Figure 2B, C).
679

680 ***Auxenochlorella* encodes a reduced set of proteins related to photosynthesis and sexual**
681 **reproduction**
682 ***The GreenCut***
683 Intrigued by the potential loss of core photosynthesis genes, we next queried the UTEX 250-A
684 and *C. vulgaris* annotated proteins against the entire GreenCut2 dataset, using *C. reinhardtii*
685 proteins as a reference set. GreenCut2 is divided into four categories: genes conserved among
686 green algae and land plants (CGL, i.e., the green lineage), the green lineage and the red alga *C.*
687 *merolae* (CPL), the green lineage and diatoms (CGLD), and all of these groups (CPLD). While
688 12.5% of the GreenCut2 genes were not found in *C. vulgaris*, 18.0% were absent in UTEX 250-
689 A, corresponding to the specific loss of 33 photosynthesis-related genes despite retaining
690 photosynthesis (Figure 5E; Dataset S12). For example, UTEX 250-A encodes a reduced
691 repertoire of light-harvesting complex proteins. Three of the *LHCA* genes that encode the
692 antenna system of photosystem I are absent, namely *LHCA4, LHCA5* and *LHCA6*, but are
693 present in all analyzed trebouxiophytes including *C. vulgaris* (Figure 5E; Dataset S13). This
694 suggests that the PSI-LHCI supercomplex of *Auxenochlorella* likely consists of a single LHCI
695 tetramer (with subunits LHCA1, LHCA8, LHCA7 and LHCA3) and one LHCI dimer (LHCA2
696 and LHCA9), as in *Dunaliella* species (class Chlorophyceae) (Caspy et al. 2020; Liu et al. 2025).
697 The *PSBS* and *LHCSR* genes, encoding light-harvesting complex-like subunits involved in
698 nonphotochemical quenching (Allorent et al. 2016), are also absent in *Auxenochlorella* UTEX
699 250-A but present in *C. vulgaris* (Cecchin et al. 2019), raising the question of how
700 *Auxenochlorella* dissipates excess excitation energy.
701
702 The absence of certain GreenCut2 genes in *Auxenochlorella* demonstrates that they are not
703 essential for photosynthesis despite their deep conservation in photosynthetic organisms.
704 Nevertheless, we note that with comparison restricted to *C. reinhardtii*, some false negatives are
705 expected. Small, divergent proteins with organellar targeting peptides would be especially
706 susceptible to exclusion in orthology analysis. One example is RBCX (Dataset S12), a protein
707 chaperone that is essential for proper assembly of the RuBisCo large subunit into the octameric
708 $RbcL_8$ intermediate, and subsequent association with RuBisCo small subunit to form the
709 holoenzyme (Saschenbrecker et al. 2007; Liu et al. 2010). Eukaryotic RBCX proteins are found
710 in two distinct isoforms; RBCX-I is most similar to cyanobacterial RbcX, and RBCX-II is more

711 divergent (Saschenbrecker et al. 2007; Bracher et al. 2015). While *A. thaliana* has both forms,
712 the *C. reinhardtii RBCX2A* and *RBCX2B* genes both encode RBCX-II proteins (Kolesiński et al.
713 2011; Bracher et al. 2015). We identified an ortholog of the *C. reinhardtii* RBCX genes in *C.*
714 *vulgaris,* but not in UTEX 250-A, despite RBCX chaperone function likely being indispensable.
715 Manual searches using the RBCX-I proteins from *A. thaliana* and *Synechocystis* sp. PCC 6803
716 did reveal a candidate protein (UTEX250_A12.36490, Figure S9A) featuring an RbcX-like
717 domain and a 56 amino acid N-terminus plastid transit peptide predicted by TargetP-2.0
718 (Almagro Armenteros et al. 2019). Interestingly, *C. vulgaris* also encodes a highly similar
719 protein, suggesting that both RBCX-I and RBCX-II isoforms are present in the species (as in *A.*
720 *thaliana,* Figure S9B), whereas UTEX 250-A has only RBCX-I and *C. reinhardtii* only RBCX-
721 II. Comparisons of monomer AlphaFold structure models of the putative RBCX-I proteins from
722 UTEX 250-A and *C. vulgaris* with the corresponding crystal structures from *Synechocystis* sp.
723 PCC 6803 (Tanaka et al. 2007) and *A. thaliana* (Kolesiński et al. 2013) confirmed structural
724 homology, supporting the contention that they perform the same RbcL chaperone function
725 (Figure S9C).
726
727 ***The CiliaCut and meiosis genes***
728 The putative hybrid origin of UTEX 250 suggests the presence of a cryptic life cycle stage in
729 *Auxenochlorella* species, which may include meiosis of otherwise vegetatively diploid cells and
730 the formation of gametes (or at least implies an asexual transition between diploid and haploid
731 cells). Direct evidence for sexual reproduction in the Trebouxiophyceae is scarce, although
732 Fučiková et al. (2015) identified a core set of nine meiosis genes that are mostly present in
733 trebouxiophyte genomes. Although *Chlorella* species have never been observed to form ciliated
734 gametes, Blanc et al. (2010) identified orthologs of *C. reinhardtii* "CiliaCut" genes, which are
735 conserved among ciliated organisms but absent from species without cilia (Merchant et al. 2007).
736 Specifically, *Chlorella variabilis* has genes encoding the outer and inner dynein arm proteins,
737 but lacks most or all of the genes encoding components of the intraflagellar transport particle,
738 radial spoke and central pair complex (Blanc et al. 2010). The cilia genes encoded by *C.*
739 *variabilis* significantly overlap with those present in the centric diatom *Thalassiosira*
740 *pseudonana* (i.e., the "CentricCut"), which produces male gametes with motile cilia (Moore et al.
741 2017). Cecchin et al. (2019) confirmed this result, finding an even more substantial intersect

742 between “MotileCut” cilia genes common to *C. vulgaris* and *T. pseudonana*, but absent in
743 *Caenorhabditis elegans* (which produces sensory cilia, but not motile cilia), suggesting that
744 *Chlorella* species may carry the necessary repertoire of genes to produce motile ciliated gametes.
745 Most recently, Gazquez et al. (2024) demonstrated that the ciliated life stage of the lichen-
746 forming *Trebouxia lynniae* corresponds to gametes, providing one of the few characterizations of
747 a sexual lifecycle in a trebouxiophyte alga.
748
749 Only 25.3% of CiliaCut genes from *C. reinhardtii* are present in the UTEX 250-A genome,
750 relative to 38.5% in *C. vulgaris* (Figure 5F). However, 58.3% of the genes that form the intersect
751 of the MotileCut and CentricCut are present in UTEX 250-A, relative to 66.7% in *C. vulgaris*,
752 corresponding to only four genes that are uniquely absent in UTEX 250-A (Dataset S14).
753 Fučiková et al. (2015) identified four of the nine core meiosis genes in the *A. protothecoides*
754 0710 genome, with *HOPP1*, *REC8*, *MER3*, *MSH4* and *MSH5* absent. We identified *HOPP1* and
755 *REC8* in UTEX 250-A (Figure 5F), suggesting that they were previously missed due to
756 misannotation. *MSH4* and *MSH5*, homologs of the bacterial mismatch repair *MutS* family, are
757 absent in some sexual species such as *Plasmodium vivax*, although *MER3*, encoding a helicase
758 involved in the formation of meiotic crossovers (Altmannova et al. 2023), is universally present
759 in sexual species (Fučiková et al. 2015). We also identified UTEX 250-A orthologs of two
760 critical genes for syngamy, *GEX1*, which functions in nuclear fusion (Ning et al. 2013), and
761 *HAP2*, which functions in gamete fusion (Fedry et al. 2017). Thus, *Auxenochlorella* species may
762 be able to reproduce sexually, possibly via the formation of ciliated gametes, although as with
763 photosynthesis the set of genes functioning in these processes has been substantially reduced.
764 Notably, our analyses confirm the prior observation that all trebouxiophytes have lost the
765 KNOX/BELL homeodomain transcription factors (Joo et al. 2018), typified by the *C. reinhardtii*
766 genes *GSP1* and *GSM1*, that regulate haploid-to-diploid transitions in Archaeplastida (Lee et al.
767 2008; Hisanaga et al. 2021; Hirooka et al. 2022). Thus, the molecular pathways underlying
768 putative sexual lifecycles in both haplontic and diplontic trebouxiophytes remain cryptic.
769
770
771
772

773 **Periodic cytosine and adenine methylation supports widespread internal antisense**
774 **transcription**
775 ***DNA methylation***
776 The DNA modification 5-methylcytosine (5mC) is widespread across all domains of life and has
777 several essential functions (Mattei et al. 2022), although relatively little is understood about
778 cytosine methylation in green algae. In *C. variabilis*, gene bodies are extensively methylated at
779 CpG dinucleotides, whereas promoter regions are hypomethylated (Zemach et al. 2010).
780 Although 5mC is generally restricted to centromeres and subtelomeres in *C. reinhardtii* (Lopez
781 et al. 2015; Chaux-Jukic et al. 2021; Craig et al. 2023), extensive gene body methylation was
782 recently shown for the chlorophyceaen alga *T. obliquus* (Biondi et al. 2024), suggesting that this
783 pattern may be widespread in chlorophytes. In the extremely compact genomes of prasinophytes
784 such as *Ostreococcus lucimarinus* and *Micromonas pusilla*, CpG dinucleotides in gene bodies
785 are also highly methylated, although 5mC occurs in periodic clusters that correspond to the
786 linker DNA between nucleosomes (Huff and Zilberman 2014). Notably, periodic adenine
787 methylation (6mA), mostly at ApT dinucleotides, was also discovered in *C. reinhardtii* (Fu et al.
788 2015). Periodicity of 6mA is also due to specific methylation of nucleosome linkers, although
789 unlike for 5mC, the adenine methylation is present around TSSs and not in the gene body.
790 Romero Charria et al. (2024) demonstrated that periodic adenine methylation of linker DNA
791 downstream of TSSs is also present in both prasinophytes and *C. variabilis*, as well as in other
792 distantly related taxa, suggesting that it is an ancient feature of eukaryotic genomes. The function
793 of 6mA in this context is not known; the biophysical properties of this modification may play
794 some role in promoting transcription (Bochtler and Fernandes 2021), and 6mA could also
795 facilitate the coordination of promoter chromatin marks such as H3K4me3 (Romero Charria et
796 al. 2024).
797
798 We used the ONT reads from CCAP 211/7C (equivalent to UTEX 250, Figure 3A) to call both
799 5mC at CpG sites and 6mA at ApT sites across the UTEX 250-A genome. Using only the high-
800 confidence TSSs from IsoSeq-based gene models, periodicity of both modifications was evident
801 (Figure 6A). Cytosine methylation is essentially absent at sites adjacent to the majority of TSSs,
802 and universal at the hypermethylated peaks within gene bodies (on average 5mC is called at
803 ~95% of CpG sites at the highest point of the peaks). Conversely, two periodic peaks of 6mA are

804 present downstream of the TSS, although far fewer ApT sites are methylated and the signal is
805 noisier between the peaks (Figure S10). Assuming that the peaks correspond to nucleosome
806 linkers, it appears that the first two linkers downstream of the TSS are specifically associated
807 with adenine methylation, whereas cytosine methylation is present at the third linker and reaches
808 a stable maximum at subsequent linkers. These patterns closely resemble those of the
809 prasinophytes, as opposed to *C. variabilis* where 6mA exhibits periodicity but 5mC is relatively
810 constant across the gene body (Romero Charria et al. 2024). Huff and Zilberman (2014)
811 proposed that 5mC periodicity contributes to nucleosome positioning and hypothesized that this
812 could be an adaptation to extremely compact nuclei and small cell sizes. Our results suggest this
813 unusual genomic architecture may have been present in the common ancestor of the core-
814 Chlorophyta (Trebouxiophyceae and Chlorophyceae in Figure 1A), or alternatively that it re-
815 evolved on the lineage leading to *Auxenochlorella*.
816
817 ***Antisense long non-coding RNAs***
818 Aside from the biological importance of periodic methylation, these signatures also provide
819 evidence for promoter regions, similar to H3K4me3 ChIP-seq datasets that only exist for a
820 limited number of green algae (Strenkert et al. 2022; Petroll et al. 2025). Figure 6B shows a
821 representative locus featuring four genes where CpG hypomethylation clearly corresponds to the
822 regions downstream of the IsoSeq-based TSSs. Notably, one of these genes
823 (UTEX250_A08.19925) is a lncRNA transcribed from a bidirectional hypomethylated promoter
824 located within the CiliaCut gene *FAP57* (UTEX250_A08.19920), which is essential for the
825 correct assembly of inner dynein arms (Lin et al. 2019). While manually curating genes in the
826 UTEX 250-A genome, we noticed many similar cases, which motivated a thorough search for
827 lncRNA genes that are supported by multiple IsoSeq reads (Table 1). After manual verification,
828 165 protein-coding genes were identified with internal bidirectional promoters that produced
829 antisense lncRNAs on at least one of the two haplotypes (Dataset S15). For genes such as
830 *FAP57*, it is unclear whether the truncated isoforms transcribed from the internal promoter could
831 be functional, while for many others the truncated isoforms clearly have no coding capacity. We
832 also identified 615 protein-coding genes that exhibit substantial overlap with an antisense
833 lncRNA (>30% of the transcript length) transcribed from an internal, or immediately adjacent,

834 unidirectional promoter. The annotated lncRNAs are expected to be polyadenylated based on the
835 IsoSeq protocol and ~59% are spliced.
836
837 To explore the potential functions of widespread antisense transcription in UTEX 250-A, we
838 analyzed the predicted functions of the genes with associated lncRNAs. Manual curation of
839 proteins encoded by genes with internal bidirectional promoters recovered potential functions
840 associated with DNA, such as repair, chromatin and replication, followed by functions related to
841 post-translational regulation, cilia, and the cell wall. A gene ontology (GO) enrichment analysis
842 focusing on the full set of 780 genes with evidence for antisense transcription supported these
843 observations. The most significant terms for biological processes were all related to DNA repair
844 and recombination, including the repair of double-strand breaks by homologous recombination
845 (Figure 6C, Dataset S16). Indeed, the list includes many fundamental DNA repair genes,
846 including *RAD51*, *RAD21*, the RECQ helicases *RECQ1* and *RECQ5*, and the DNA polymerases
847 *POLK1* and *POLQ2* (Dataset S15). For cellular compartments, the two most significant terms,
848 "condensed nuclear chromosome" and "cohesin complex", are both related to mitosis, meiosis
849 and recombination, whereas the next two significant terms are related to cilia. The most
850 significant molecular function term, "calcium ion binding" (Dataset S16), is also associated with
851 cilia genes. Indeed 14 of the CiliaCut genes (including *FAP57*) exhibit antisense transcription,
852 alongside five of the eight UTEX 250-A core meiotic and syngamy genes (Figure 5F),
853 representing a significant enrichment ($\chi^2$= 32.05, $p$ = 1.5x10$^{-8}$). The nuclear fusion gene *GEX1*
854 represents an extreme case where all of the IsoSeq reads at the locus are antisense (Figure S11).
855
856 Based on these results, we speculate that antisense lncRNAs may play a regulatory role in
857 *Auxenochlorella*. In synchronized cultures of *C. reinhardtii*, some DNA repair genes are
858 specifically upregulated at the transition from dark to light (e.g., the ribonucleotide reductase
859 subunit *RIR2L*, which features an antisense lncRNA in UTEX 250-A), suggesting a possible role
860 in photodamage response (Strenkert et al. 2019). Many human DNA repair genes are also
861 upregulated under genotoxic stresses such as UV radiation and ROS (Christmann and Kaina
862 2013). In *C. reinhardtii*, several of the core meiotic genes are among a small set of genes that
863 lack H3K4me3 at their TSSs during vegetative growth, supporting expression restricted to the
864 sexual cycle (Strenkert et al. 2022). It therefore seems plausible that genes overlapped by

865 antisense lncRNAs must be specifically translated at certain points of the cell and life cycles, or
866 under specific conditions. While this gene set is enriched for putative functions in DNA repair
867 and sex, we also observe this phenomenon for many genes that are predicted to have unrelated
868 functions, suggesting that this may be a general regulatory mechanism. Antisense lncRNAs can
869 regulate expression of associated protein-coding genes by several mechanisms (Werner et al.
870 2024). For example, in compact genomes such as *S. cerevisiae*, transcriptional interference can
871 occur via the physical collision of polymerase complexes on opposite strands (Prescott and
872 Proudfoot 2002; Hobson et al. 2012). Experimental characterization, including differential
873 expression analyses of the protein-coding and lncRNA gene pairs at these loci, will be required
874 to test this hypothesis. However, we are not aware of similar observations in other green algae,
875 suggesting widespread antisense transcription may have emerged as an unusual gene regulatory
876 mechanism during *Auxenochlorella* evolution. One notable observation is that the UTEX 250-A
877 genome encodes only 15 predicted F-box domain proteins, relative to 43 in *C. vulgaris* and 38 in
878 *C. reinhardtii* (based on the Pfam domains PF12937 and PF00646). F-box proteins function in
879 post-translational regulation as a component of the SCF (Skp, Cullin, F-box containing)
880 ubiquitin-ligase complex, with the F-box protein responsible for binding specific substrates
881 destined for proteolysis (Kipreos and Pagano 2000). Although currently speculative, it is possible
882 that transcriptional regulation of at least a subset of fundamental genes by lncRNAs emerged in
883 the compact *Auxenochlorella* genome in contrast to a decrease in regulation via protein
884 degradation.
885
886 **An *Auxenochlorella* genetic toolkit: reverse genetics, selectable markers, inducible**
887 **promoters, and fluorescent reporters for cellular localization**
888 In combination with the genomic resources presented thus far, nuclear gene targeting by
889 homologous recombination in *Auxenochlorella* facilitates reverse genetic analysis of biochemical
890 pathways and processes. To illustrate this capability, we targeted *CHL27*, encoding the
891 chlorophyll (Chl) biosynthesis enzyme Mg-protoporphyrin IX monomethylester (MgPMME)
892 aerobic cyclase, reasoning that loss of Chl would provide a clear visual phenotype, and that non-
893 photosynthetic mutants could be maintained heterotrophically on glucose in the dark. The UTEX
894 250-A *CHL27* locus is heterozygous and exemplifies the hybrid origin of the strain. The *CHL27-*
895 *1* allele on chromosome B12 is closely related to the corresponding heterozygous *CHL27* alleles
896 in UTEX 25, and the homozygous *CHL27* locus in UTEX 2341; conversely, *CHL27-2* on A07

897 has more polymorphisms in common with the heterozygous CCAP 211/61 *CHL27* alleles
898 (Figure 7A, S12A). *CHL27* polymorphisms cluster in the introns and in the flanking intergenic
899 region between the *CHL27* 3' UTR and the downstream gene, and all but one of the coding
900 sequence polymorphisms are synonymous, so six of the seven alleles encode identical
901 polypeptides (Figure S12A, B). A non-synonymous polymorphism in UTEX 2341 *CHL27*
902 substitutes Arg for Gly at position 33, but this change is in the region predicted to encode the
903 plastid transit peptide, so all of the alleles are expected to produce identical mature polypeptides
904 after chloroplast import and transit peptide cleavage (Figure S12B).
905
906 We designed UTEX 250 *CHL27-1* and *CHL27-2* targeting constructs to fully eliminate the
907 coding sequence of each allele (Figure S12C, D). Since thiamine auxotrophy is a characteristic of
908 all *Auxenochlorella* and *Prototheca* species (Pore 1972), we used *THIC* from *A. thaliana*
909 (Franklin et al. 2013; Moseley et al. 2024) as the marker for transformation targeting *CHL27-1*.
910 Neomycin phosphotransferase II (*nptII*), conferring resistance to G418 (Geneticin) (Franklin et
911 al. 2011; Moseley et al. 2024; Dueñas et al. 2025a), was the selectable marker for targeting
912 *CHL27-2*. We did not make allele-specific targeting constructs for UTEX 25, UTEX 2341 or
913 CCAP 211/61, reasoning that the UTEX 250 *CHL27* homology arms should have sufficient
914 sequence identity to enable recombination (Figure S12A). Representative *chl27-1::THIC*
915 transformants from UTEX 25, UTEX 250 and UTEX 2341 are thiamine prototrophs, while
916 *chl27-2::nptII* integration into UTEX 250 and CCAP 211/61 conferred resistance to G418
917 (Figure 7B). Integration of the constructs at the *CHL27* loci was confirmed by PCR amplification
918 of genomic DNA using primers flanking the homology arms (Figure S12E), and subsequent
919 sequencing of the PCR products. *CHL27-1* was disrupted in 8/11 thiamine prototrophic UTEX
920 250 transformants and 3/11 had the integrating construct mis-targeted to the *CHL27-2* allele,
921 while integration at the *CHL27-2* locus was confirmed in 9/9 G418-resistant UTEX 250
922 transformants. Genetic tractability appears to be a feature of the *Auxenochlorella* genus, as
923 evidenced by successful transformation and targeted gene replacement by homologous
924 recombination in both *A. protothecoides* and *A. symbiontica* strains, and allele-specific targeting
925 in the UTEX 250 hybrid. Integration into the nuclear genome occurs via homologous
926 recombination in almost all *Auxenochlorella* transformants, in contrast to 2% of transformants
927 reported for *Ostreococcus* (Lozano et al, 2014). We have documented a step by step guide to

928 transformation of *Auxenochlorella* using a lithium acetate/polyethylene glycol method on
929 protocols.io (Dueñas et al. 2025b).

930 Representative *chl27* loss-of-function mutants (*chl27-1::THIC*/*chl27-2::nptII*) were generated by
931 sequential targeting of the *CHL27* alleles. Double knockout strains were both thiamine prototrophs
932 and G418-resistant (Figure 7B). Heterotrophic growth was robust for all strains, but *chl27* mutants
933 bleached and were incapable of photoautotrophic growth on minimal media (Figure 7B).
934 Photoautotrophic growth of single allele knockouts was equivalent to the wild-type parents,
935 demonstrating that one copy of *CHL27* was sufficient to maintain adequate Chl for photosynthesis
936 under the growth conditions (Figure 7B). In agreement with the observations of Shihira-Ishikawa
937 and Hase (1964), wild-type UTEX 250-A and the *chl27* heterozygous mutants grown under
938 heterotrophic conditions with a high C/N ratio (2% glucose) contained minimal amounts of Chl,
939 with the yellow color attributed to xanthophyll pigments (Figure 7C, D). Partial greening was
940 observed in wild-type and single-allele knockout cultures grown with limiting glucose (0.5%) in
941 the dark, and full greening was stimulated by just 5 $\mu mol.m^{-2}.s^{-1}$ of white light (Figure 7C). In
942 contrast, representative *chl27* double knockout cultures accumulated a red pigment under the same
943 growth conditions (Figure 7C). Absorbance spectra from cell extracts of wild-type and single allele
944 knockout strains grown in the dark with 2% glucose had minor Chl peaks at 663 nm, and were
945 consistent with lutein and zeaxanthin as the major xanthophyll pigments (Figure 7D). Chl was not
946 detected in the *chl27* double knockouts, which instead had a major peak at 418 nm with minor
947 peaks at 551 and 590 nm (Figure 7D), indicative of Mg-protoporphyrin IX or MgPMME
948 accumulation (Tottey et al. 2003; Hollingshead et al. 2012). This is the expected result of blocking
949 the aerobic cyclase that converts red MgPMME to green divinyl protochlorophyllide (Chen et al.
950 2021). The photosensitive phenotype of *chl27* mutants (Figure 7B) is likely a consequence of
951 reactive oxygen species damage caused by the highly photoactive protoporphyrin pigments (Mock
952 2001).
953
954 To expand our capabilities for gene targeting and metabolic engineering, we developed additional
955 selectable markers for transformation (Figure 7E). Two additional selectable markers have been
956 described in *Prototheca moriformis*: *SUC2*, encoding secreted sucrose invertase from *S.*
957 *cerevisiae*, enables heterotrophic growth on sucrose (Franklin et al. 2011), whereas *MEL1*,
958 encoding secreted α-galactosidase from *Saccharomyces carlsbergensis*, enables heterotrophic

959 growth on melibiose (Franklin et al. 2013). We demonstrated that *Auxenochlorella* transformants
960 expressing *SUC2* or *MEL1* consume sucrose or melibiose, respectively, for heterotrophic growth
961 (Figure 7E, S13A). The secreted SUC2 and MEL1 glycosyl hydrolases act in a non-cell-
962 autonomous manner, making rigorous single-colony purification essential to avoid persistence of
963 untransformed parental cells. Another aminoglycoside 3'-phosphotransferase antibiotic resistance
964 gene, *aph7"*, from *Streptomyces hygroscopicus*, is commonly used as a *C. reinhardtii*
965 transformation marker (Berthold et al. 2002), and confers resistance to 300 µg/mL hygromycin B
966 to *Auxenochlorella* (Figure 7E, S13A). As was reported for *C. reinhardtii* (Loera-Quezada et al.
967 2016), phosphite oxidoreductase, encoded by *ptxD* from *Pseudomonas stutzeri* (Costas et al.
968 2001), enables mixotrophically grown *Auxenochlorella* to utilize phosphite as a P-source (Figure
969 7E, S13A). Together with *nptII* and *THIC*, described above, the four selectable markers illustrated
970 in Figure 7E can be used in serial transformations to generate homozygous knockout mutants at
971 up to three loci, or to make heterozygous mutations at six independent loci.
972
973 We next set out to establish fluorescent protein expression in *Auxenochlorella* for intracellular
974 localization studies. *GFP* and *GFP* fusions were driven by the RuBisCo small subunit (*RBCS1*)
975 promoter and targeted to the neutral *DAO1* locus (Dueñas et al. 2025a). Strains were cultured in
976 the light under mixotrophic conditions such that transgene expression was activated in green
977 cells. Chl and GFP fluorescence imaging by confocal microscopy showed that untargeted GFP
978 was located in the cytoplasm (Figure 7F, S13B). Fusion of GFP to the mitochondrial targeting
979 sequence of the OXA3 membrane insertase revealed a mitochondrial network encircling the
980 cytoplasm (Figure 7F, S13B), while the chloroplast ribosomal protein L35 (PRPL35) plastid
981 transit peptide targeted GFP to the chloroplast, as evidenced by co-localization of GFP
982 fluorescence and Chl autofluorescence (Figure 7F, S13B). These strains may serve as references
983 for the localization of cytoplasmic, mitochondrial and plastid-targeted proteins in future research.
984
985 Next, we asked whether one of the inferred trisomies in UTEX 250-A (Figure 4A) could be
986 confirmed by allele-specific targeting. To do so, we took advantage of the characteristic
987 metabolic switch of *Auxenochlorella*. The ammonium transporter gene *AMT1B*, which is highly
988 upregulated under N-depletion in heterotrophic growth in UTEX 25 (Yan et al. 2013), is located
989 on the right arms of chromosomes A03 and B03, with a putative third copy on the duplicate of

990 B03 (termed “C03”, Figure 8A). Since A03 and B03/C03 are heterozygous at this region, we
991 designed an allele-specific construct with homology arms that specifically corresponded to the
992 flanking sequences of the *AMT1B* allele of B03/C03 (termed *AMT1B-1*, Figure 8A). The
993 construct replaced the *AMT1B-1* ORF with an ORF encoding the Venus fluorescent protein,
994 which is under the control of the endogenous *AMT1B-1* promoter. As above, *nptII* served as a
995 transformation marker conferring resistance to G418. We designed allele-specific PCR primers
996 that targeted the left and right flanks of the endogenous *AMT1B* alleles, in addition to primer
997 pairs that targeted the construct, with one primer within the construct and the other in the
998 sequence flanking the homology arms (Figure 8A). Wild-type sequences for both *AMT1B* alleles
999 (*AMT1B-1* on B03 or C03, *AMT1B-2* on A03) could be amplified from wild-type UTEX 250-A
1000 and from three representative transformants, but the transformants also contained mutant *amt1B-*
1001 *1::Venus* integrations, consistent with a trisomic state featuring one copy of *AMT1B-2* and two
1002 copies of *AMT1B-1* in the wild-type nuclear genome (Figure 8B). We next compared Venus
1003 fluorescence between transformants and the UTEX 250-A wild-type negative control grown
1004 under photoautotrophic and heterotrophic conditions (Figure 8C, D). Since biomass
1005 concentration is higher in late log-phase heterotrophic cell cultures than in photoautotrophic
1006 cultures, the greater depletion of N from the medium leads to the activation of the N-deficiency-
1007 responsive *AMT1B* promoter. As a consequence, transformants displayed 19-fold higher Venus
1008 fluorescence in heterotrophy than in phototrophy, and on average 66-fold higher Venus
1009 fluorescence compared to the wild-type negative control (Figure 8C, D). We previously utilized
1010 the promoter of Photosystem I subunit D (*PSAD1*) to induce expression of a synthetic construct
1011 in phototrophy (Dueñas et al. 2025a). In addition to confirming trisomy, we here show that the
1012 *AMT1B* promoter represents a convenient counterpart for inducing expression in heterotrophy.
1013
1014 Finally, given the effectiveness of homologous recombination in *Auxenochlorella* species, we
1015 assessed the UTEX 250-A genome for genes involved in non-homologous end joining. All of the
1016 core non-homologous end joining genes (Chang et al. 2017) are present in UTEX 250-A and *C.*
1017 *reinhardtii*, namely Ku70 and Ku80, and DNA ligase IV (*LIG4*) and *XRCC4* (Dataset S17). A
1018 gene encoding the DNA dependent protein kinase catalytic subunit (DNA-PKcs) is present in *C.*
1019 *reinhardtii* but is absent in UTEX 250-A, as it also is absent in *A. thaliana* (Khan & Ochi 2023).
1020 DNA-PKcs appears to be absent in all Chlorellales genomes analyzed, and its absence is

1021 therefore unlikely to explain the effectiveness of homologous recombination in *Auxenochlorella*.
1022 Overall, we conclude that *Auxenochlorella* has the core complement of genes functioning in non-
1023 homologous end joining.
1024
1025 **DISCUSSION**
1026
1027 Species that reproduce clonally as diploids may experience specific phenomena including
1028 allodiploid hybridization, mitotic recombination, loss-of-heterozygosity and aneuploidy.
1029 Although these processes can act as strong evolutionary forces, for example in speciation or
1030 adaptation to new environments, most of our knowledge is presently based upon yeasts and
1031 filamentous fungi. In green algae, observations consistent with these phenomena have been
1032 reported sporadically, including trisomy in *C. primus* (Lemieux et al. 2019), and chromosome-
1033 scale LOH and possible hybridization in *T. obliquus* (Biondi et al. 2024). Here, we find that each
1034 of these phenomena occurs in the trebouxiophyte genus *Auxenochlorella*, demonstrating the
1035 generality of these important "yeast-like" molecular and evolutionary processes in vegetatively
1036 diploid eukaryotes.
1037
1038 *Auxenochlorella* UTEX 250 is an allodiploid hybrid of two closely related species, *A.*
1039 *protothecoides* and *A. symbiontica*, that are differentiated by extensive karyotypic variation and
1040 possibly ecological niches. Following hybridization, the parental haplotypes have presumably
1041 been shuffled by mitotic crossovers, and homogenized via LOH events mediated by
1042 recombination, DNA repair, and potentially transient aneuploidy. Post hybridization, the
1043 outcomes of these events can be driven by selection to resolve genomic incompatibilities
1044 between the divergent parental genomes in a process termed genome stabilization (Sipiczki
1045 2008; Gabaldón 2020). It is presently unclear to what extent the crossovers, LOH events and
1046 aneuploidies of UTEX 250 are involved in stabilization, although at least some of the LOH
1047 events that occurred prior to laboratory culture were presumably beneficial (e.g., at rDNA
1048 arrays). However, our observations from authentic strains of *A. protothecoides* and *A.*
1049 *symbiontica* suggest that LOH and aneuploidy are not just associated with hybridization: they are
1050 generally expected to be prominent evolutionary forces in *Auxenochlorella* that can occur on the
1051 timescale of laboratory culture. It is also unclear whether the sampling of UTEX 250 was a rare
1052 chance event, or whether it may represent a successful hybrid lineage that is persistent and

1053 potentially abundant (albeit likely sterile, assuming a sexual life cycle). Given the number of
1054 *Auxenochlorella* strains in culture, and the potential to isolate new strains (Asker and Awad
1055 2019; Chen et al. 2024), it would be interesting to apply population genomics approaches to
1056 investigate the prevalence of LOH, aneuploidy, and possibly other hybridization events, in the
1057 genus.
1058
1059 The resources and genetic toolkit presented here will serve as a foundation for functional
1060 genomics and systems biology analyses, as well as the design of transformation constructs for
1061 molecular manipulation of the genome. *Auxenochlorella* has great potential for development as
1062 an algal model for fundamental discovery research and bioengineering, and we hope that it will
1063 complement existing reference systems. As introduced, *C. reinhardtii* is the premier green alga
1064 for forward and reverse genetics, with well-developed classical haploid genetics (Harris 2001),
1065 established methods for transforming the nuclear, plastid and mitochondrial genomes (Boynton
1066 et al. 1988; Kindle et al. 1989; Kindle 1990; Randolph-Anderson et al. 1993; Shimogawara et al.
1067 1998), well-annotated genome assemblies (Merchant et al. 2007; Craig et al. 2023), a plethora of
1068 genetic markers and molecular components (promoters, UTRs, etc.) (Crozet et al. 2018),
1069 ribonucleoprotein (RNP)-mediated gene editing (Nievergelt 2025), and extensive community
1070 resources including wild-type and mutant collections (Li et al. 2016; Li et al. 2019). Nonetheless,
1071 two major hurdles have inhibited routine transformation of the *C. reinhardtii* nuclear genome:
1072 the essentially random integration of transgenes by non-homologous end joining, and the
1073 transcriptional silencing of transgenes (Schroda 2019; Nievergelt 2025). Recently, RNP-
1074 mediated gene-editing has been combined with scar-less homology-directed repair for transgene
1075 insertion at specific loci (Ferenczi et al. 2017; Akella et al. 2021; Nievergelt et al. 2023),
1076 substantially reducing position effects on transgene expression compared to un-targeted
1077 integration, and improving the proportion of transformants with seamless insertions at the
1078 recombination site up to 27% (Jacobebbinghaus et al. 2025). Mutant strains that robustly express
1079 transgenes have also been characterized (Neupert et al. 2020; Schroda and Remacle 2022).
1080
1081 Despite these advances, targeting of the *Auxenochlorella* nuclear genome by homologous
1082 recombination provides several advantages. Transformation only requires linear double-stranded
1083 DNA, in contrast to endonuclease-based methods that also depend on guide RNAs and DNA

1084 repair templates. Homologous recombination is not constrained by the requirement for double-
1085 strand breaks adjacent to Protospacer Adjacent Motif (PAM) sequences and thus provides greater
1086 flexibility in integration sites. Efficient integration by homologous recombination appears to be a
1087 general property of wild-type *Auxenochlorella* strains, and extends to species from the related
1088 genus *Prototheca* (Franklin et al. 2013; Moseley et al. 2024). Transgene integration occurs at
1089 high frequency via homologous recombination in the red algae *C. merolae* and *Galdieria partita*
1090 (Minoda et al. 2004; Fujiwara et al. 2013; Hirooka et al. 2022), but among green algae appears to
1091 be unique to *Auxenochlorella* and *Prototheca*. We have observed that the majority of
1092 *Auxenochlorella* transformants, typically 70-100%, exhibit precise, allele-specific integration at
1093 the targeted locus (Figure 7, 8). The upper limit for deletions is not yet defined, but in this work
1094 homology arms that are separated by more than 2 kb were used to delete the entire CDS of
1095 *CHL27* and *AMT1B* (Figure 7, 8, S12). Moreover, large cassettes containing up to four transgene
1096 expression modules, each with its own promoter, CDS and UTRs, have been integrated
1097 successfully (Bhat et al. 2025). However, we presently lack the ability in *Auxenochlorella* to
1098 simultaneously target multiple loci or both alleles of the same locus in a single transformation –
1099 an advantage offered by RNP-mediated approaches. Ideally, both methods would serve
1100 complementary roles depending on the application.
1101
1102 Beyond the versatility of genetic manipulation, several other features make *Auxenochlorella* an
1103 attractive reference system. Gene families containing multiple members in green algae, and
1104 especially in *C. reinhardtii*, more frequently feature just a single gene in *Auxenochlorella* UTEX
1105 250-A. The reduced genetic redundancy is expected to improve the likelihood of achieving
1106 informative phenotypes following gene knockout or knockdown, analogous to the lower
1107 redundancy of regulatory gene families in the liverwort *Marchantia polymorpha* relative to *A.*
1108 *thaliana* and other angiosperms (Bowman et al. 2017). Although hundreds of genes and gene
1109 families have been entirely lost, including many genes that are expected to play roles in
1110 photosynthesis, these genes could be re-introduced to the *Auxenochlorella* genome to probe their
1111 functions in the species where they do occur. Mutations in genes involved in photosynthesis,
1112 such as *CHL27* demonstrated herein, are also possible due to the facultative autotrophy of the
1113 species. The metabolic switch of the organism enables chloroplast biogenesis to be studied, and
1114 lends itself to the synthetic use of promoters induced in phototrophy or heterotrophy. Genomic

1115 manipulation also presents opportunities to address fundamental questions related to genetic
1116 mechanisms, as demonstrated by the dissection of the mechanism underlying the translation of
1117 bicistronic genes (Dueñas et al. 2025a). *Auxenochlorella* species are highly oleaginous and have
1118 GRAS status, and the genomic features reported here could also find utility in bioengineering.
1119 The introduction of artificial chromosomes is a powerful technique that enables multiple traits to
1120 be stacked at a single locus (Birchler et al. 2024). If the short AT-rich regions present on
1121 *Auxenochlorella* chromosomes can be confirmed to function as centromeres, it may be possible
1122 to maintain minichromosomes by introducing an AT-rich region, similar to the synthetic
1123 episomes that can be maintained in diatoms (Diner et al. 2017). Alternatively, the aneuploidy of
1124 UTEX 250-A and UTEX 25 could be exploited. We demonstrated allele-specific knock-in to a
1125 trisomic chromosome, and presumably it would be possible to sequentially manipulate this
1126 chromosome with minimal impact on fitness. Finally, since *Auxenochlorella* species are
1127 generally transformable, including the hydra symbiont *A. symbiontica* CCAP 211/61, the
1128 genetics of host/symbiont establishment and maintenance could be studied (Huss et al. 1994).
1129 Collectively, the yeast-like evolutionary biology and genetic manipulation of *Auxenochlorella*,
1130 thus far unique among green algae, presents many exciting prospects for discovery in plant
1131 biology and utilization in bioengineering.
1132

1133 **METHODS**

1134 **Strains, media and growth conditions**

1135 *Auxenochlorella* strains UTEX 250, UTEX 25 and UTEX 2341 were originally obtained from
1136 the University of Texas Culture Collection of Algae. A single colony of UTEX 250 was isolated
1137 and used in all experimental work (i.e., UTEX 250-A) except for *CHL27* gene targeting, which
1138 was carried out in the original UTEX 250 strain. *A. symbiontica* CCAP 211/61 and
1139 *Auxenochlorella* CCAP 211/7C were obtained from the Culture Collection of Algae and
1140 Protozoa. Strains were typically grown in TAP (for DNA extraction) or HPv1 (for RNA) culture
1141 media using the trace element recipe of Kropat et al. (2011). HPv1 is a modified version of TP
1142 that replaces 20 mM Tris with 20 mM HEPES. Standard liquid cultures were generally grown
1143 under constant light (100 µmol photons $m^{-2}$ $sec^{-1}$) and shaken at 200 rpm. Specific growth
1144 conditions were used for preparing RNA from UTEX 250-A for the generation of IsoSeq and
1145 RNAseq datasets. These included a range of photoautotrophic, mixotrophic or heterotrophic

1146 conditions, different light conditions, $CO_2$ levels, and elemental adjustments to media (Dataset
1147 S18).
1148
1149 **Extraction and sequencing of nucleic acids**
1150 UTEX 250-A DNA was extracted at UC Davis DNA Technologies Core from a frozen cell pellet
1151 and used to prepare and sequence PacBio HiFi and OmniC linked-read libraries. Individual
1152 IsoSeq libraries were produced from RNA extracted from cultures grown in autotrophic replete
1153 and heterotrophic with 2% (w/v) glucose conditions (HPv1 media), and each library was
1154 sequenced on a Sequel II SMRT cell at UC Davis. For all remaining conditions (Dataset S18),
1155 Illumina RNAseq with poly(A) selection was performed on a Novoseq 6000 platform at UC
1156 Davis, generating stranded 150 bp paired-end reads.
1157
1158 UTEX 25, UTEX 2341, CCAP 211/61 and CCAP 211/7C were cultured mixotrophically in TAP
1159 medium. High molecular weight DNA was extracted using a modified version of a CTAB and
1160 Phenol:Chloroform protocol (Camacho et al. 2025). Cell pellets were ground in liquid nitrogen
1161 in a mortar and pestle prior to the CTAB incubation. Size selection was performed using the
1162 PacBio short read eliminator (SRE) kit following user instructions.
1163
1164 DNA from UTEX 25, UTEX 2341 and CCAP 211/61 was prepared in a multiplexed library
1165 along with five other unrelated samples and sequenced on a single PacBio Sequel II SMRT cell
1166 at QB3 Genomics, UC Berkeley. ONT sequencing for CCAP 211/61 and CCAP 211/7C was
1167 performed using the Ligation Sequencing Kit V14, two independent R10.1.4 flow cells, and a
1168 MinION Mk1B device following user instructions.
1169
1170 **Genome assembly of *Auxenochlorella* UTEX 250-A**
1171 The UTEX 250-A PacBio HiFi reads were first subsampled to remove any reads shorter than 18
1172 kb, resulting in ~0.98 Gb of total sequence. An initial phased diploid assembly was produced by
1173 passing the filtered HiFi reads and Omni-C reads to Hifiasm v0.16.1-r375 (Cheng et al. 2022),
1174 which was run with default parameters. The initial assembly consisted of two sets of phased
1175 contigs, haplotype 1 and 2, featuring 17 and 21 nuclear contigs, respectively.
1176

1177 Several steps of manual assessment and assembly were then performed to achieve a final
1178 telomere-to-telomere phased diploid assembly. First, haplotype 1 and haplotype 2 contigs were
1179 aligned to each other using minimap2 v2.23-r1117 (Li 2021) with the parameter "-x asm20". The
1180 raw PacBio HiFi reads were also mapped against each haplotype assembly with minimap2 ("-x
1181 map-hifi"). Assembly and read-based alignments were manually inspected using IGV v2.16.2
1182 (Robinson et al. 2011). Due to the rearranged nature of the parental haplotypes, we found that the
1183 two automated haplotype sets did not individually represent a complete copy of the haploid
1184 genome (i.e. some genomic regions were present twice in one haplotype assembly, and absent
1185 from the other). We therefore transferred contigs between the two haplotype assemblies to arrive
1186 at two sets of contigs that each represented a complete haploid genome, which were arbitrarily
1187 labelled haplotype A and B.
1188
1189 Via comparison of the haplotype A and B assemblies it was possible to identify contigs that
1190 could potentially be fused i.e. at locations where the ends of two contigs from one haplotype
1191 mapped to a single assembled region from the other haplotype. Based on manual inspection of
1192 the raw reads, four pairs of contigs were successfully fused. This process involved either
1193 removing redundant sequence from one of the contig ends, or adding additional sequence that
1194 filled the gap between the two contigs. Where gap filling was required, haplotype-specific raw
1195 reads spanning the gap were extracted, aligned with MAFFT v7.490 (Katoh and Standley 2013),
1196 and reduced to a single consensus sequence that was subsequently trimmed and inserted between
1197 the two contigs. Following a similar approach, the termini of several chromosomes were
1198 manually extended to reach the telomeric repeats. Uniquely mapped reads that exhibited soft-
1199 clipped bases extending beyond the assembled chromosome were extracted, aligned, and reduced
1200 to a consensus sequence that was trimmed and appended to the chromosome. These manual steps
1201 resulted in the A and B haplotypes each being represented by 12 gapless chromosomes that
1202 either terminated in telomeric repeats or rDNA arrays.
1203
1204 Two additional contigs represented the third copies of the trisomic chromosomes, C03 and C05.
1205 The contig corresponding to chromosome C05 was manually truncated at the location of the
1206 internal telomere (see Figure S7), since the original contig was misassembled as an exact
1207 duplicate of chromosome B05. Finally, eight short contigs were removed from the assembly; six

1208 contigs entirely featured rDNA that is already represented as truncated arrays on the assembled
1209 chromosomes, and two contigs were identified as duplicate redundant sequences that are
1210 similarly represented in the chromosomal assembly. The assembly and phasing of the entire final
1211 assembly was manually verified by examining read alignments to each haplotype.
1212
1213 The plastome and mitogenome were each identified as single linear contigs in the Hifiasm
1214 assembly. Circularization was performed manually by identifying and removing redundant
1215 sequence from one end of each of the organellar contigs.
1216
1217 **Oxford Nanopore basecalling and methylation analyses**
1218 ONT raw reads derived from CCAP 211/61 were basecalled using Dorado v0.3.2
1219 (https://github.com/nanoporetech/dorado), which was run in super accuracy basecalling mode
1220 using the command "basecaller" and the model "dna_r10.4.1_e8.2_400bps_sup@v4.2.0". ONT
1221 reads for CCAP 211/7C were basecalled using Dorado v0.6.0 using the options
1222 "sup,5mCG_5hmCG" or "sup,6mA" to perform basecalling in super accuracy mode with the
1223 detection of 5mC and 5-hydroxymethylcytosine (5hmC) calls at CpG dinucleotides, or 6mA
1224 calls, respectively. The resulting BAM files were converted to FASTQ using samtools 1.18-1
1225 (Danecek et al. 2021) using the "-T '*'" option to retain base modifications. The resulting
1226 FASTQ files were mapped against the UTEX 250-A assembly using minimap2 with
1227 recommended parameters for ONT 10.4.1 chemistry ("-x map-ont -k19 -w19 -U50,500 -g10k -a
1228 -y"). The 5mC status of each CpG site was then called using modkit v0.2.7
1229 (https://github.com/nanoporetech/modkit) with the command "pileup --preset traditional", which
1230 aggregates methylation calls on both strands and returns only 5mC information. The 6mA status
1231 of each ApT site was similarly called using the command "pileup --motif AT 0". Sites with less
1232 than 5 mapped reads were removed. For each site, the fraction of modified bases at each site was
1233 then extracted, and the median for each nucleotide position relative to the TSS was calculated
1234 across all genes that were predicted using IsoSeq data (see below).
1235
1236
1237

1238 **Genome assembly of *Auxenochlorella protothecoides* UTEX 25 and UTEX 2341, and**
1239 ***Auxenochlorella symbiontica* CCAP 211/61**
1240 The UTEX 25, UTEX 2341 and CCAP 211/61 genome assemblies were produced using Hifiasm
1241 v0.19.8-r603. For UTEX 25 and UTEX 2341, all PacBio HiFi reads were passed to Hifiasm,
1242 which was run with default parameters. For CCAP 211/61, PacBio HiFi reads were
1243 supplemented with ONT reads, which were passed to Hifiasm with the flag "--ul".
1244
1245 When run without linked-read data (e.g. Omni-C), Hifiasm produces a primary haploid
1246 assembly, as well as a diploid assembly consisting of two partially phased haplotypes. For each
1247 strain, we first mapped the two (pseudo)haplotypes of the diploid assembly to each other using
1248 minimap2 ("-x asm20") and manually inspected the alignments to confirm that all contigs
1249 exhibited a one-to-one syntenic relationship between the two haplotypes (i.e. they are
1250 homologous without interchromosomal rearrangements as observed in UTEX 250-A). We then
1251 produced a representative haploid assembly by mapping the primary contigs to the UTEX 250-A
1252 genome assembly with minimap2 ("-x asm20") and manually inspecting alignments. For CCAP
1253 211/61, 14 primary contigs corresponded to 12 nuclear chromosomes and two circular organellar
1254 genomes. For both UTEX 25 and UTEX 2341, one pair of nuclear contigs and one pair of
1255 plastome contigs were each fused to produce the final assemblies, with all other major contigs
1256 corresponding to near-complete chromosomes. Contig fusion and circularization of the
1257 organellar contigs was performed using the manual assembly approaches described above. As
1258 with the UTEX 250-A assembly, additional short contigs featuring rDNA or redundant sequence
1259 were removed.
1260
1261 **Quantification of heterozygosity and divergence among strains**
1262 To quantify heterozygosity (i.e. individual-level genetic diversity) between the A and B
1263 haplotypes of the UTEX 250-A assembly, the haplotype A and B chromosomes were aligned
1264 against each other using minimap2 ("-x asm20"). Variants segregating between the A and B
1265 haplotypes were then identified using the minimap2 script paftools.js and the parameters "call -l
1266 2000 -L 10000" (i.e., a minimum alignment length to compute coverage of 2 kb, and a minimum
1267 alignment length to call variants of 10 kb). Chromosomes were divided in to 10 kb windows, and
1268 heterozygosity was calculated based on the number of single nucleotide variants (mapping

1269 quality of 60) relative to the number of alignable sites per window. Windows with <2000 aligned
1270 sites were excluded. Zero-fold, two-fold and four-fold degenerate sites were extracted from the
1271 coding sequence of the primary transcript of each A haplotype gene model using the degenotate
1272 tool (https://github.com/harvardinformatics/degenotate). All overall heterozygosity values were
1273 reported after removing homozygous genomic tracts that are putatively the product of LOH (see
1274 below).
1275
1276 Since the UTEX 25, UTEX 2341, and CCAP 211/61 genomes were assembled to the haploid-
1277 level, a different approach was used to quantify heterozygosity. Raw PacBio HiFi reads for each
1278 strain were mapped against the A haplotype of the UTEX 250-A assembly using minimap2 ("-x
1279 map-hifi"). DeepVariant v1.6.0 (Poplin et al. 2018) was then used to call variants directly from
1280 the PacBio HiFi read alignments for each strain independently, using the parameter "model_type
1281 PACBIO". Heterozygosity was calculated in 10 kb windows by identifying heterozygous single
1282 nucleotide variants segregating among the reads of the strain in question. Callable sites for each
1283 strain were determined by mapping the haploid genome assembly of the relevant strain against
1284 the A haplotype of UTEX 250 with minimap2 ("-x asm20") and identifying alignment blocks
1285 with paftools.js ("call -l 2000 -L 10000"). Genetic divergence for a given strain was calculated in
1286 10 kb windows as the proportion of variant sites relative to the UTEX 250-A haplotype A.
1287 Divergence was averaged over both haplotypes for each strain (i.e., a site where only one of the
1288 two haplotypes varied relative to the UTEX 250-A reference contributed half as much to
1289 divergence as a site where both haplotypes varied). Divergence was also calculated relative to
1290 the UTEX 250 B haplotype using the same approach.
1291
1292 Neighbor joining trees for nuclear genome, plastome and mitogenome sites were produced using
1293 MEGA v11.0.13 (Tamura et al. 2021) with the Tamura Nei substitution model and default
1294 parameters. For the organelles, either the plastome or mitogenome assembly of each strain was
1295 aligned to the UTEX 250-A plastome or mitogenome assembly with minimap2 ("-x asm20"),
1296 and variant and invariant sites were called with paftools.js ("call -l 2000 -L 10000"). To be
1297 included in the nuclear genome analysis, a site had to meet two criteria, i) to be on a UTEX 250-
1298 A chromosome that had not undergone any reciprocal crossovers, and ii) to be in a region that
1299 was not affected by a LOH event in any of the strains. These criteria were enforced to ensure that

1300 ancestral evolutionary relationships among haplotypes were recovered, and resulted in the
1301 analysis of 651,399 sites on chromosome A11. Variants were extracted from either the
1302 paftools.js analysis (for variants segregating between the UTEX 250 A and B haplotypes) or the
1303 DeepVariant analyses (for variants segregating within and between the other strains). Variants in
1304 UTEX 25, UTEX 2341 and CCAP 211/61 were phased relative to their respective genome
1305 assemblies (note that these may represent pseudohaplotypes with a small number of phase
1306 changes per chromosome).
1307
1308 **Loss-of-heterozygosity**
1309 LOH events in the UTEX 250-A genome were called from the alignment of the A and B
1310 haplotype assemblies and paftools.js analysis (see above). LOH events were arbitrarily assigned
1311 to regions that featured no heterozygous sites over a stretch of 1 kb or more. Events were
1312 designated as terminal if they extended to a chromosome end, or interstitial if they fell within a
1313 chromosome.
1314
1315 LOH events were similarly called as invariant regions of at least 1 kb for UTEX 25, UTEX 2341
1316 and CCAP 211/61 based on heterozygous variants identified from the DeepVariant analysis.
1317 CCAP 211/7C heterozygous sites and LOH events were also identified using DeepVariant, ran
1318 with the parameter "model_type ONT_R104" and supplied with an alignment of the raw ONT
1319 reads against the UTEX 250-A haplotype A assembly (minimap2 "-x map-ont"). These
1320 alignments were manually inspected to confirm that the karyotype of UTEX 250-A and CCAP
1321 211/7C are identical.
1322
1323 **Aneuploidy**
1324 The raw reads for all available strains were first mapped against the relevant genome assembly to
1325 calculate coverage. For PacBio datasets, the UTEX 250-A reads were mapped against the phased
1326 UTEX 250-A assembly, whereas UTEX 25, UTEX 2341 and CCAP 211/61 reads were mapped
1327 against the representative haplotype from their respective haploid assemblies, all using minimap2
1328 ("-x map-hifi"). CCAP 211/7C ONT reads were mapped against the UTEX 250-A assembly
1329 using minimap2 ("-x map-ont"), and the "0710" reads were mapped against the UTEX 25

1330 assembly using bwa mem (Li 2013). The coverage at each nucleotide was calculated using the
1331 samtools "depth" command, and the average coverage per 20-kb sliding window was calculated.
1332
1333 ***Auxenochlorella* UTEX 250-A gene model annotation**
1334 Gene models were annotated using a hybrid approach that utilized both the IsoSeq and RNAseq
1335 datasets, followed by extensive manual curation. The A and B haplotypes of the UTEX 250
1336 assembly were annotated independently, except for the entirely homozygous chromosomes (04,
1337 09 and 10).
1338
1339 IsoSeq reads were pre-processed using the IsoSeq3 pipeline v3.8.1
1340 (https://github.com/PacificBiosciences/IsoSeq). Primers and poly(A) tails were first removed
1341 from circular consensus sequencing (CCS) reads. The resulting reads from the two libraries were
1342 then combined and mapped against the UTEX 250-A genome assembly using minimap2 ("-x
1343 splice:hq"). These read mappings were used to partition the reads into sets that either primarily
1344 mapped to the A haplotype (≥20 mapping quality), primarily mapped to the B haplotype, or
1345 mapped to neither with high quality (<20 mapping quality; this set was enriched for reads
1346 derived from homozygous genomic regions). Read clustering was then performed individually
1347 on each read set ("isoseq3 cluster –use-qvs"), to avoid clustering reads that derived from
1348 different alleles of the same gene. The three clustered read sets were then re-combined and
1349 aligned collectively to the UTEX 250-A assembly using pbmm2 v1.9.0 ("--preset ISOSEQ").
1350 Unique isoforms were called using the isoseq3 collapse command with the options "--do-not-
1351 collapse-extra-5exons --max-5p-diff 20 --max-3p-diff 20". These options were selected
1352 following manual assessment of the quality of the IsoSeq data and effectively enabled alternative
1353 TSSs (within a different exon of a longer isoform, or >20 bp away if on the same exon) and
1354 termination sites (>20 bp if on the same exon) to be preserved as independent isoforms.
1355
1356 Gene models were predicted from the isoseq3 output using SQANTI3 v5.1 in QC mode with
1357 default parameters (Pardo-Palacios et al. 2024), which utilizes GeneMarkS-T (Tang et al. 2015)
1358 to predict ORFs from isoforms. The isoseq3 pipeline groups neighboring genes under a single
1359 gene model if any of their isoforms overlap on the same strand, which is a frequent occurrence in
1360 compact genomes such as that of *Auxenochlorella*. We thus had to "decouple" the isoforms of

1361 incorrectly merged genes based on comparison of the protein-coding coordinates of each isoform
1362 (i.e. isoforms which did not share coding sequence were separated to independent gene models).
1363 Next, we removed any isoforms that were supported by <5 reads or that were derived from <5%
1364 of the total reads associated with a given gene. This step minimized the number of false isoforms
1365 derived from 5' truncated reads. Finally, the primary isoform for each gene was selected based
1366 on a combination of read abundance and length criteria. For the majority of genes with multiple
1367 isoforms, the isoform with the longest coding sequence was also the most abundant and was
1368 deemed the primary isoform (where multiple isoforms had the same coding length but differed in
1369 their UTRs, the most abundant was selected). For genes where the longest coding sequence was
1370 not the most abundant, it was deemed the primary isoform if it was supported by ≥50% of the
1371 reads associated with the most abundant isoform. If this was not the case, the most abundant
1372 isoform was selected if its coding length was ≥80% of the isoform with the longest coding
1373 length. In all other cases, the longest isoform was selected to avoid overly short isoforms being
1374 selected. since they may be overrepresented in IsoSeq data.
1375
1376 We noticed that many ORFs were truncated due to predicted initiation at an internal ATG codon.
1377 To perform ORF extension, we first built a model of the Kozak sequence surrounding annotated
1378 initiation sites from primary transcripts that had no in-frame upstream ATG codons (i.e. ORFs
1379 that could not be extended), based on the assumption that most of these annotations represented
1380 true initiation sites. The strength of the Kozak sequence surrounding in-frame upstream ATG
1381 codons for the remaining transcripts was then calculated by comparison to this Kozak model (see
1382 Cross (2015) and Dueñas et al. (2025a) for details). ORF extension was performed if an
1383 upstream in-frame start codon existed with a Kozak score greater than the 25$^{th}$ percentile of the
1384 scores from which the Kozak model was built, or if the upstream Kozak score was simply higher
1385 than that of the annotated start codon.
1386
1387 In parallel to the above annotation, we used high confidence gene models derived from the
1388 photoautotrophic IsoSeq dataset to train the AUGUSTUS v3.3.2 gene predictor tool (Stanke et
1389 al. 2008), following the protocols described by Hoff and Stanke (2019). Pre-processing and gene
1390 model annotation using isoseq3 and SQANTI3 was performed as above. The training set of gene
1391 models was produced by extracting the most abundant isoform supported by a minimum of 10

1392 reads from gene models derived from IsoSeq reads that mapped uniquely to the A haplotype.
1393 Accurate AUGUSTUS training requires information on the noncoding DNA flanking the
1394 training set genes. To avoid including unannotated genes in the flanking regions, we performed a
1395 preliminary de novo gene annotation using BRAKER v2.1.2 (Hoff et al. 2016). Each RNAseq
1396 dataset was mapped against the UTEX 250-A assembly using STAR v2.7.10a ("--twopassMode
1397 Basic --alignIntronMax 5000") (Dobin et al. 2013), and subsequently passed to the braker.pl
1398 script using default parameters. We then extracted BRAKER gene models that had no overlap
1399 with the IsoSeq-derived training gene set using bedtools v2.30.0 (Quinlan and Hall 2010). The
1400 combined dataset of IsoSeq training genes and nonoverlapping BRAKER genes was then used to
1401 compute the flanking region length (1,366 bp), and the final training set was produced by
1402 extracting the training set genes with flanking DNA. This dataset was then split randomly into a
1403 gene set for training (N=3,444 genes) and for testing the trained parameters (N=1,000). Training
1404 was performed using both the standard hidden-Markov model and the conditional random field
1405 (CRF) approach, which can result in more accurate gene prediction but is less robust to errors in
1406 the training dataset. We selected the CRF parameters based on increased sensitivity when applied
1407 to the test dataset. A final RNAseq-based gene model prediction was then performed by passing
1408 the RNAseq BAM files (see above) to BRAKER using the trained CRF parameters ("—
1409 useexisting").
1410
1411 The IsoSeq and AUGUSTUS-based gene models were then semi-manually integrated. IsoSeq-
1412 based models were given precedence, and an AUGUSTUS model was automatically retained if it
1413 did not share any same-strand coding exons with an IsoSeq model. AUGUSTUS models that
1414 exactly matched an IsoSeq model (with the exception of the first and/or last exon, which was
1415 often incorrectly predicted by AUGUSTUS) were immediately filtered. All other AUGUSTUS
1416 models, which could have partial overlap with an IsoSeq model, were manually curated using all
1417 IsoSeq and RNAseq evidence. Alternative transcripts were analyzed by first removing isoforms
1418 that differ only in their UTR sequences, before passing the resulting annotation to SUPPA v2.4
1419 (Trincado et al. 2018) run with the command "-f ioe -e SE RI SS" to classify alternative isoforms
1420 into the categories "skipping exon", "retained intron" and "alternative splice site".
1421

1422 Annotation of the plastome and mitogenome was performed using the GeSeq webserver (Tillich
1423 et al. 2017), which was provided with existing annotations from *A. protothecoides* "0710" and
1424 *Prototheca wickerhamii* (Yan et al. 2015). Additional genes were manually annotated by
1425 searching for long ORFs in the unannotated sequence.
1426
1427 **Repeat annotation**
1428 Preliminary annotation of interspersed repeats was performed by running RepeatModeler 2.0.5
1429 (Flynn et al. 2020) on the UTEX 250-A genome with the parameter "–LTRStruct". Consensus
1430 sequences for specific transposon families were then manually curated following Goubert et al.
1431 (2022) (see Dataset S3, Figure S2). The manually curated models were then combined with the
1432 automated models and passed to RepeatMasker v4.1.6 (https://github.com/Dfam-
1433 consortium/RepeatMasker) ("–gccalc"). Tandem Repeats Finder v4.10.0 was run on the UTEX
1434 250-A genome to identify tandem repeats using the parameters "2 7 7 80 10 50 2000 -f -d -m -
1435 ngs". Total repeat density was calculated by combining the output of RepeatMasker and TRF.
1436 Transposon genes (see Table 1, Figure S2) were identified and marked in the GFF3 annotation
1437 by intersecting the coordinates of manually curated transposons and gene model coding
1438 sequence, followed by manual inspection.
1439
1440 Repeat densities for the genome assemblies in Table 1 were similarly performed by combining
1441 the output of RepeatModeler/RepeatMasker and TRF. RepeatModeler was not run for *A.*
1442 *protothecoides* 0710, which was repeat masked using the UTEX 250-A repeat library (see
1443 above), nor *C. reinhardtii*, for which an extensive repeat library exists (Craig et al. 2021).
1444
1445 **Gene family and phylogenetic analyses**
1446 Gene family analyses were performed using the 19 species shown in Figure 1A. Protein
1447 sequences for each species were reduced to their primary isoforms, and where possible genes on
1448 the organelle genomes were removed. Annotated transposon proteins were included if known.
1449 For UTEX 250-A, all A haplotype genes were arbitrarily selected and supplemented by genes
1450 that are putatively unique to the B haplotype (see below). For *Picochlorum* sp. BPE23, the genes
1451 from only haplotype A were arbitrarily used. Metadata for protein datasets are presented in
1452 Dataset S19.

1453 OrthoFinder v2.5.5 was run on the 19 protein sets using the Diamond search method ("-S
1454 diamond_ultra_sens") (Emms and Kelly 2019). Orthogroups featuring a known transposon
1455 protein were filtered out (the results of which are reflected in Table 1). GreenCut2 and CiliaCut
1456 orthologs were extracted for UTEX 250-A and *C. vulgaris* by comparison to *C. reinhardtii* (i.e.,
1457 from the "orthologues" output directory of OrthoFinder). For the specific genes shown in Figure
1458 5E and F (LHCA, NPQ, meiosis and syngamy), the gene trees for the underlying orthogroups
1459 were manually validated.
1460
1461 The species tree in Figure 1A was produced from a concatenated alignment of the 669 single-
1462 copy orthologs identified by OrthoFinder. For each ortholog, proteins were aligned using
1463 MAFFT v7.525 ("–maxiterate 1000 –localpair"), and the resulting alignments were trimmed
1464 using trimAl v1.4.rev22 ("–automated1") (Capella-Gutiérrez et al. 2009). The trimmed
1465 alignments were concatenated and a maximum likelihood phylogeny was produced using IQ-
1466 TREE v2.3.0 ("-m MFP -bb 1000") (Minh et al. 2020). For Figure 1B, rDNA sequences were
1467 accessed from NCBI (Dataset S1), aligned with MAFFT ("L-INS-i"), manually trimmed, and
1468 passed to IQ-TREE as above.
1469
1470 **Gene IDs and functional annotation**
1471 Allelic gene pairs were determined by running SynChro (Drillon et al. 2014), a tool for detecting
1472 syntenic orthologs between genomes, on the A and B haplotype genomes and gene models.
1473 Genes were assigned a unique 5-digit number starting with "00005" for the first gene on
1474 chromosome A01 (i.e., UTEX250_A01.00005), with each subsequent gene along the A
1475 chromosomes receiving a number that increased in increments of five (i.e.,
1476 UTEX250_A01.00010, UTEX250_A01.00015, etc.). Genes on the B haplotype that had been
1477 identified as part of an allelic pair were assigned the gene number of their corresponding A gene,
1478 regardless of their order on the B chromosomes. Genes that were determined to be unique to
1479 haplotype B were given new gene numbers that extended beyond the last number that was used
1480 on haplotype A. Genes on the plastome were given unique numbers starting at "80000" (e.g.,
1481 UTEX250_xCp.80000), and genes on the mitogenome starting at "90000" (e.g.,
1482 UTEX250_xMt.90000). Isoforms were distinguished by adding a number to the gene ID, with
1483 the primary isoform labelled as "1", e.g., UTEX250_A01.00030.1.

1484 In addition to IDs, genes were also assigned a unique name or "gene symbol" where available
1485 following the nomenclature of *C. reinhardtii* (Craig et al. 2023). Gene symbols were transferred
1486 based on the gene orthology relationships between *C. reinhardtii* and *Auxenochlorella* UTEX
1487 250-A determined by OrthoFinder ("Orthologues" output directory). For genes with 1:1
1488 orthology the *C. reinhardtii* symbol was directly transferred to UTEX 250-A. For multicopy
1489 gene families in *C. reinhardtii* that are represented by a single gene in UTEX 250-A, either the
1490 symbol with the lowest number was used (e.g., *FOX1*/*FOX2* gene pair orthologous to *FOX1* in
1491 UTEX 250-A), or if letters are used, they were removed (e.g., *CMT1A*/*CMT1B* gene pair
1492 orthologous to *CMT1* in UTEX 250-A). For the reverse situation, letters were appended to the
1493 UTEX 250-A gene symbol where possible (e.g., *LPAAT1* orthologous to *LPAAT1A*/*LPAAT1B*
1494 gene pair in UTEX 250-A). Genes on the nuclear genome in *C. reinhardtii* but on an organelle
1495 genome in UTEX 250-A were changed accordingly (e.g., *CHLI1* orthologous to *chlI* in UTEX
1496 250-A plastome). Several gene symbols were manually curated, including a small number of
1497 genes that are absent from *C. reinhardtii* (e.g., the photosystem I subunit gene *psaM*), and all of
1498 the histone genes (for which orthology relationships are difficult to determine). All gene symbol
1499 relationships between *C. reinhardtii* and UTEX 250-A are presented in Dataset S7.
1500 Protein domains were identified by running InterProScan v5.67-99.0 ("-dp -goterms").
1501

1502 **Identification of antisense long noncoding RNA genes**

1503 LncRNA genes were annotated based on IsoSeq data using two approaches. First, IsoSeq-based
1504 gene models that had initially been annotated as protein-coding but had no orthology to proteins
1505 from other species, nor a recognized Pfam or CDD domain, were reassigned as lncRNAs if they
1506 overlapped coding sequence of another gene that did meet one of these criteria. Second, IsoSeq
1507 gene models supported by at least 5 reads for which no ORF had been annotated by SQANTI3
1508 were added. The first category of genes had been assigned gene IDs as described above, and
1509 retained these IDs as lncRNAs instead of protein-coding genes. New gene IDs starting from
1510 "50000" and increasing in increments of 5 were introduced for the second category of lncRNA
1511 genes.
1512

1513 The coordinates of protein-coding genes with either a homolog in at least one other species or a
1514 recognized domain were then intersected with the lncRNA gene coordinates, and those with

1515 antisense overlap spanning at least 30% of the protein-coding gene length were selected for
1516 manual curation. Validated gene models with antisense lncRNAs were then divided into those
1517 with internal birdirectional promoters, or those with internal or immediately adjacent promoters
1518 without evidence for birdirectional activity. Gene ontology enrichment analysis was performed
1519 using topGO (Alexa and Rahnenfuhrer 2023) using the GO term output from InterProScan and
1520 the classic Fisher Test option.
1521
1522 **Genetic toolkit**
1523 Transformation, fluorescence measurements and microscopy were performed following Dueñas
1524 et al. (2025a) and Dueñas et al. (2025b). Full details are provided in Supplementary File S1, the
1525 sequence of all primers used in this study can be found in Dataset S20, and the sequences of
1526 plasmids as a GenBank file in Supplementary File S2.
1527
1528 **DATA AVAILABILITY**
1529 All sequencing data associated with *Auxenochlorella* UTEX 250-A is available from NCBI
1530 under the BioSample SAMN45466464. The UTEX 250-A genome and gene annotations are
1531 specifically available under the BioProjects PRJNA1195245 (Haplotype A) and PRJNA1195244
1532 (Haplotype B). Haploid genome assemblies and sequencing data for UTEX 25, UTEX 2341,
1533 CCAP 211/61 and CCAP 211/7A (sequencing data only) are available from NCBI under the
1534 BioProject PRJNA1328465.
1535
1536 **ACCESSION NUMBERS**
1537 NCBI accession numbers for all *Auxenochlorella* UTEX 250-A genes named in the main text are
1538 available in Dataset S21.
1539
1540
1541
1542
1543
1544
1545

# FIGURES

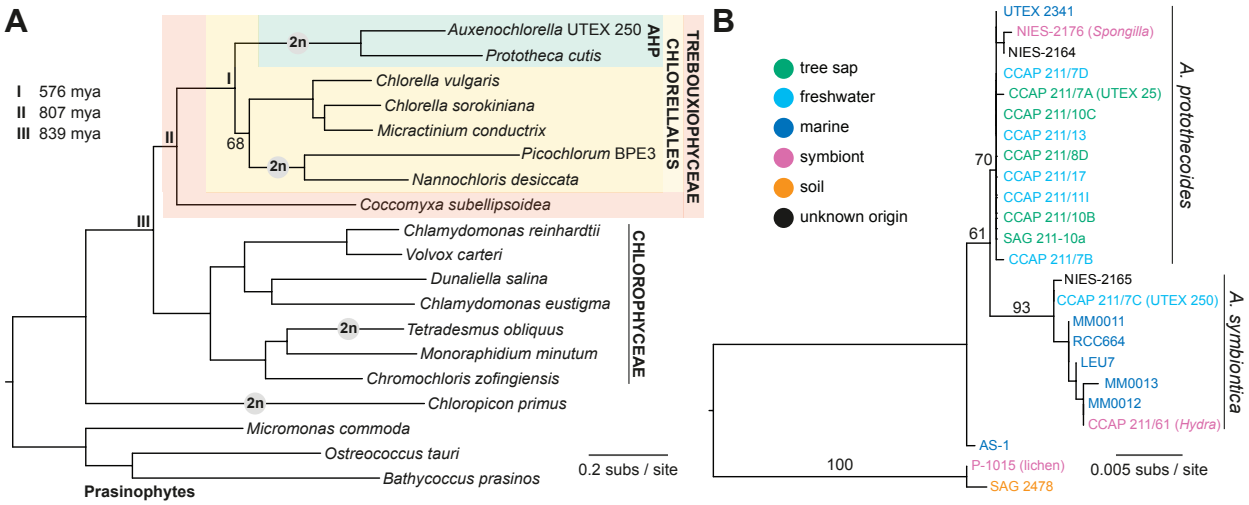
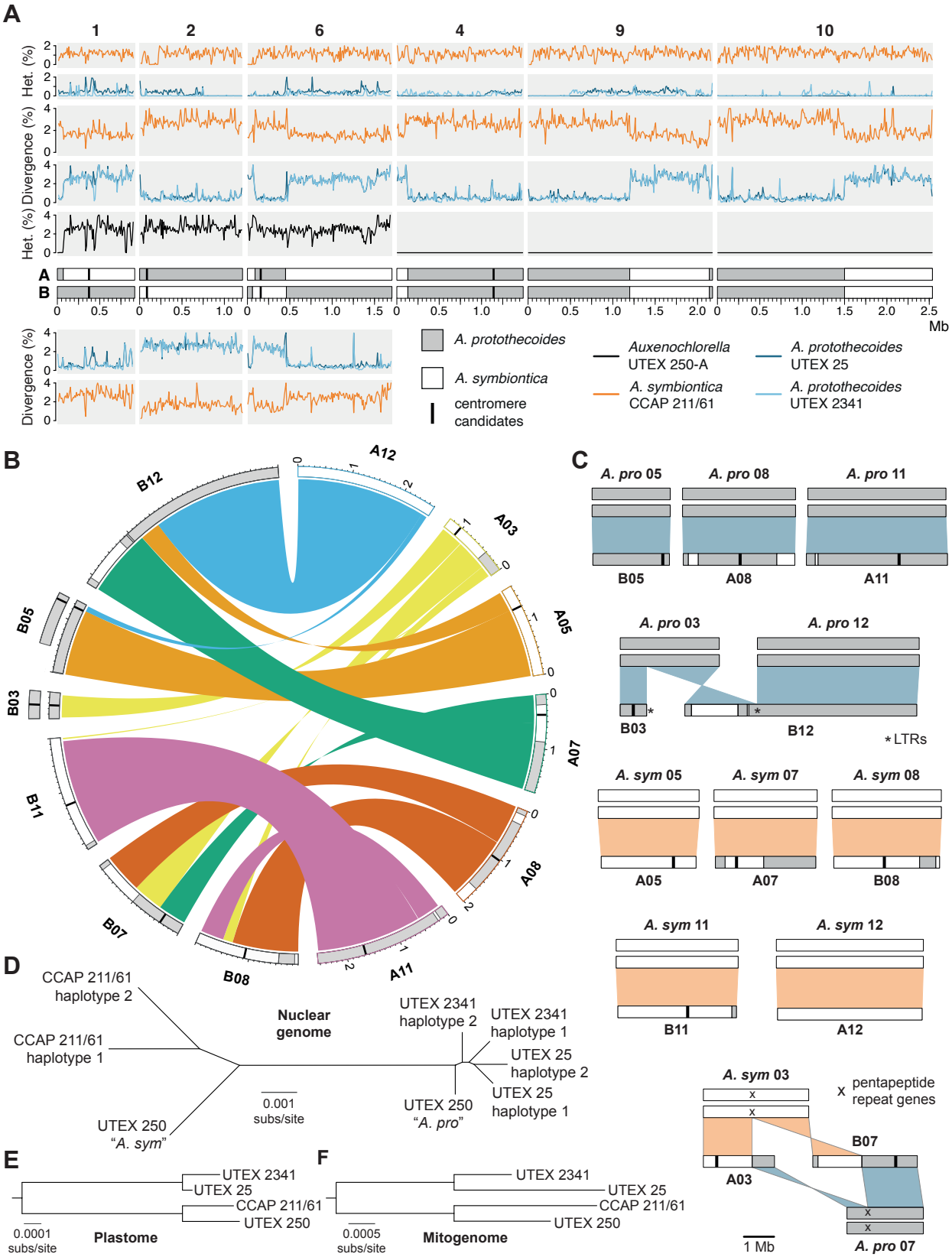



**Figure 1.** Phylogenetic relationships of *Auxenochlorella*. A) Maximum likelihood phylogeny based on concatenated protein alignment of 669 single-copy orthologs, using the LG+F+R5 substitution model. All nodes received ultrafast bootstrap values >95%, except the node connecting *Picochlorum* and *Nannochloris* to *Chlorella* and *Micractinium* (68% ultrafast bootstrap support). Putative transitions to at least partial vegetative diploidy are shown by the "2n" symbol. Divergence estimates were taken from Del Cortona et al. (2020). B) Maximum likelihood phylogeny based on 18S ribosomal DNA alignment, using the TNe+I substitution model. See Dataset S1 for strain metadata.


A
1
2
6
4
9
10
Het. (%)
Divergence (%)
Mb
A. protothecoides
A. symbiontica
centromere candidates
Auxenochlorella UTEX 250-A
A. symbiontica CCAP 211/61
A. protothecoides UTEX 25
A. protothecoides UTEX 2341
B
A12
A03
A05
A07
A08
A11
B08
B07
B11
B03
B05
B12
C
A. pro 05
A. pro 08
A. pro 11
B05
A08
A11
A. pro 03
A. pro 12
B03
B12
* LTRs
A. sym 05
A. sym 07
A. sym 08
A05
A07
B08
A. sym 11
A. sym 12
B11
A12
A. sym 03
x pentapeptide repeat genes
B07
A03
A. pro 07
1 Mb
D
Nuclear genome
CCAP 211/61 haplotype 2
CCAP 211/61 haplotype 1
UTEX 250 "A. sym"
0.001 subs/site
UTEX 2341 haplotype 2
UTEX 2341 haplotype 1
UTEX 25 haplotype 2
UTEX 25 haplotype 1
UTEX 250 "A. pro"
E
UTEX 2341
UTEX 25
CCAP 211/61
UTEX 250
0.0001 subs/site
Plastome
F
UTEX 2341
UTEX 25
CCAP 211/61
UTEX 250
0.0005 subs/site
Mitogenome

1572

1573 **Figure 2.** Allodiploid hybrid origin of UTEX 250. A) Heterozygosity and divergence for four
1574 *Auxenochlorella* strains in 10 kb windows across the six UTEX 250-A chromosomes with one-to-one
1575 homology. For divergence metrics the upper panels refer to haplotype A, lower panels to haplotype B. B)
1576 Genomic rearrangements among the remaining six diploid UTEX 250-A chromosomes. C) Comparison
1577 between rearranged UTEX 250-A chromosomes and diploid *A. protothecoides* and *A. symbiontica*
1578 chromosomes. D) Neighbor joining phylogeny of all *Auxenochlorella* haplotypes (all bootstrap values
1579 100%). E) Neighbor joining phylogeny of plastome sequences (all bootstrap values 100%). F) Neighbor
1580 joining phylogeny of mitogenome sequences (all bootstrap values 100%).
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606

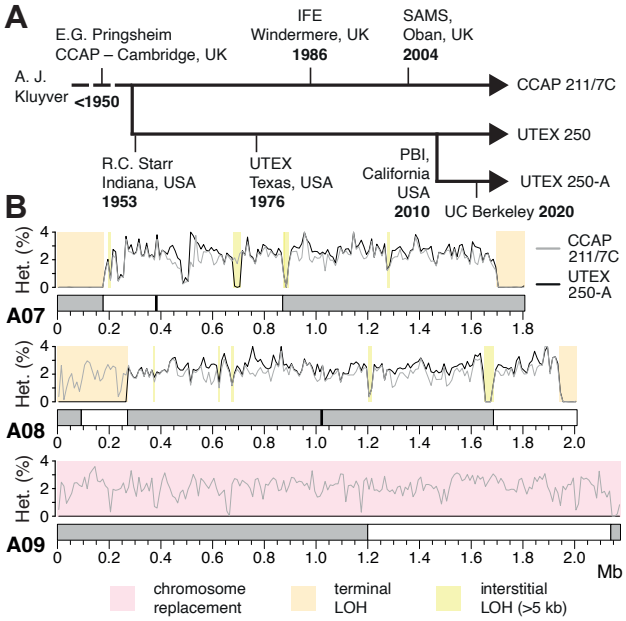



**Figure 3.** Loss-of-heterozygosity and evolution in the laboratory. A) Putative strain history of UTEX 250. IFE = Institute of Freshwater Ecology, SAMS = Scottish Association for Marine Science, PBI = Phycoil Biotechnology International, Inc. B) Heterozygosity in 10 kb windows across three example chromosomes. Chromosomes are shaded according to parental species of origin and candidate centromeres are marked where available. LOH = loss-of-heterozygosity. Different LOH categories in UTEX 250-A are shown by colored boxes.

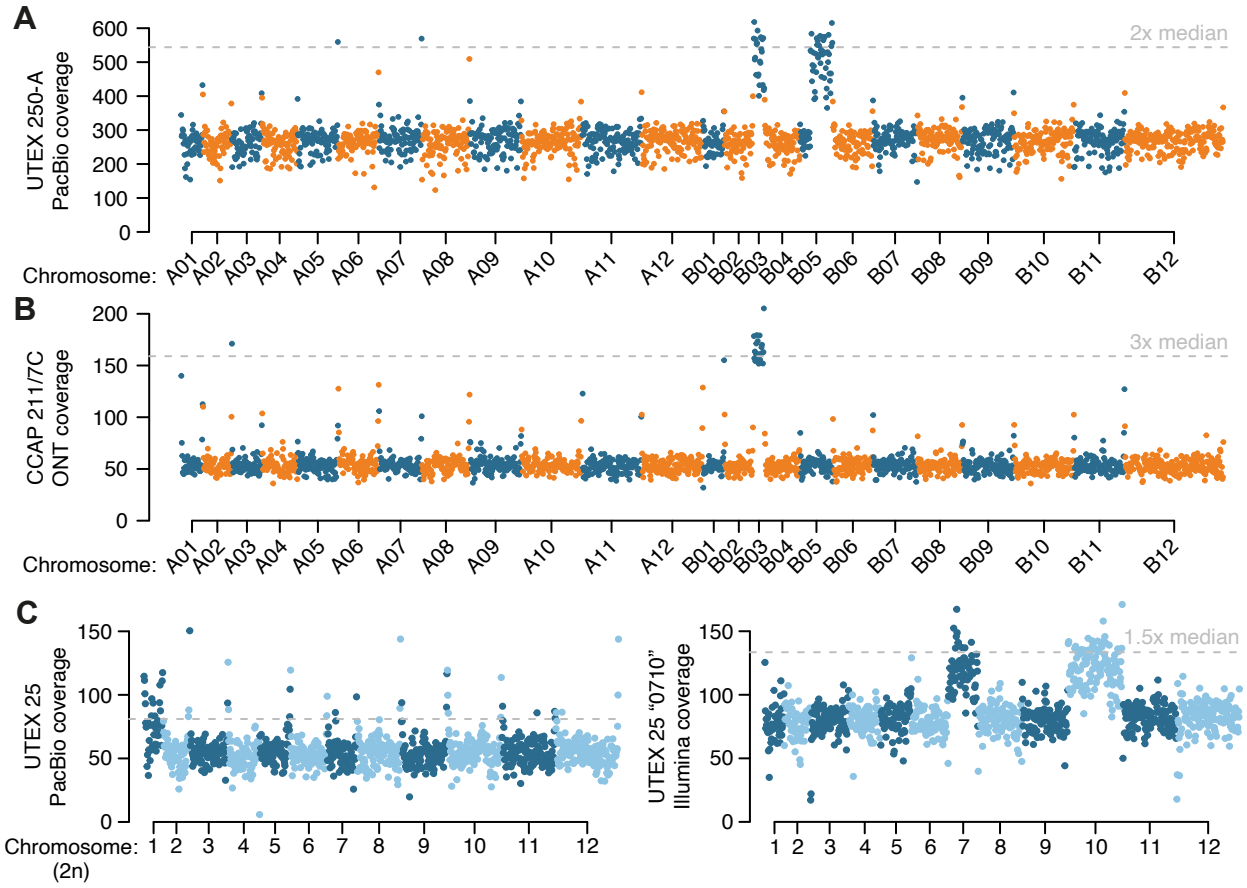



**Figure 4.** Aneuploidy in *Auxenochlorella.* A) UTEX 250-A PacBio read coverage for 20-kb windows across the phased diploid genome assembly of UTEX 250-A. B) CCAP 211/7C ONT read coverage for 20-kb windows across the phased diploid genome assembly of UTEX 250-A. C) UTEX 25 PacBio read coverage and strain "0710" Illumina read coverage for 20-kb windows across a representative haplotype of the UTEX 25 genome assembly.

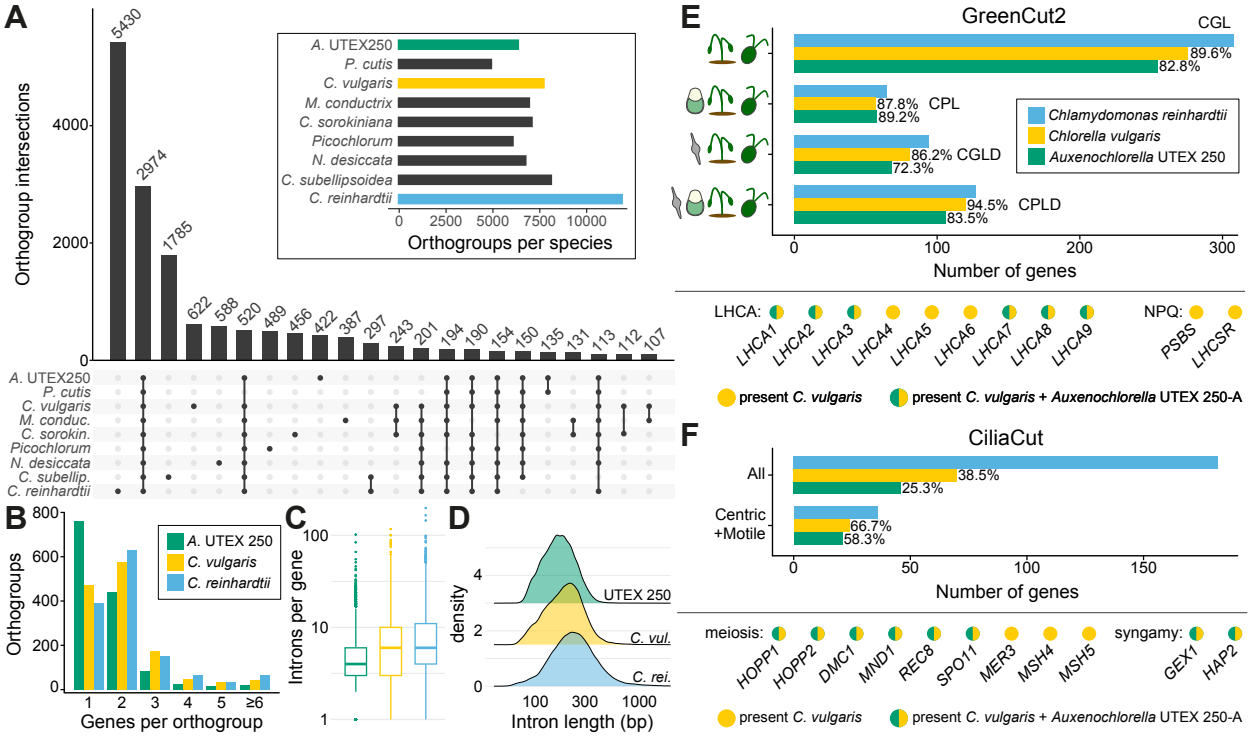



**Figure 5.** Gene family loss and reduction in *Auxenochlorella* UTEX 250-A. A) Upset plot showing intersection between the presence and absence of orthogroups among Trebouxiophyte algae, with *C. reinhardtii* as an outgroup. Only intersects with more than 100 orthogroups are shown. B) Number of genes per orthogroup for orthogroups that are present in each of *Auxenochlorella* UTEX 250-A, *C. vulgaris* and *C. reinhardtii*, and have more than one gene in at least one of the species. C) Introns per gene for multi-exonic genes. D) Intron length distributions. E) Presence of GreenCut2 genes in UTEX 250-A and *C. vulgaris* relative to *C. reinhardtii*. Light-harvesting complex genes of the photosystem I antenna (*LHCA*) and genes functioning in nonphotochemical quenching (NPQ) are highlighted; *PSBS* and *LHCSR* are present as two and three paralogous genes in *C. reinhardtii* and single orthologs in *C. vulgaris* (see Dataset S13). The CGL, CPL, CGLD and CPLD divisions are defined in the main text. F) Presence of CiliaCut, meiosis and syngamy genes in UTEX 250-A and *C. vulgaris* relative to *C. reinhardtii*.

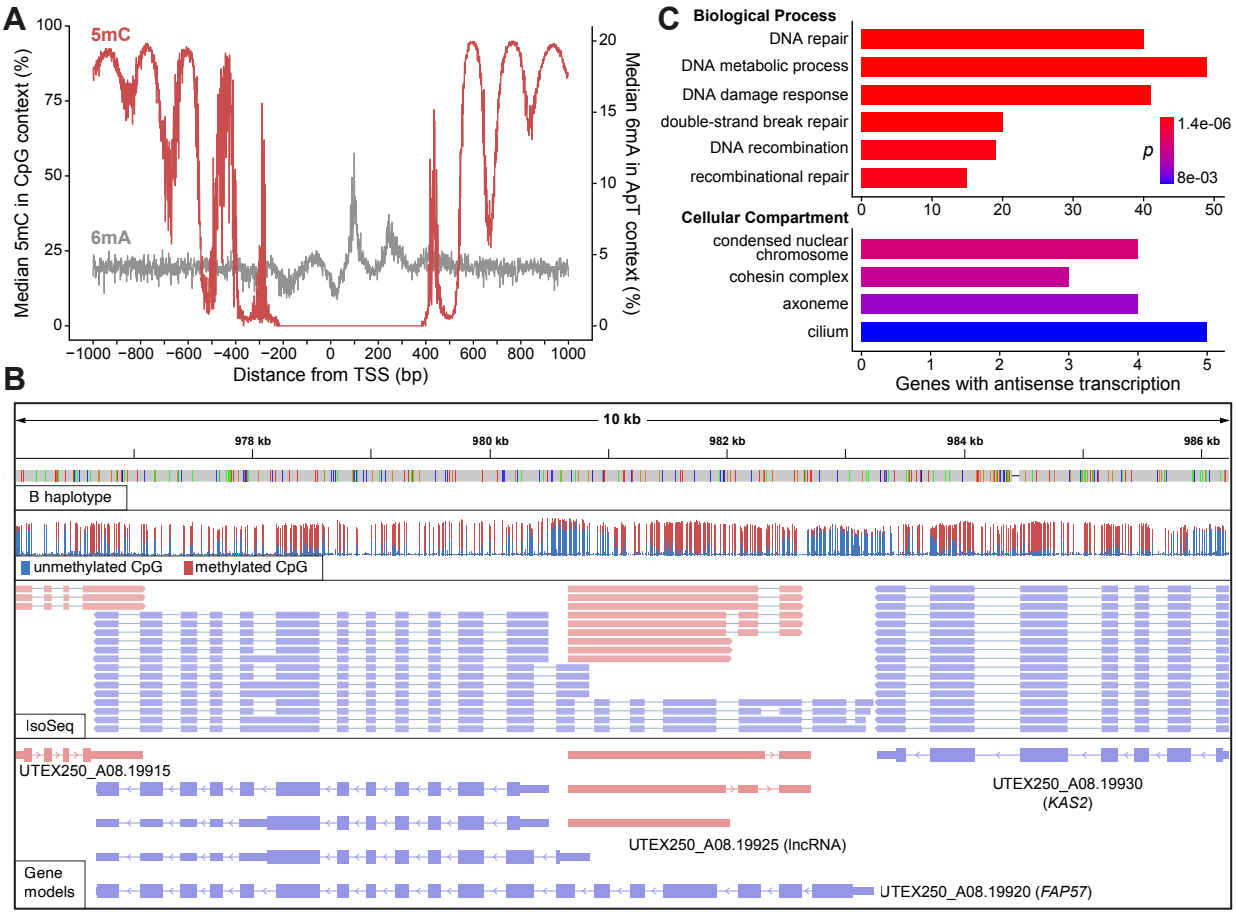
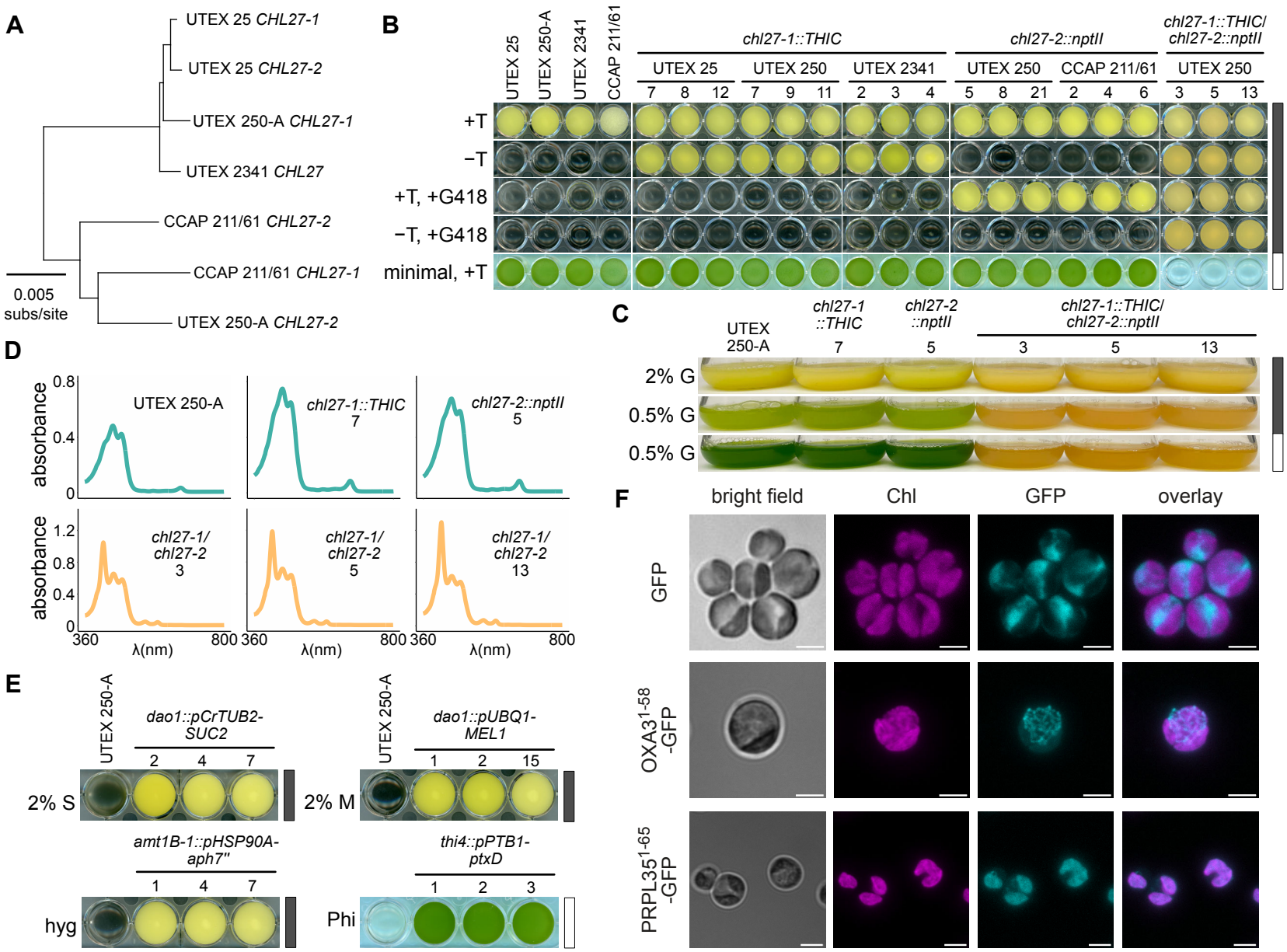



**Figure 6.** Promoter methylation and antisense lncRNAs in *Auxenochlorella* UTEX 250-A. A) Median per site 5mC (red) and 6mA (grey) around transcription start sites. B) IGV screenshot showing example of internal bidirectional promoter producing antisense lncRNA and truncated isoform for *FAP57* (UTEX_A08.19920). The top row shows the B haplotype of the UTEX 250-A genome mapped against the A haplotype, with colored bands corresponding to single nucleotide polymorphisms. The second row shows read coverage of ONT reads from CCAP 211/7C where only CpG sites are shown. The proportion of red to blue corresponds to the proportion of 5mC calls in the ONT reads. The third row shows IsoSeq data, where pink reads are on the forward strand, and blue reads the reverse. The final row shows the annotated gene models, where the thin lines correspond to introns, intermediate lines to UTRs, and thick lines to coding sequence. C) GO term enrichment analysis for 780 genes overlapped by antisense lncRNAs. See Dataset S16 for full results.


A
UTEX 25 CHL27-1
UTEX 25 CHL27-2
UTEX 250-A CHL27-1
UTEX 2341 CHL27
CCAP 211/61 CHL27-2
CCAP 211/61 CHL27-1
UTEX 250-A CHL27-2
0.005 subs/site
B
UTEX 25
UTEX 250-A
UTEX 2341
CCAP 211/61
chl27-1::THIC
chl27-2::nptII
chl27-1::THIC/ chl27-2::nptII
UTEX 25
UTEX 250
UTEX 2341
UTEX 250
CCAP 211/61
UTEX 250
7 8 12 7 9 11 2 3 4 5 8 21 2 4 6 3 5 13
+T
−T
+T, +G418
−T, +G418
minimal, +T
C
UTEX 250-A
chl27-1 ::THIC
chl27-2 ::nptII
chl27-1::THIC/ chl27-2::nptII
7 5 3 5 13
2% G
0.5% G
0.5% G
D
absorbance
UTEX 250-A
chl27-1::THIC 7
chl27-2::nptII 5
chl27-1/ chl27-2 3
chl27-1/ chl27-2 5
chl27-1/ chl27-2 13
λ(nm)
E
UTEX 250-A
dao1::pCrTUB2-SUC2
2 4 7
2% S
dao1::pUBQ1-MEL1
1 2 15
2% M
amt1B-1::pHSP90A-aph7"
1 4 7
hyg
thi4::pPTB1-ptxD
1 2 3
Phi
F
bright field
Chl
GFP
overlay
GFP
OXA3 1-58 -GFP
PRPL35 1-65 -GFP

1692

1693 **Figure 7.** *Auxenochlorella* reverse genetics and genetic toolkit. A) Neighbor-joining phylogeny of *CHL27*
1694 alleles in UTEX 25, 250-A, 2341 and CCAP 211/61. Sequence comparisons encompassed the *CHL27*
1695 alleles and the 5' and 3' flanking regions used to make targeting constructs (see Figure S12A). All bootstrap
1696 values >70%. B) Selection and photosynthetic growth phenotypes of representative thiamine auxotrophic
1697 and G418-resistant transformants versus the corresponding wild-type controls. Cultures were grown for
1698 five days at 26°C with 140 rpm shaking in the dark, or for ten days at 24°C with approximately 40 μmol.m$^{-}$
1699 $^{2}$.s$^{-1}$ of white light provided by cool white fluorescent bulbs. All strains grew in medium supplemented with
1700 2 μM thiamine (+T), but only transformants expressing *THIC* grew in medium without added thiamine (-
1701 T). Resistance to 100 μg/mL of G418 (+T, +G418) was conferred by *nptII*. *chl27* double knockout strains
1702 grew under both selection conditions (-T, +G418). Wild-type and *chl27* heterozygotes grew
1703 photoautotrophically, but *chl27* homozygous mutants were non-photosynthetic (minimal, +T). C) Pigment
1704 accumulation under different trophic conditions in 25 mL flask cultures of UTEX 250-derived *chl27*
1705 heterozygous and homozygous mutants versus UTEX 250-A. Cultures were grown with 2 μM thiamine and
1706 20 g/L (2% G) or 5 g/L (5% G) of glucose for three days at 28°C, with 200 rpm shaking, in the dark or with
1707 approximately 5 μmol.m$^{-2}$.s$^{-1}$ of white light provided by warm white LEDs. D) Absorption spectra of
1708 extracts from UTEX 250-A, *chl27* heterozygous and homozygous mutant strains. Cultures were grown for
1709 four days in the dark with 20 g/L glucose at 24°C, 160 rpm in 24-well plates. Shoulders at 421 nm and
1710 peaks around 450 nm and 475 nm are consistent with absorption by lutein and zeaxanthin, and a minor peak
1711 at 663 nm reflects a small amount of chlorophyll in heterotrophic wild-type and *chl27* heterozygotes. Peaks
1712 at 418, 551 and 590 nm in extracts from chl27 double knockouts are indicative of absorption by Mg-
1713 protoporphyrin IX or Mg-PMME. E) Selectable marker growth phenotypes of transformants compared to
1714 UTEX 250-A. Strains were cultured for four days in the dark at 26°C with 140 rpm shaking, in 20 g/L
1715 sucrose (2% S), 20 g/L melibiose (2% M), 300 μg/mL hygromycin B (hyg), or for 5 days with
1716 approximately 40 μmol.m$^{-2}$.s$^{-1}$ of white light provided by cool white fluorescent bulbs, at 24°C with 140
1717 rpm shaking, in tris-acetate medium containing 1.1 mM phosphite (Phi). *SUC2* was targeted to the *DAO1*
1718 locus, and driven by the *TUB2* promoter from *C. reinhardtii* (*pCrTUB2*); *MEL1* was integrated at *DAO1*
1719 and driven by the endogenous *UBQ1* promoter (*pUBQ1*); *aph7''* was controlled by the promoter of *HSP90A*
1720 (*pHSP90A*) and targeted to *AMT1B-1*; and *ptxD*, under the control of the *PTB1* promoter (*pPTB1*), was
1721 integrated at *THI4*. F) Fluorescence imaging of UTEX 250-A transformants expressing sucrose invertase
1722 and GFP. Cells were from cultures grown with 5 g/L of sucrose for three to four days at 25°C, shaking at
1723 160 rpm, and approximately 40 μmol.m$^{-2}$.s$^{-1}$ of white light provided by cool white fluorescent bulbs.
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743

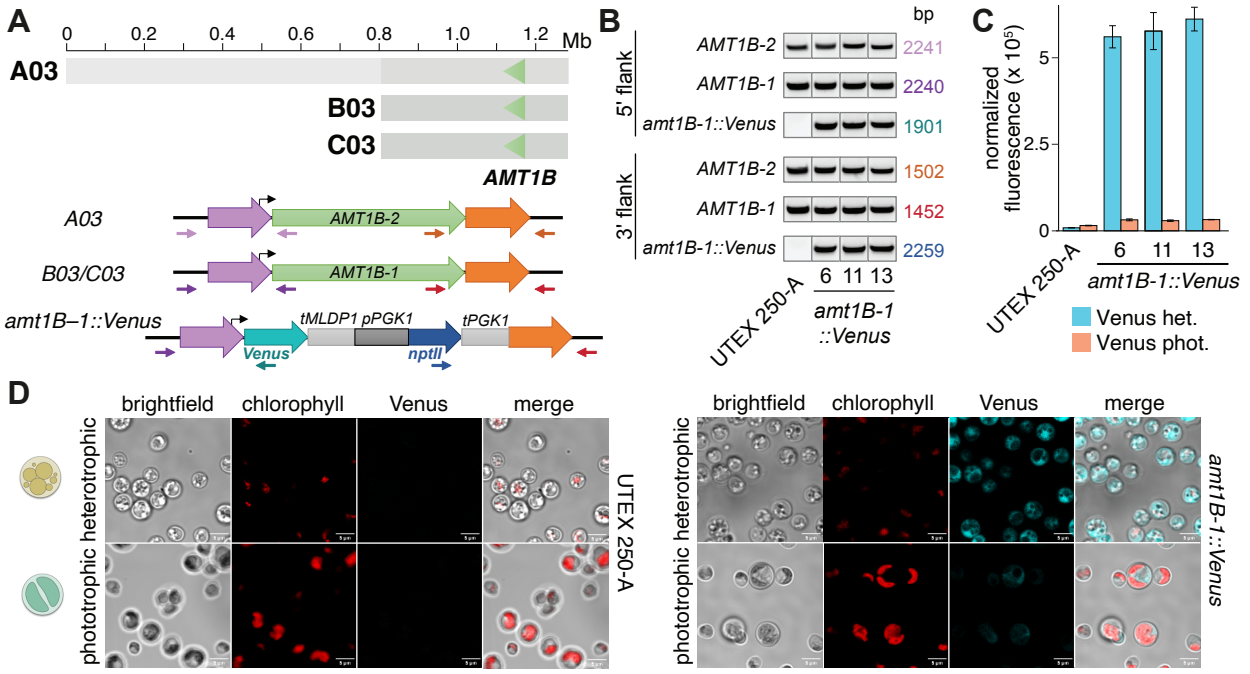



**Figure 8.** Allele-specific targeting of a trisomic chromosome. A) Schematic of chromosome 3 copies (A03 and B03/C03) with green triangles marking the location of *AMT1B*. The mutant locus features a gene cassette comprised of a *Venus* ORF and a 537 bp terminator sequence containing the endogenous 3' UTR of *MLDP1* (*MAJOR LIPID DROPLET PROTEIN 1*), and a gene cassette comprised of a *nptII* ORF flanked by the endogenous promoter/5' UTR and terminator/3' UTR of *PGK1B* (*PHOSPHOGLYCERATE KINASE 1B*), which is targeted to the *AMT1B-1* allele of B03/C03. *Venus* expression is under the control of the endogenous *AMT1B-1* promoter. B) PCR amplification of *AMT1B-1* (B03/C03) and *AMT1B-2* (A03) 5' and 3' flanking regions using allele-specific primers in wild-type UTEX 250-A and three representative *amt1B-1::Venus* transformants, with additional amplification of *amt1B-1::Venus* integration in the transformants. C) Venus fluorescence detection in late-log heterotrophic transformant cultures compared to wildtype UTEX 250-A negative control. D) Comparison of chlorophyll autofluorescence (red) and Venus fluorescence (cyan) in wild-type UTEX 250-A versus a representative transformant under heterotrophic or photoautotrophic conditions using confocal fluorescence microscopy.

## SUPPLEMENTARY DATA

**Figure S1.** Permuted tRNA genes in *Auxenochlorella* UTEX 250-A.

**Figure S2.** IGV screenshots of transposable element insertions that are polymorphic between the A and B haplotypes.

**Figure S3.** Putative AT-rich short regional centromeres in *Auxenochlorella* UTEX 250-A Haplotype A.

**Figure S4.** Map of *Auxenochlorella* UTEX250-A plastome.

**Figure S5.** Map of *Auxenochlorella* UTEX250-A mitogenome.

**Figure S6.** UTEX 250-A and CCAP 211/7C chromosome schematics and heterozygosity.

**Figure S7.** IGV screenshot of putative de novo telomere on duplicate fragment of chromosome B05.

**Figure S8.** PacBio read coverage in 20-kb windows across a representative haplotype of the *Auxenochlorella symbiontica* CCAP 211/61 and *Auxenochlorella protothecoides* UTEX 2341 genome assemblies.

**Figure S9**. Putative RBCX orthologs in *Auxenochlorella* UTEX 250-A and *Chlorella vulgaris*.

**Figure S10.** 5-methylcytosine and $N^6$-methyladenine around transcription start sites of UTEX 250-A genes.

**Figure S11.** IGV screenshot of *GEX1* and antisense lncRNA.

**Figure S12.** *Auxenochlorella* reverse genetics.

**Figure S13.** *Auxenochlorella* selection markers and intracellular localization.

**Dataset S1**. *Auxenochlorella* strain metadata for Figure 1B.

**Dataset S2**. Summary statistics for existing *Auxenochlorella* genome assemblies.

**Dataset S3**. Annotation notes for manually curated transposable elements and other interspersed repeats in the *Auxenochlorella* UTEX 250-A genome.

**Dataset S4**. Possible centromere coordinates for the *Auxenochlorella* UTEX 250-A genome.

**Dataset S5**. Summary statistics for *Auxenochlorella protothecoides* and *Auxenochlorella symbiontica* haploid genome assemblies.

**Dataset S6**. Percentage of UTEX 250-A proteins with functional domains as determined by InterProScan.

**Dataset S7**. *Auxenochlorella* UTEX 250-A genes and gene symbols and their orthology to *C. reinhardtii* genes and gene symbols.

**Dataset S8**. *Auxenochlorella* UTEX 250-A heterozygosity by site class.

1820 **Dataset S9**. *Auxenochlorella* UTEX 250-A orthogroup analysis.
1821
1822 **Dataset S10**. BUSCO genes that have been lost from the *Auxenochlorella* UTEX 250-A genome.
1823
1824 **Dataset S11**. Expanded orthogroups in *Auxenochlorella* UTEX 250-A genome featuring tandem
1825 duplication of genes.
1826
1827 **Dataset S12**. GreenCut2 genes that are absent in the *Auxenochlorella* UTEX 250-A genome.
1828
1829 **Dataset S13**. Presence and absence of *LHCA* genes and specific genes involved in nonphotochemical
1830 quenching across Trebouxiophyceae.
1831
1832 **Dataset S14**. CiliaCut genes that are absent in the *Auxenochlorella* UTEX 250-A genome.
1833
1834 **Dataset S15**. UTEX 250-A protein-coding genes with IsoSeq evidence for antisense lncRNA
1835 transcription.
1836
1837 **Dataset S16**. Significant GO terms for genes with antisense lncRNA transcription.
1838
1839 **Dataset S17**. Core genes involved in non-homologous end joining.
1840
1841 **Dataset S18**. UTEX 250-A growth conditions for RNA extraction.
1842
1843 **Dataset S19**. Species used in Orthofinder analyses.
1844
1845 **Dataset S20**. Primers for genotyping.
1846
1847 **Dataset S21**. NCBI accessions for *Auxenochlorella* UTEX 250-A genes named in the main text.
1848
1849 **Supplementary File S1**. Supplementary methods for the *Auxenochlorella* genetic toolkit.
1850
1851 **Supplementary File S2**. Genbank file of plasmid sequences used in this study.
1852

1853 **FUNDING**

1854 This work was supported by the Department of Energy (DOE) Office of Science, Biological and
1855 Environmental Research program under award no. DE-SC0023027 (to SSM and JLM) with
1856 support from the Gordon and Betty Moore Foundation (to SSM) for the work
1857 on *A. symbiontica* (9203). RJC was supported for his work on *Auxenochlorella* genome
1858 sequencing and assembly in Berkeley, in part, through the Laboratory Directed Research and
1859 Development Program of Lawrence Berkeley National Laboratory (to SSM) under U.S.
1860 Department of Energy Contract No. DE-AC02-05CH11231. MAD was supported, in part, by the
1861 NIH T32 Genetics Dissection of Cells and Organisms Grant (1T32GM132022-01) and the
1862 Newton Graduate Fellowship in Synthetic Biology (QB3-Berkeley). DJC was supported in part
1863 by the NIH T32 Molecular Basis of Cell Function Training Grant (5T32GM007232-44) and the

University of California, Berkeley, Chancellor's Fellowship. Work at the Molecular Foundry was supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 (CB-H). Work at the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231 (CLB-H).

## ACKNOWLEDGMENTS

Confocal fluorescence microscopy was performed at the CRL Molecular Imaging Center, RRID:SCR_017852.

## CONFLICT OF INTEREST STATEMENT

JLM served as a paid consultant for Phycoil Biotechnology International, Inc. (PBI), a company that uses *Auxenochlorella* and other microorganisms to develop therapeutic and nutritional oils. PBI had no role in the design, execution, data collection, analysis, or interpretation of this study, nor was the company involved in the preparation, review, or approval of this manuscript. All authors declare that they have no other conflicts of interest to disclose related to this publication.

## AUTHOR CONTRIBUTIONS

RJC, JLM and SSM designed the research. RJC, MAD, DLC, SDG, MCA-M, Y-TL and JLM performed the research. RJC, MAD, SDG, MCA-M, CLB-H and JLM analyzed the data. RJC, MAD, JLM and SSM wrote the paper.

## REFERENCES

Akella S, Ma X, Bacova R, Harmer ZP, Kolackova M, Wen X, Wright DA, Spalding MH, Weeks DP, Cerutti H. 2021. Co-targeting strategy for precise, scarless gene editing with CRISPR/Cas9 and donor ssODNs in *Chlamydomonas*. *Plant Physiol* **187**: 2637–2655.

Alexa A, Rahnenfuhrer J. 2023. topGO: enrichment analysis for gene ontology. doi:10.18129/B9.bioc.topGO.R package version 2.54.0

Allen MD, del Campo JA, Kropat J, Merchant SS. 2007. *FEA1*, *FEA2*, and *FRE1*, encoding two homologous secreted proteins and a candidate ferrireductase, are expressed coordinately with *FOX1* and *FTR1* in iron-deficient *Chlamydomonas reinhardtii*. *Eukaryot Cell* **6**: 1841–1852.

1897 Allorent G, Lefebvre-Legendre L, Chappuis R, Kuntz M, Truong TB, Niyogi KK, Ulm R,
1898 Goldschmidt-Clermont M. 2016. UV-B photoreceptor-mediated protection of the
1899 photosynthetic machinery in *Chlamydomonas reinhardtii*. *Proc Natl Acad Sci U S A* **113**:
1900 14864–14869.
1901 Almagro Armenteros JJ, Salvatore M, Emanuelsson O, Winther O, von Heijne G, Elofsson A,
1902 Nielsen H. 2019. Detecting sequence signals in targeting peptides using deep learning.
1903 *Life Sci Alliance* **2**: e201900429.
1904 Altmannova V, Firlej M, Muller F, Janning P, Rauleder R, Rousova D, Schaffler A, Bange T,
1905 Weir JR. 2023. Biochemical characterisation of Mer3 helicase interactions and the
1906 protection of meiotic recombination intermediates. *Nucleic Acids Res* **51**: 4363–4384.
1907 Andersen MP, Nelson ZW, Hetrick ED, Gottschling DE. 2008. A genetic screen for increased
1908 loss of heterozygosity in *Saccharomyces cerevisiae*. *Genetics* **179**: 1179–1195.
1909 Arriola MB, Velmurugan N, Zhang Y, Plunkett MH, Hondzo H, Barney BM. 2018. Genome
1910 sequences of *Chlorella sorokiniana* UTEX 1602 and *Micractinium conductrix* SAG
1911 241.80: implications to maltose excretion by a green alga. *Plant J* **93**: 566–586.
1912 Asker D, Awad TS. 2019. Isolation and characterization of a novel lutein-producing marine
1913 microalga using high throughput screening. *Food Res Int* **116**: 660–667.
1914 Bakuła Z, Siedlecki P, Gromadka R, Gawor J, Gromadka A, Pomorski JJ, Panagiotopoulou H,
1915 Jagielski T. 2021. A first insight into the genome of *Prototheca wickerhamii*, a major
1916 causative agent of human protothecosis. *BMC Genomics* **22**: 168.
1917 Bao W, Jurka J. 2013. Homologues of bacterial TnpB_*IS605* are widespread in diverse
1918 eukaryotic transposable elements. *Mob DNA* **4**: 12.
1919 Barten R, van Workum DM, de Bakker E, Risse J, Kleisman M, Navalho S, Smit S, Wijffels RH,
1920 Nijveen H, Barbosa MJ. 2022. Genetic mechanisms underlying increased microalgal
1921 thermotolerance, maximal growth rate, and yield on light following adaptive laboratory
1922 evolution. *BMC Biol* **20**: 242.
1923 Becker SA, Spreafico R, Kit JL, Brown R, Likhogrud M, Fang W, Posewitz MC, Weissman JC,
1924 Radakovits R. 2020. Phased diploid genome sequence for the fast-growing microalga
1925 *Picochlorum celeri*. *Microbiol Resour Announc* **9**: e00087–20.
1926 Benites LF, Bucchini F, Sanchez-Brosseau S, Grimsley N, Vandepoele K, Piganeau G. 2021.
1927 Evolutionary genomics of sex-related chromosomes at the base of the green lineage.
1928 *Genome Biol Evol* **13**: evab216.
1929 Berthold P, Schmitt R, Mages W. 2002. An engineered *Streptomyces hygroscopicus aph 7″* gene
1930 mediates dominant resistance against hygromycin B in *Chlamydomonas reinhardtii*.
1931 *Protist* **153**: 401–412.
1932 Bhat R, Im BH, Kim J, Im CS. 2025. Microalgae compositions and methods for treating disease.
1933 World Intellectual Property Organization application number WO2025101f37A1, filed
1934 November 1, 2024, and published May 15, 2025
1935 Biondi TC, Kruse CPS, Koehler SI, Kwon T, Davis AK, Eng W, Kunde Y, Gleasner CD, Mak
1936 KTY, Polle J et al. 2024. The telomere-to-telomere, gapless, phased diploid genome and
1937 methylome of the green alga *Scenedesmus obliquus* UTEX 3031 reveals significant
1938 heterozygosity and genetic divergence of the haplotypes. *Algal Research* **79**: 103431.
1939 Birchler JA, Kelly J, Singh J, Liu H, Zhang Z, Char SN, Sharma M, Yang H, Albert PS, Yang B.
1940 2024. Synthetic minichromosomes in plants: past, present, and promise. *Plant J* **120**:
1941 2356-2366

1942 Blaby IK, Blaby-Haas CE, Tourasse N, Hom EF, Lopez D, Aksoy M, Grossman A, Umen J,
1943 Dutcher S, Porter M et al. 2014. The Chlamydomonas genome project: a decade on.
1944 *Trends Plant Sci* **19**: 672–680.
1945 Blaby-Haas CE, Merchant SS. 2019. Comparative and functional algal genomics. *Annu Rev*
1946 *Plant Biol* **70**: 605–638.
1947 Blanc G, Agarkova I, Grimwood J, Kuo A, Brueggeman A, Dunigan DD, Gurnon J, Ladunga I,
1948 Lindquist E, Lucas S et al. 2012. The genome of the polar eukaryotic microalga
1949 *Coccomyxa subellipsoidea* reveals traits of cold adaptation. *Genome Biol* **13**: R39
1950 Blanc G, Duncan G, Agarkova I, Borodovsky M, Gurnon J, Kuo A, Lindquist E, Lucas S,
1951 Pangilinan J, Polle J et al. 2010. The *Chlorella variabilis* NC64A genome reveals
1952 adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. *Plant Cell* **22**:
1953 2943–2955.
1954 Bochtler M, Fernandes H. 2021. DNA adenine methylation in eukaryotes: Enzymatic mark or a
1955 form of DNA damage? *Bioessays* **43**: e2000243.
1956 Bowman JL Kohchi T Yamato KT Jenkins J Shu S Ishizaki K Yamaoka S Nishihama R
1957 Nakamura Y Berger F et al. 2017. Insights into land plant evolution garnered from the
1958 *Marchantia polymorpha* genome. *Cell* **171**: 287–304 e15.
1959 Boynton JE, Gillham NW, Harris EH, Hosler JP, Johnson AM, Jones AR, Randolph-Anderson
1960 BL, Robertson D, Klein TM, Shark KB et al. 1988. Chloroplast transformation in
1961 *Chlamydomonas* with high velocity microprojectiles. *Science* **240**: 1534–1538.
1962 Bracher A, Hauser T, Liu C, Hartl FU, Hayer-Hartl M. 2015. Structural analysis of the Rubisco-
1963 assembly chaperone RbcX-II from *Chlamydomonas reinhardtii*. *PLoS One* **10**: e0135448.
1964 Brooks G, Franklin S, Avila J, Decker SM, Baliu E, Rakitsky W, Piechocki J, Zdanis D, Norris
1965 LM. 2024. Microalgal flour. United States US12059006B2, filed November 5 2021, and
1966 issued July 26, 2024.
1967 Bulankova P, Sekulić M, Jallet D, Nef C, van Oosterhout C, Delmont TO, Vercauteren I, Osuna-
1968 Cruz CM, Vancaester E, Mock T et al. 2021. Mitotic recombination between homologous
1969 chromosomes drives genomic diversity in diatoms. *Curr Biol* **31**: 3221–3232.e9.
1970 Calhoun S, Bell TAS, Dahlin LR, Kunde Y, LaButti K, Louie KB, Kuftin A, Treen D, Dilworth
1971 D, Mihaltcheva S et al. 2021. A multi-omic characterization of temperature stress in a
1972 halotolerant *Scenedesmus* strain for algal biotechnology. *Commun Biol* **4**: 333.
1973 Camacho DJ, Craig RJ, Merchant SS. 2025. Extraction and sequencing of high molecular weight
1974 genomic DNA from *Auxenochlorella protothecoides* using the Oxford Nanopore
1975 Technologies MinION platform. doi.org/10.17504/protocols.io.14egn6r8ql5d/v1.
1976 Protocols.io
1977 Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated
1978 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.
1979 Caspy I, Malavath T, Klaiman D, Fadeeva M, Shkolnisky Y, Nelson N. 2020. Structure and
1980 energy transfer pathways of the *Dunaliella Salina* photosystem I supercomplex. *Biochim*
1981 *Biophys Acta Bioenerg* **1861**: 148253.
1982 Cecchin M, Marcolungo L, Rossato M, Girolomoni L, Cosentino E, Cuine S, Li-Beisson Y,
1983 Delledonne M, Ballottari M. 2019. *Chlorella vulgaris* genome assembly and annotation
1984 reveals the molecular basis for metabolic acclimation to high light conditions. *Plant J*
1985 **100**: 1289–1305.
1986 Chan PP, Lin BY, Mak AJ, Lowe TM. 2021. tRNAscan-SE 2.0: improved detection and
1987 functional classification of transfer RNA genes. *Nucleic Acids Res* **49**: 9077–9096.

1988 Chang HHY, Pannunzio NR, Adachi N, Lieber MR. 2017. Non-homologous DNA end joining
1989 and alternative pathways to double-strand break repair. *Nat Rev Mol Cell Biol* **18**: 495–
1990 506.
1991 Charron G, Marsit S, Henault M, Martin H, Landry CR. 2019. Spontaneous whole-genome
1992 duplication restores fertility in interspecific hybrids. *Nat Commun* **10**: 4126.
1993 Chaux F, Agier N, Eberhard S, Xu Z. 2024. Extraction and selection of high-molecular-weight
1994 DNA for long-read sequencing from *Chlamydomonas reinhardtii*. *PLoS One* **19**:
1995 e0297014.
1996 Chaux-Jukic F, O'Donnell S, Craig RJ, Eberhard S, Vallon O, Xu Z. 2021. Architecture and
1997 evolution of subtelomeres in the unicellular green alga *Chlamydomonas reinhardtii*.
1998 *Nucleic Acids Res* **49**: 7571–7587.
1999 Chen GYE, Adams NBP, Jackson PJ, Dickman MJ, Hunter CN. 2021. How the O2-dependent
2000 Mg-protoporphyrin monomethyl ester cyclase forms the fifth ring of chlorophylls. *Nat*
2001 *Plants* **7**: 365–375.
2002 Chen H, Sosa A, Chen F. 2024. Growth and cell size of microalga *Auxenochlorella*
2003 *protothecoides* AS-1 under different trophic modes. *Microorganisms* **12**: 835.
2004 Cheng H, Jarvis ED, Fedrigo O, Koepfli KP, Urban L, Gemmell NJ, Li H. 2022. Haplotype-
2005 resolved assembly of diploid genomes without parental data. *Nat Biotechnol* **40**: 1332–
2006 1335.
2007 Christmann M, Kaina B. 2013. Transcriptional regulation of human DNA repair genes following
2008 genotoxic stress: trigger mechanisms, inducible responses and genotoxic adaptation.
2009 *Nucleic Acids Res* **41**: 8403–8420.
2010 Costas AM, White AK, Metcalf WW. 2001. Purification and characterization of a novel
2011 phosphorus-oxidizing enzyme from *Pseudomonas stutzeri* WM88. *J Biol Chem* **276**:
2012 17429–17436.
2013 Craig RJ, Gallaher SD, Shu S, Salome PA, Jenkins JW, Blaby-Haas CE, Purvine SO, O'Donnell
2014 S, Barry K, Grimwood J et al. 2023. The Chlamydomonas Genome Project, version 6:
2015 Reference assemblies for mating-type plus and minus strains reveal extensive structural
2016 mutation in the laboratory. *Plant Cell* **35**: 644–672.
2017 Craig RJ, Hasan AR, Ness RW, Keightley PD. 2021. Comparative genomics of *Chlamydomonas*.
2018 *Plant Cell* **33**: 1016–1041.
2019 Cross FR. 2015. Tying down loose ends in the *Chlamydomonas* genome: functional significance
2020 of abundant upstream open reading frames. *G3 (Bethesda)* **6**: 435–446.
2021 Crozet P, Navarro FJ, Willmund F, Mehrshahi P, Bakowski K, Lauersen KJ, Perez-Perez ME,
2022 Auroy P, Gorchs Rovira A, Sauret-Gueto S et al. 2018. Birth of a photosynthetic chassis:
2023 A MoClo Toolkit enabling synthetic biology in the microalga *Chlamydomonas*
2024 *reinhardtii*. *ACS Synth Biol* **7**: 2074–2086.
2025 da Roza PA, Muller H, Sullivan GJ, Walker RSK, Goold HD, Willows RD, Palenik B, Paulsen
2026 IT. 2024. Chromosome-scale assembly of the streamlined picoeukaryote *Picochlorum* sp.
2027 SENEW3 genome reveals Rabl-like chromatin structure and potential for C(4)
2028 photosynthesis. *Microb Genom* **10**: 001223.
2029 Dale AL, Feau N, Everhart SE, Dhillon B, Wong B, Sheppard J, Bilodeau GJ, Brar A, Tabima
2030 JF, Shen D et al. 2019. Mitotic recombination and rapid genome evolution in the invasive
2031 forest pathogen *Phytophthora ramorum*. *mBio* **10**: e02452–02418.

2032 Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T,
2033 McCarthy SA, Davies RM et al. 2021. Twelve years of SAMtools and BCFtools.
2034 *Gigascience* **10**: giab008.
2035 Darienko T, Pröschold T. 2015. Genetic variability and taxonomic revision of the genus
2036 *Auxenochlorella* (Shihira et Krauss) Kalina et Puncocharova (Trebouxiophyceae,
2037 Chlorophyta). *J Phycol* **51**: 394–400.
2038 Del Cortona A, Jackson CJ, Bucchini F, Van Bel M, D'Hondt S, Skaloud P, Delwiche CF, Knoll
2039 AH, Raven JA, Verbruggen H et al. 2020. Neoproterozoic origin and multiple transitions
2040 to macroscopic growth in green seaweeds. *Proc Natl Acad Sci U S A* **117**: 2551–2559.
2041 Diner RE, Nodding CM, Lian NC, Kang AK, McQuaid JB, Jablanovic J, Espinoza JL, Nguyen
2042 NA, Anzelmatti MA, Jansson J et al. 2017. Diatom centromeres suggest a mechanism for
2043 nuclear DNA acquisition. *Proc Natl Acad Sci U S A* **114**: E6015–E6024.
2044 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras
2045 TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.
2046 Drillon G, Carbone A, Fischer G. 2014. SynChro: a fast and easy tool to reconstruct and
2047 visualize synteny blocks along eukaryotic chromosomes. *PLoS One* **9**: e92621.
2048 Dueñas MA, Craig RJ, Gallaher SD, Moseley JL, Merchant SS. 2025a. Leaky ribosomal
2049 scanning enables tunable translation of bicistronic ORFs in green algae. *Proc Natl Acad*
2050 *Sci U S A* **122**: e2417695122.
2051 Dueñas MA, Lin Y-T, Moseley JL, Merchant SS. 2025b. Transformation protocol for targeted
2052 homologous recombination in *Auxenochlorella protothecoides*.
2053 doi.org/10.17504/protocols.io.x54v922mql3e/v1. Protocols.io
2054 Dutta A, Dutreux F, Schacherer J. 2021. Loss of heterozygosity results in rapid but variable
2055 genome homogenization across yeast genetic backgrounds. *Elife* **10**: e70339.
2056 Dutta A, Schacherer J. 2025. The dynamics of loss of heterozygosity events in genomes. *EMBO*
2057 *Rep* **26**: 602–612.
2058 Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative
2059 genomics. *Genome Biol* **20**: 238.
2060 Fedry J, Liu Y, Pehau-Arnaudet G, Pei J, Li W, Tortorici MA, Traincard F, Meola A, Bricogne
2061 G, Grishin NV et al. 2017. The ancient gamete fusogen HAP2 is a eukaryotic class II
2062 fusion protein. *Cell* **168**: 904–915 e10.
2063 Feng L, Mundy JEA, Stevenson CEM, Mitchenall LA, Lawson DM, Mi K, Maxwell A. 2021.
2064 The pentapeptide-repeat protein, MfpA, interacts with mycobacterial DNA gyrase as a
2065 DNA T-segment mimic. *Proc Natl Acad Sci U S A* **118**: e2016705118.
2066 Ferenczi A, Pyott DE, Xipnitou A, Molnar A. 2017. Efficient targeted DNA editing and
2067 replacement in *Chlamydomonas reinhardtii* using Cpf1 ribonucleoproteins and single-
2068 stranded DNA. *Proc Natl Acad Sci U S A* **114**: 13567–13572.
2069 Figueroa-Martinez F, Nedelcu AM, Smith DR, Reyes-Prieto A. 2015. When the lights go out:
2070 the evolutionary fate of free-living colorless green algae. *New Phytologist* **206**: 972–982.
2071 Fischer G, Liti G, Llorente B. 2021. The budding yeast life cycle: More complex than
2072 anticipated? *Yeast* **38**: 5–11.
2073 Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020.
2074 RepeatModeler2 for automated genomic discovery of transposable element families. *Proc*
2075 *Natl Acad Sci U S A* **117**: 9451–9457.

Foflonker F, Mollegard D, Ong M, Yoon HS, Bhattacharya D. 2018. Genomic analysis of *Picochlorum* species reveals how microalgae may adapt to variable environments. *Mol Biol Evol* **35**: 2702–2711.

Franklin S, Somanchi A, Espina K, Rudenko G, Chua P. 2011. Renewable chemical production from novel fatty acid feedstocks. United States US7883882B2, filed November 30, 2009, and issued February 9, 2011.

Franklin S, Somanchi A, Wee J, Rudenko G, Moseley J, Rakitsky W. 2013. Tailored oils produced from recombinant heterotrophic microorganisms. United States US8592188B2, filed May 27, 2011, and issued November 26, 2023.

Fu Y, Luo GZ, Chen K, Deng X, Yu M, Han D, Hao Z, Liu J, Lu X, Dore LC et al. 2015. *N*6-methyldeoxyadenosine marks active transcription start sites in *Chlamydomonas*. *Cell* **161**: 879–892.

Fučiková K, Pažoutová M, Rindi F. 2015. Meiotic genes and sexual reproduction in the green algal class Trebouxiophyceae (Chlorophyta). *J Phycol* **51**: 419–430.

Fujiwara T, Ohnuma M, Yoshida M, Kuroiwa T, Hirano T. 2013. Gene targeting in the red alga *Cyanidioschyzon merolae*: single- and multi-copy insertion using authentic and chimeric selection markers. *PLoS One* **8**: e73608.

Fulnečková J, Hasíková T, Fajkus J, Lukešová A, Eliáš M, Sýkorová E. 2012. Dynamic evolution of telomeric sequences in the green algal order Chlamydomonadales. *Genome Biol Evo* **4**: 248–264.

Gabaldón T. 2020. Hybridization and the origin of new yeast lineages. *FEMS Yeast Res* **20**: foaa040.

Gallaher SD, Craig RJ, Ganesan I, Purvine SO, McCorkle SR, Grimwood J, Strenkert D, Davidi L, Roth MS, Jeffers TL et al. 2021. Widespread polycistronic gene expression in green algae. *Proc Natl Acad Sci U S A* **118**.

Gallone B, Steensels J, Mertens S, Dzialo MC, Gordon JL, Wauters R, Thesseling FA, Bellinazzo F, Saels V, Herrera-Malaver B et al. 2019. Interspecific hybridization facilitates niche adaptation in beer yeast. *Nat Ecol Evol* **3**: 1562–1575.

Gao C, Wang Y, Shen Y, Yan D, He X, Dai J, Wu Q. 2014. Oil accumulation mechanisms of the oleaginous microalga *Chlorella protothecoides* revealed through its genome, transcriptomes, and proteomes. *BMC Genomics* **15**: 582.

Garagna S, Page J, Fernandez-Donoso R, Zuccotti M, Searle JB. 2014. The Robertsonian phenomenon in the house mouse: mutation, meiosis and speciation. *Chromosoma* **123**: 529–544.

Gazquez A, Bordenave CD, Montero-Pau J, Pérez-Rodrigo M, Marco F, Martínez-Alberola F, Muggia L, Barreno E, Carrasco P. 2024. From spores to gametes: A sexual life cycle in a symbiotic *Trebouxia* microalga. *Algal Research* **84**: 103744.

Goold HD, Moseley JL, Lauersen KJ. 2024. The synthetic future of algal genomes. *Cell Genom* **4**: 100505.

Goubert C, Craig RJ, Bilat AF, Peona V, Vogan AA, Protasio AV. 2022. A beginner's guide to manual curation of transposable elements. *Mob DNA* **13**: 7.

Guo J, Jian J, Wang L, Xiong L, Lin H, Zhou Z, Sonnenschein EC, Wu W. 2022. Genome sequences of two strains of *Prototheca wickerhamii* provide insight into the protothecosis evolution. *Front Cell Infect Microbiol* **12**: 797017.

Harris EH. 2001. *Chlamydomonas* as a Model Organism. *Annu Rev Plant Physiol Plant Mol Biol* **52**: 363–406.

2122 He X, Dai J, Wu Q. 2016. Identification of sporopollenin as the outer layer of vell eall in
2123 microalga *Chlorella protothecoides*. *Front Microbiol* **7**: 1047.
2124 Heimerl N, Hommel E, Westermann M, Meichsner D, Lohr M, Hertweck C, Grossman AR,
2125 Mittag M, Sasso S. 2018. A giant type I polyketide synthase participates in zygospore
2126 maturation in *Chlamydomonas reinhardtii*. *Plant J* **95**: 268–281.
2127 Hirakawa MP, Martinez DA, Sakthikumar S, Anderson MZ, Berlin A, Gujja S, Zeng Q, Zisson
2128 E, Wang JM, Greenberg JM et al. 2015. Genetic and phenotypic intra-species variation in
2129 *Candida albicans*. *Genome Res* **25**: 413–425.
2130 Hirooka S, Itabashi T, Ichinose TM, Onuma R, Fujiwara T, Yamashita S, Jong LW, Tomita R,
2131 Iwane AH, Miyagishima SY. 2022. Life cycle and functional genomics of the unicellular
2132 red alga *Galdieria* for elucidating algal and plant evolution and industrial use. *Proc Natl*
2133 *Acad Sci U S A* **119**: e2210665119.
2134 Hisanaga T, Fujimoto S, Cui Y, Sato K, Sano R, Yamaoka S, Kohchi T, Berger F, Nakajima K.
2135 2021. Deep evolutionary origin of gamete-directed zygote activation by KNOX/BELL
2136 transcription factors in green plants. *Elife* **10**: e57090.
2137 Hobson DJ, Wei W, Steinmetz LM, Svejstrup JQ. 2012. RNA polymerase II collision interrupts
2138 convergent transcription. *Mol Cell* **48**: 365–374.
2139 Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised
2140 RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS.
2141 *Bioinformatics* **32**: 767–769.
2142 Hoff KJ, Stanke M. 2019. Predicting genes in single genomes with AUGUSTUS. *Curr Protoc*
2143 *Bioinformatics* **65**: e57.
2144 Hollingshead S, Kopecna J, Jackson PJ, Canniffe DP, Davison PA, Dickman MJ, Sobotka R,
2145 Hunter CN. 2012. Conserved chloroplast open-reading frame *ycf54* is required for
2146 activity of the magnesium protoporphyrin monomethylester oxidative cyclase in
2147 *Synechocystis* PCC 6803. *J Biol Chem* **287**: 27823–27833.
2148 Huff JT, Zilberman D. 2014. Dnmt1-independent CG methylation contributes to nucleosome
2149 positioning in diverse eukaryotes. *Cell* **156**: 1286–1297.
2150 Huss VAR, Frank C, Hartmann EC, Hirmer M, Kloboucek A, Seidel BM, Wenzeler P, Kessler
2151 E. 1999. Biochemical taxonomy and molecular phylogeny of the genus *Chlorella* sensu
2152 lato (Chlorophyta). *Journal of Phycology* **35**: 587–598.
2153 Huss VAR, Holweg C, Seidel B, Reich V, Rahat M, Kessler E. 1994. There is an ecological
2154 basis for host symbiont specificity in *Chlorella*/*Hydra* symbioses. *Endocyt Cell Res* **10**:
2155 35–46.
2156 Jacobebbinghaus N, Bigge F, Saudhof M, Hubner W, Kruse O, Baier T. 2025. Transcriptional
2157 gene fusions via targeted integration at safe harbors for high transgene expression in
2158 *Chlamydomonas reinhardtii*. *New Phytol* doi:10.1111/nph.70368.
2159 Jagielski T, Bakuła Z, Gawor J, Maciszewski K, Kusber WH, Dylag M, Nowakowska J,
2160 Gromadka R, Karnkowska A. 2019. The genus *Prototheca* (Trebouxiophyceae,
2161 Chlorophyta) revisited: Implications from molecular taxonomic studies. *Algal Res* **43**:
2162 101639.
2163 Joo S, Nishimura Y, Cronmiller E, Hong RH, Kariyawasam T, Wang MH, Shao NC, El Akkad
2164 SE, Suzuki T, Higashiyama T et al. 2017. Gene regulatory networks for the haploid-to-
2165 diploid transition of *Chlamydomonas reinhardtii*. *Plant Physiol* **175**: 314–332.

2166 Joo S, Wang MH, Lui G, Lee J, Barnas A, Kim E, Sudek S, Worden AZ, Lee JH. 2018.
2167 Common ancestry of heterodimerizing TALE homeobox transcription factors across
2168 Metazoa and Archaeplastida. *BMC Biol* **16**: 136.
2169 Kalina T, Punčochárová M. 1987. Taxonomy of the subfamily Scotiellocystoideae Fott 1976
2170 (Chlorellaceae, Chlorophyceae). *Algol Stud* **45**: 473–521.
2171 Kanesaki Y, Imamura S, Matsuzaki M, Tanaka K. 2015. Identification of centromere regions in
2172 chromosomes of a unicellular red alga, *Cyanidioschyzon merolae*. *Febs Letters* **589**:
2173 1219–1224.
2174 Karpowicz SJ, Prochnik SE, Grossman AR, Merchant SS. 2011. The GreenCut2 resource, a
2175 phylogenomically derived inventory of proteins specific to the plant lineage. *J Biol Chem*
2176 **286**: 21427–21439.
2177 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
2178 improvements in performance and usability. *Mol Biol Evol* **30**: 772–780.
2179 Khan H, Ochi T. 2023. Plant PAXX has an XLF-like function and stimulates DNA end joining
2180 by the Ku-DNA ligase IV/XRCC4 complex. *Plant J* **116**: 58–68.
2181 Kieselbach T, Mant A, Robinson C, Schroder WP. 1998. Characterisation of an *Arabidopsis*
2182 cDNA encoding a thylakoid lumen protein related to a novel 'pentapeptide repeat' family
2183 of proteins. *FEBS Lett* **428**: 241–244.
2184 Kim SS, Grienenberger E, Lallemand B, Colpitts CC, Kim SY, Souza Cde A, Geoffroy P, Heintz
2185 D, Krahn D, Kaiser M et al. 2010. *LAP6/POLYKETIDE SYNTHASE A* and
2186 *LAP5/POLYKETIDE SYNTHASE B* encode hydroxyalkyl alpha-pyrone synthases
2187 required for pollen development and sporopollenin biosynthesis in *Arabidopsis thaliana*.
2188 *Plant Cell* **22**: 4045–4066.
2189 Kindle KL. 1990. High-frequency nuclear transformation of *Chlamydomonas reinhardtii*. *Proc*
2190 *Natl Acad Sci U S A* **87**: 1228–1232.
2191 Kindle KL, Schnell RA, Fernandez E, Lefebvre PA. 1989. Stable nuclear transformation of
2192 *Chlamydomonas* using the *Chlamydomonas* gene for nitrate reductase. *J Cell Biol* **109**:
2193 2589–2601.
2194 Kipreos ET, Pagano M. 2000. The F-box protein family. *Genome Biol* **1**: REVIEWS3002.
2195 Kolesiński P, Golik P, Grudnik P, Piechota J, Markiewicz M, Tarnawski M, Dubin G,
2196 Szczepaniak A. 2013. Insights into eukaryotic Rubisco assembly - crystal structures of
2197 RbcX chaperones from *Arabidopsis thaliana*. *Biochim Biophys Acta* **1830**: 2899–2906.
2198 Kolesiński P, Piechota J, Szczepaniak A. 2011. Initial characteristics of RbcX proteins from
2199 *Arabidopsis thaliana*. *Plant Mol Biol* **77**: 447–459.
2200 Krasovec M, Merret R, Sanchez F, Sanchez-Brosseau S, Piganeau G. 2023. A high frequency of
2201 chromosomal duplications in unicellular algae is compensated by translational regulation.
2202 *Genome Biol Evol* **15**: evad086.
2203 Krishnan A, Dahlin LR, Guarnieri MT, Weissman JC, Posewitz MC. 2025. Small cells with big
2204 photosynthetic productivities: biotechnological potential of the *Picochlorum* genus.
2205 *Trends Biotechnol* **43**: 759–772.
2206 Kropat J, Hong-Hermesdorf A, Casero D, Ent P, Castruita M, Pellegrini M, Merchant SS,
2207 Malasarn D. 2011. A revised mineral nutrient supplement increases biomass and growth
2208 rate in *Chlamydomonas reinhardtii*. *Plant J* **66**: 770–780.
2209 Krüger W. 1894a. Beiträge zur Kenntnis der Organismen des Saftflusses (sog. Schleimflusses)
2210 der Laubbäume. II. Über zwei aus Saftflüssen rein gezüchtete Algen. *Zopf Beitr Physiol*
2211 *Morphol Organ* **4**: 91–116.

2212 Krüger W. 1894b. Kurze Charakteristik einiger niederer Organismen im Saftflusse der
2213 Laubbäume. *Hedwigia* **33**: 241–266.
2214 Kwon T, Hanschen ER, Hovde BT. 2023. Addressing the pervasive scarcity of structural
2215 annotation in eukaryotic algae. *Sci Rep* **13**: 1687.
2216 Lancaster SM, Payen C, Smukowski Heil C, Dunham MJ. 2019. Fitness benefits of loss of
2217 heterozygosity in *Saccharomyces* hybrids. *Genome Res* **29**: 1685–1692.
2218 Langdon QK, Peris D, Baker EP, Opulente DA, Nguyen HV, Bond U, Goncalves P, Sampaio JP,
2219 Libkind D, Hittinger CT. 2019. Fermentation innovation through complex hybridization
2220 of wild and domesticated yeasts. *Nat Ecol Evol* **3**: 1576–1586.
2221 Lee JH, Lin H, Joo S, Goodenough U. 2008. Early sexual origins of homeoprotein
2222 heterodimerization and evolution of the plant KNOX/BELL family. *Cell* **133**: 829–840.
2223 Lemieux C, Turmel M, Otis C, Pombert JF. 2019. A streamlined and predominantly diploid
2224 genome in the tiny marine green alga *Chloropicon primus*. *Nat Commun* **10**: 4061.
2225 Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
2226 *Arxiv* doi:10.48550/arXiv.1303.3997.
2227 Li H. 2021. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**: 4572–
2228 4574.
2229 Li X, Patena W, Fauser F, Jinkerson RE, Saroussi S, Meyer MT, Ivanova N, Robertson JM, Yue
2230 R, Zhang R et al. 2019. A genome-wide algal mutant library and functional screen
2231 identifies genes required for eukaryotic photosynthesis. *Nat Genet* **51**: 627–635.
2232 Li X, Zhang R, Patena W, Gang SS, Blum SR, Ivanova N, Yue R, Robertson JM, Lefebvre PA,
2233 Fitz-Gibbon ST et al. 2016. An indexed, mapped mutant library enables reverse genetics
2234 studies of biological processes in *Chlamydomonas reinhardtii*. *Plant Cell* **28**: 367–387.
2235 Lin J, Le TV, Augspurger K, Tritschler D, Bower R, Fu G, Perrone C, O'Toole ET, Mills KV,
2236 Dymek E et al. 2019. FAP57/WDR65 targets assembly of a subset of inner arm dyneins
2237 and connects to regulatory hubs in cilia. *Mol Biol Cell* **30**: 2659–2680.
2238 Lin WR, Ng IS. 2020. Development of CRISPR/Cas9 system in *Chlorella vulgaris* FSP-E to
2239 enhance lipid accumulation. *Enzyme Microb Technol* **133**: 109458.
2240 Liu C, Young AL, Starling-Windhof A, Bracher A, Saschenbrecker S, Rao BV, Rao KV,
2241 Berninghausen O, Mielke T, Hartl FU et al. 2010. Coupled chaperone action in folding
2242 and assembly of hexadecameric Rubisco. *Nature* **463**: 197–202.
2243 Liu HW, Khera R, Grob P, Gallaher SD, Purvine SO, Nicora CD, Lipton MS, Niyogi KK,
2244 Nogales E, Iwai M et al. 2025. A distinct LHCI arrangement is recruited to photosystem I
2245 in Fe-starved green algae. *Proc Natl Acad Sci U S A* **122**: e2500621122.
2246 Loera-Quezada MM, Leyva-Gonzalez MA, Velazquez-Juarez G, Sanchez-Calderon L, Do
2247 Nascimento M, Lopez-Arredondo D, Herrera-Estrella L. 2016. A novel genetic
2248 engineering platform for the effective management of biological contaminants for the
2249 production of microalgae. *Plant Biotechnol J* **14**: 2066–2076.
2250 Lopez D, Hamaji T, Kropat J, De Hoff P, Morselli M, Rubbi L, Fitz-Gibbon S, Gallaher SD,
2251 Merchant SS, Umen J et al. 2015. Dynamic changes in the transcriptome and methylome
2252 of *Chlamydomonas reinhardtii* throughout Its life cycle. *Plant Physiol* **169**: 2730–2743.
2253 Lopez-Cortegano E, Craig RJ, Chebib J, Balogun EJ, Keightley PD. 2023. Rates and spectra of
2254 de novo structural mutations in *Chlamydomonas reinhardtii*. *Genome Res* **33**: 45–60.
2255 Lozano JC, Schatt P, Botebol H, Verge V, Lesuisse E, Blain S, Carre IA, Bouget FY. 2014.
2256 Efficient gene targeting and removal of foreign DNA by homologous recombination in
2257 the picoeukaryote *Ostreococcus*. *Plant J* **78**: 1073–1083.

2258 MacIntosh GC. 2011. RNase T2 family: enzymatic properties, functional diversity, and evolution
2259 of ancient ribonucleases. In *Nucleic Acids Mol Bi*, Vol 26 (ed. AW Nicholson), pp. 89–
2260 114.
2261 Mancera E, Bourgon R, Brozzi A, Huber W, Steinmetz LM. 2008. High-resolution mapping of
2262 meiotic crossovers and non-crossovers in yeast. *Nature* **454**: 479–485.
2263 Manni M, Berkeley MR, Seppey M, Simao FA, Zdobnov EM. 2021. BUSCO update: novel and
2264 streamlined workflows along with broader and deeper phylogenetic coverage for scoring
2265 of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* **38**: 4647–4654.
2266 Marcet-Houben M, Gabaldón T. 2015. Beyond the whole-genome duplication: phylogenetic
2267 evidence for an ancient interspecies hybridization in the baker's yeast lineage. *PLoS Biol*
2268 **13**: e1002220.
2269 Marcolungo L, Bellamoli F, Cecchin M, Lopatriello G, Rossato M, Cosentino E, Rombauts S,
2270 Delledonne M, Ballottari M. 2024. *Haematococcus lacustris* genome assembly and
2271 annotation reveal diploid genetic traits and stress-induced gene expression patterns. *Algal*
2272 *Res* **80**.
2273 Maruyama S, Sugahara J, Kanai A, Nozaki H. 2010. Permuted tRNA genes in the nuclear and
2274 nucleomorph genomes of photosynthetic eukaryotes. *Mol Biol Evol* **27**: 1070–1076.
2275 Matsuka M, Otsuka H, Hase E. 1966. Changes in contents of carbohydrate and fatty acid in cells
2276 of *Chlorella protothecoides* during processes of de- and re-generation of chloroplasts.
2277 *Plant Cell Physiol* **7**: 651–662.
2278 Mattei AL, Bailly N, Meissner A. 2022. DNA methylation: a historical perspective. *Trends*
2279 *Genet* **38**: 676–707.
2280 Merchant SS Prochnik SE Vallon O Harris EH Karpowicz SJ Witman GB Terry A Salamov A
2281 Fritz-Laylin LK Marechal-Drouard L et al. 2007. The *Chlamydomonas* genome reveals
2282 the evolution of key animal and plant functions. *Science* **318**: 245–250.
2283 Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear
2284 R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the
2285 genomic era. *Mol Biol Evol* **37**: 1530–1534.
2286 Minoda A, Sakagami R, Yagisawa F, Kuroiwa T, Tanaka K. 2004. Improvement of culture
2287 conditions and evidence for nuclear transformation by homologous recombination in a
2288 red alga, *Cyanidioschyzon merolae* 10D. *Plant Cell Physiol* **45**: 667–671.
2289 Mock H-P. 2001. Photosensitizing tetrapyrroles induce antioxidative and pathogen defense
2290 responses in plants. In *Oxidative Stress in Plants*, (ed. D Inze, M Van Montagu), pp.
2291 207–226. CRC Press, London.
2292 Moore ER, Bullington BS, Weisberg AJ, Jiang Y, Chang J, Halsey KH. 2017. Morphological
2293 and transcriptomic evidence for ammonium induction of sexual reproduction in
2294 *Thalassiosira pseudonana* and other centric diatoms. *PLoS One* **12**: e0181098.
2295 Moseley JL, Lee B, Im CS, Kim J, Kim D, Bhat R. 2024. Production of lipids and terpenoids in
2296 *Auxenochlorella protothecoides*. United States US12037630B2, filed November 5, 2021,
2297 and issued July 16, 2024.
2298 Neupert J, Gallaher SD, Lu Y, Strenkert D, Segal N, Barahimipour R, Fitz-Gibbon ST, Schroda
2299 M, Merchant SS, Bock R. 2020. An epigenetic gene silencing pathway selectively acting
2300 on transgenic DNA in the green alga *Chlamydomonas*. *Nat Commun* **11**: 6269.
2301 Nievergelt AP. 2025. Genome editing in the green alga *Chlamydomonas*: past, present practice
2302 and future prospects. *Plant J* **122**: e70140.

2303 Nievergelt AP, Diener DR, Bogdanova A, Brown T, Pigino G. 2023. Efficient precision editing
2304 of endogenous *Chlamydomonas reinhardtii* genes with CRISPR-Cas. *Cell Rep Methods*
2305 **3**: 100562.
2306 Ning J, Otto TD, Pfander C, Schwach F, Brochet M, Bushell E, Goulding D, Sanders M,
2307 Lefebvre PA, Pei J et al. 2013. Comparative genomics in *Chlamydomonas* and
2308 *Plasmodium* identifies an ancient nuclear envelope protein family essential for sexual
2309 reproduction in protists, fungi, plants, and vertebrates. *Genes Dev* **27**: 1198–1215.
2310 Orr-Weaver TL, Szostak JW, Rothstein RJ. 1981. Yeast transformation: a model system for the
2311 study of recombination. *Proc Natl Acad Sci U S A* **78**: 6354–6358.
2312 Pardo-Palacios FJ, Arzalluz-Luque A, Kondratova L, Salguero P, Mestre-Tomas J, Amorin R,
2313 Estevan-Morio E, Liu T, Nanni A, McIntyre L et al. 2024. SQANTI3: curation of long-
2314 read transcriptomes for accurate identification of known and novel isoforms. *Nat*
2315 *Methods* **21**:793-797.
2316 Park SH, Kyndt JA, Brown JK. 2022. Comparison of *Auxenochlorella protothecoides* and
2317 *Chlorella* spp. chloroplast genomes: evidence for endosymbiosis and horizontal virus-like
2318 gene transfer. *Life (Basel)* **12**: 458.
2319 Peris D, Perez-Torrado R, Hittinger CT, Barrio E, Querol A. 2018. On the origins and industrial
2320 applications of *Saccharomyces cerevisiae* x *Saccharomyces kudriavzevii* hybrids. *Yeast*
2321 **35**: 51–69.
2322 Petroll R, Papareddy RK, Krela R, Laigle A, Riviere Q, Bišová K, Mozgová I, Borg M. 2025.
2323 The expansion and diversification of epigenetic regulatory networks underpins major
2324 transitions in the evolution of land plants. *Mol Biol Evol* **42**: msaf064.
2325 Pombert JF, Blouin NA, Lane C, Boucias D, Keeling PJ. 2014. A lack of parasitic reduction in
2326 the obligate parasitic green alga *Helicosporidium*. *PLoS Genet* **10**: e1004355.
2327 Pontes A, Cadez N, Goncalves P, Sampaio JP. 2019. A quasi-domesticate relic hybrid population
2328 of *Saccharomyces cerevisiae* x *S. paradoxus* adapted to olive brine. *Front Genet* **10**: 449.
2329 Poplin R, Chang PC, Alexander D, Schwartz S, Colthurst T, Ku A, Newburger D, Dijamco J,
2330 Nguyen N, Afshar PT et al. 2018. A universal SNP and small-indel variant caller using
2331 deep neural networks. *Nat Biotechnol* **36**: 983–987.
2332 Pore RS. 1972. Nutritional basis for relating *Prototheca* and *Chlorella*. *Can J Microbiol* **18**:
2333 1175–1177.
2334 Prescott EM, Proudfoot NJ. 2002. Transcriptional collision between convergent genes in budding
2335 yeast. *Proc Natl Acad Sci U S A* **99**: 8796–8801.
2336 Pryszcz LP, Nemeth T, Saus E, Ksiezopolska E, Hegedusova E, Nosek J, Wolfe KH, Gacser A,
2337 Gabaldon T. 2015. The genomic aftermath of hybridization in the opportunistic pathogen
2338 *Candida metapsilosis*. *PLoS Genet* **11**: e1005626.
2339 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic
2340 features. *Bioinformatics* **26**: 841–842.
2341 Randolph-Anderson BL, Boynton JE, Gillham NW, Harris EH, Johnson AM, Dorthu MP,
2342 Matagne RF. 1993. Further characterization of the respiratory deficient dum-1 mutation
2343 of *Chlamydomonas reinhardtii* and its use as a recipient for mitochondrial
2344 transformation. *Mol Gen Genet* **236**: 235–244.
2345 Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP.
2346 2011. Integrative Genomics Viewer. *Nat Biotechnol* **29**: 24–26.
2347 Romero Charria P, Navarrete C, Ovchinnikov V, Sarre LA, Shabardina V, Casacuberta E, Lara-
2348 Astiaso D, Sebé-Pedrós A, de Mendoza A. 2024. Adenine DNA methylation associated

2349 to transcription is widespread across eukaryotes. *bioRxiv*
2350 doi:10.1101/2024.10.28.620566: 2024.2010.2028.620566.
2351 Saito M, Xu P, Faure G, Maguire S, Kannan S, Altae-Tran H, Vo S, Desimone A, Macrae RK,
2352 Zhang F. 2023. Fanzor is a eukaryotic programmable RNA-guided endonuclease. *Nature*
2353 **620**: 660–668.
2354 Salome PA, Merchant SS. 2019. A series of fortunate events: introducing Chlamydomonas as a
2355 reference organism. *Plant Cell* **31**: 1682–1707.
2356 Sanders CK, Hanschen ER, Biondi TC, Hovde BT, Kunde YA, Eng WL, Kwon T, Dale T. 2022.
2357 Phylogenetic analyses and reclassification of the oleaginous marine species *Nannochloris*
2358 sp. "desiccata" (Trebouxiophyceae, Chlorophyta), formerly *Chlorella desiccata*,
2359 supported by a high-quality genome assembly. *J Phycol* **58**: 436–448.
2360 Santaguida S, Amon A. 2015. Short- and long-term effects of chromosome mis-segregation and
2361 aneuploidy. *Nat Rev Mol Cell Biol* **16**: 473–485.
2362 Saschenbrecker S, Bracher A, Rao KV, Rao BV, Hartl FU, Hayer-Hartl M. 2007. Structure and
2363 function of RbcX, an assembly chaperone for hexadecameric Rubisco. *Cell* **129**: 1189–
2364 1200.
2365 Sasso S, Stibor H, Mittag M, Grossman AR. 2018. From molecular manipulation of
2366 domesticated *Chlamydomonas reinhardtii* to survival in nature. *Elife* **7**: e39233.
2367 Schroda M. 2019. Good news for nuclear transgene expression in *Chlamydomonas*. *Cells* **8**:
2368 1534.
2369 Schroda M, Remacle C. 2022. Molecular advancements establishing Chlamydomonas as a host
2370 for biotechnological exploitation. *Front Plant Sci* **13**: 911483.
2371 Schroder MS, Martinez de San Vicente K, Prandini TH, Hammel S, Higgins DG, Bagagli E,
2372 Wolfe KH, Butler G. 2016. Multiple origins of the pathogenic yeast *Candida*
2373 *orthopsilosis* by separate hybridizations between two parental species. *PLoS Genet* **12**:
2374 e1006404.
2375 Schubert M, Petersson UA, Haas BJ, Funk C, Schroder WP, Kieselbach T. 2002. Proteome map
2376 of the chloroplast lumen of *Arabidopsis thaliana*. *J Biol Chem* **277**: 8354–8365.
2377 Selmecki A, Forche A, Berman J. 2006. Aneuploidy and isochromosome formation in drug-
2378 resistant *Candida albicans*. *Science* **313**: 367–370.
2379 Seto A, Wang HL, Hesseltine CW. 1984. Culture conditions affect eicosapentaenoic acid content
2380 of *Chlorella minutissima*. *J Am Oil Chem Soc* **61**: 892–894.
2381 Severgnini M, Lazzari B, Capra E, Chessa S, Luini M, Bordoni R, Castiglioni B, Ricchi M,
2382 Cremonesi P. 2018. Genome sequencing of *Prototheca zopfii* genotypes 1 and 2 provides
2383 evidence of a severe reduction in organellar genomes. *Sci Rep* **8**: 14637.
2384 Shihira-Ishikawa I, Hase E. 1964. Nutritional control of cell pigmentation in *Chlorella*
2385 *protothecoides* with special reference to the degeneration of chloroplast induced by
2386 glucose. *Plant Cell Physiol* **5**: 227–240.
2387 Shimogawara K, Fujiwara S, Grossman A, Usuda H. 1998. High-efficiency transformation of
2388 *Chlamydomonas reinhardtii* by electroporation. *Genetics* **148**: 1821–1828.
2389 Sipiczki M. 2008. Interspecies hybridization and recombination in *Saccharomyces* wine yeasts.
2390 *FEMS Yeast Res* **8**: 996–1007.
2391 Smukowski Heil CS, DeSevo CG, Pai DA, Tucker CM, Hoang ML, Dunham MJ. 2017. Loss of
2392 heterozygosity drives adaptation in hybrid yeast. *Mol Biol Evol* **34**: 1596–1612.
2393 Soltis PS, Soltis DE. 2009. The role of hybridization in plant speciation. *Annu Rev Plant Biol* **60**:
2394 561–588.

Soma A, Onodera A, Sugahara J, Kanai A, Yachie N, Tomita M, Kawamura F, Sekine Y. 2007. Permuted tRNA genes expressed via a circular RNA intermediate in *Cyanidioschyzon merolae*. *Science* **318**: 450–453.

Sproles AE, Fields FJ, Smalley TN, Le CH, Badary A, Mayfield SP. 2021. Recent advancements in the genetic engineering of microalgae. *Algal Research* **53**: 102158.

Stadler R, Wolf K, Hilgarth C, Tanner W, Sauer N. 1995. Subcellular localization of the inducible *Chlorella* HUP1 monosaccharide-H+ symporter and cloning of a co-Induced galactose-H+ symporter. *Plant Physiol* **107**: 33–41.

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.

Steenwyk JL, Lind AL, Ries LNA, Dos Reis TF, Silva LP, Almeida F, Bastos RW, Fraga da Silva TFC, Bonato VLD, Pessoni AM et al. 2020. Pathogenic allodiploid hybrids of *Aspergillus* fungi. *Curr Biol* **30**: 2495–2507.e7.

Strenkert D, Schmollinger S, Gallaher SD, Salome PA, Purvine SO, Nicora CD, Mettler-Altmann T, Soubeyrand E, Weber APM, Lipton MS et al. 2019. Multiomics resolution of molecular events during a day in the life of Chlamydomonas. *Proc Natl Acad Sci U S A* **116**: 2374–2383.

Strenkert D, Yildirim A, Yan J, Yoshinaga Y, Pellegrini M, O'Malley RC, Merchant SS, Umen JG. 2022. The landscape of Chlamydomonas histone H3 lysine 4 methylation reveals both constant features and dynamic changes during the diurnal cycle. *Plant J* **112**: 352–368.

Sui Y, Qi L, Wu JK, Wen XP, Tang XX, Ma ZJ, Wu XC, Zhang K, Kokoska RJ, Zheng DQ et al. 2020. Genome-wide mapping of spontaneous genetic alterations in diploid yeast cells. *Proc Natl Acad Sci U S A* **117**: 28191–28200.

Suzuki S, Endoh R, Manabe R, Ohkuma M, Hirakawa Y. 2018. Multiple losses of photosynthesis and convergent reductive genome evolution in the colourless green algae *Prototheca*. *Sci Rep* **8**: 940.

Tamura K, Stecher G, Kumar S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol* **38**: 3022–3027.

Tanaka S, Sawaya MR, Kerfeld CA, Yeates TO. 2007. Structure of the RuBisCO chaperone RbcX from *Synechocystis* sp. PCC6803. *Acta Crystallogr D Biol Crystallogr* **63**: 1109–1112.

Tang S, Lomsadze A, Borodovsky M. 2015. Identification of protein coding regions in RNA transcripts. *Nucleic Acids Res* **43**: e78.

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 Res* **45**: W6–W11.

Torres-Machorro AL, Hernandez R, Cevallos AM, López-Villaseñor I. 2010. Ribosomal RNA genes in eukaryotic microorganisms: witnesses of phylogeny? *FEMS Microbiol Rev* **34**: 59–86.

Tottey S, Block MA, Allen M, Westergren T, Albrieux C, Scheller HV, Merchant S, Jensen PE. 2003. *Arabidopsis* CHL27, located in both envelope and thylakoid membranes, is required for the synthesis of protochlorophyllide. *Proc Natl Acad Sci U S A* **100**: 16119–16124.

2439 Trincado JL, Entizne JC, Hysenaj G, Singh B, Skalic M, Elliott DJ, Eyras E. 2018. SUPPA2:
2440 fast, accurate, and uncertainty-aware differential splicing analysis across multiple
2441 conditions. *Genome Biol* **19**: 40.
2442 Tsai IJ, Bensasson D, Burt A, Koufopanou V. 2008. Population genomics of the wild yeast
2443 *Saccharomyces paradoxus*: Quantifying the life cycle. *Proc Natl Acad Sci U S A* **105**:
2444 4957–4962.
2445 Ueno R, Hanagata N, Urano N, Suzuki M. 2005. Molecular phylogeny and phenotypic variation
2446 in the heterotrophic green algal genus *Prototheca* (Trebouxiophyceae, Chlorophyta).
2447 *Journal of Phycology* **41**: 1268–1280.
2448 Vande Zande P, Zhou X, Selmecki A. 2023. The dynamic fungal genome: polyploidy,
2449 aneuploidy and copy number variation in response to stress. *Annu Rev Microbiol* **77**:
2450 341–361.
2451 Vetting MW, Hegde SS, Fajardo JE, Fiser A, Roderick SL, Takiff HE, Blanchard JS. 2006.
2452 Pentapeptide repeat proteins. *Biochemistry* **45**: 1–10.
2453 Vogler BW, Starkenburg SR, Sudasinghe N, Schambach JY, Rollin JA, Pattathil S, Barry AN.
2454 2018. Characterization of plant carbon substrate utilization by *Auxenochlorella*
2455 *protothecoides*. *Algal Research* **34**: 37–48.
2456 Wang B, Jia Y, Dang N, Yu J, Bush SJ, Gao S, He W, Wang S, Guo H, Yang X et al. 2024. Near
2457 telomere-to-telomere genome assemblies of two *Chlorella* species unveil the composition
2458 and evolution of centromeres in green algae. *BMC Genomics* **25**: 356.
2459 Werner A, Kanhere A, Wahlestedt C, Mattick JS. 2024. Natural antisense transcripts as versatile
2460 regulators of gene expression. *Nat Rev Genet* **25**: 730–744.
2461 Wolff G, Burger G, Lang BF, Kück U. 1993. Mitochondrial genes in the colourless alga
2462 *Prototheca wickerhamii* resemble plant genes in their exons but fungal genes in their
2463 introns. *Nucleic Acids Res* **21**: 719–726.
2464 Yan D, Dai J, Wu Q. 2013. Characterization of an ammonium transporter in the oleaginous alga
2465 *Chlorella protothecoides*. *Appl Microbiol Biotechnol* **97**: 919–928.
2466 Yan D, Wang Y, Murakami T, Shen Y, Gong JH, Jiang HF, Smith DR, Pombert JF, Dai JBA,
2467 Wu QY. 2015. *Auxenochlorella protothecoides* and *Prototheca wickerhamii* plastid
2468 genome sequences give insight into the origins of non-photosynthetic algae. *Sci Rep* **5**:
2469 14465.
2470 Yang B, Liu J, Jiang Y, Chen F. 2016. *Chlorella* species as hosts for genetic engineering and
2471 expression of heterologous proteins: Progress, challenge and perspective. *Biotechnol J*
2472 **11**: 1244–1261.
2473 Yona AH, Manor YS, Herbst RH, Romano GH, Mitchell A, Kupiec M, Pilpel Y, Dahan O. 2012.
2474 Chromosomal duplication is a transient evolutionary solution to stress. *Proc Natl Acad*
2475 *Sci U S A* **109**: 21010–21015.
2476 Yoon PH, Skopintsev P, Shi H, Chen L, Adler BA, Al-Shimary M, Craig RJ, Loi KJ, DeTurk
2477 EC, Li Z et al. 2023. Eukaryotic RNA-guided endonucleases evolved from a unique clade
2478 of bacterial enzymes. *Nucleic Acids Res* **51**: 12414–12427.
2479 Yuen KW, Warren CD, Chen O, Kwok T, Hieter P, Spencer FA. 2007. Systematic genome
2480 instability screens in yeast and their potential relevance to cancer. *Proc Natl Acad Sci U S*
2481 *A* **104**: 3925–3930.
2482 Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-wide evolutionary analysis of
2483 eukaryotic DNA methylation. *Science* **328**: 916–919.
2484