A streamlined and predominantly diploid genome in the tiny marine green alga Chloropicon primus

Claude Lemieux1,3, Monique Turmel1,3, Christian Otis1 & Jean-François Pombert2

Tiny marine green algae issued from two deep branches of the Chlorophyta, the Mamiellophyceae and Chloropicophyceae, dominate different regions of the oceans and play key roles in planktonic communities. Considering that the Mamiellophyceae is the sole lineage of prasinophyte algae that has been intensively investigated, the extent to which these two algal groups differ in their metabolic capacities and cellular processes is currently unknown. To address this gap of knowledge, we investigate here the nuclear genome sequence of a member of the Chloropicophyceae, Chloropicon primus. Among the main biological insights that emerge from this 17.4 Mb genome, we find an unexpected diploid structure for most chromosomes and a propionate detoxification pathway in green algae. Our results support the notion that separate events of genome minimization, which entailed differential losses of genes/pathways, have occurred in the Chloropicophyceae and Mamiellophyceae, suggesting different strategies of adaptation to oceanic environments.

1Département de biochimie, de microbiologie et de bio-informatique, Institut de Biologie Intégrative et des Systèmes, Université Laval, Québec, QC G1V 0A6, Canada. 2Department of Biology, Illinois Institute of Technology, Chicago, IL 60616, USA. 3These authors contributed equally: Claude Lemieux, Monique Turmel. Correspondence and requests for materials should be addressed to C.L. (email: claude.lemieux@bcm.ulaval.ca)
Green algae are morphologically, ecologically, and phylogenetically diverse, and represent one of the most successful groups of photosynthetic eukaryotes; yet our knowledge about the genetic factors determining their ecological success is still limited. Soon after their emergence about 1000 Mya, early green algae split into two phyla: Chlorophyta, comprising the vast majority of extant species, and Streptophyta, comprising charophyte algae and all land plants.1,2 The genomes of a number of photosynthetic green algae, especially those from the Chlorophyta, have been sequenced to unravel the evolutionary trajectories followed by ancestral green algae,4–7 decipher cellular mechanisms such as the basis of multicellularity,8,9 and investigate metabolic networks, including routes for the biosynthesis of highly valuable compounds for the industry.16–19 Known for their important role in the global carbon cycle, the green algae traditionally considered as prasinophytes constitute a genetically diverse group that encompass the vast majority of extant species, and Streptophyta, which comprises charophyte algae and all land plants.1,2 The algae of our study group belong to the highlighted lineages.2,3

**Fig. 1** Phylogenetic position and genome features of *Chloropicon* relative to the 12 green algae selected for comparative analyses. **a** Schematic diagram depicting the main lineages of the Chloroplastida. The algae of our study group belong to the highlighted lineages. **b** Nuclear phylogenomic tree inferred from 792 concatenated proteins encoded by single-copy genes and comparison of genome sizes, numbers of chromosomes/DNA assemblies, and sizes of protein-coding gene repertoires. **c** Venn diagram showing the numbers of protein-coding genes that *Chloropicon* shares with the Mamiellophyceae, core chlorophytes and *Klebsormidium*. **d** Density of SNPs and small indels in coding (blue bars) and non-coding (intron, orange bars; intergenic sequences, gray bars) regions of individual *Chloropicon* chromosomes. The source data of **d** are provided as a Source Data file.
flagella)—are typically found in coastal waters although they have been observed in open oceanic waters\(^2\). For each of these genera, cryptic species derived from separate clades or strains representing different ecotypes were targeted for genome analysis\(^1,7,24,25\). All examined mamiellaleans possess reduced genomes (12–22 Mb with only ~8000–11,000 predicted genes, Fig. 1b) compared to freshwater/terrestrial green algae and the prasinophyte *Cymbomonas tetramitiformis*\(^26\), implying that genome minimization occurred along cell reduction and simplification in the course of prasinophyte evolution\(^22\). The analysis\(^4\) presenting different ecotypes were targeted for genome which includes two genera (*Chloropicon* and *Chloroparvula*), viruses\(^27\), while the larger appears to contain the sex-determining chromosomes displaying lower GC content. The small outlier species divergence or adaptation to ecological niches was linked to chromosome losses, all of which were sequenced from telomere to telomere, nuclear genome was assembled into 20 chromosomes, with averages of 0.50 gene per kb and 0.50 predicted intron (average size of 152 bp) per gene (Supplementary Data 4), guanylyl cyclases (30 members, Supplementary Data 5), and sialyltransferases (59 members, Supplementary Data 3). The latter interpretation is supported by the analysis of the genome, an intermediate value between the *Micromonas* and *Chloropicon* genomes (12 times, Supplementary Fig. 1). Most Chloropicon chromosomes have a diploid structure. Polyomorphisms are pervasive in the *Chloropicon* genome, an unexpected discovery considering that all unicellular green algae analyzed so far are haploids and that the *Chloropicon* culture was derived from a clone. A total of 7282 single-nucleotide polymorphisms (SNPs) and 2048 small insertions/deletions (indels) were identified by aligning the Illumina reads against the genome sequence, for an average of 0.54 variants per kb (i.e. 0.05% divergence), with minimum and maximum values of 0.17 and 1.24 for chromosomes XVII and XX, respectively (Fig. 1d and Supplementary Data 7). Each polymorphic site features two alleles and two allelic combinations are readily distinguishable in aligned reads spanning two or more SNP sites. Except for a few regions, the polymorphic sites are distributed uniformly along each chromosome (Fig. 3). In coding regions which represent 73.8% of the genome, SNPs and small indels showed a 3.5- to 3.8-fold lower density than in intergenic regions and introns, respectively (Fig. 1d). Furthermore, 73 indels larger than 100 bp were identified by mapping PacBio reads onto the chromosomal assemblies (Fig. 3 and Supplementary Data 8). For all chromosomes, except IV and XVII, plots of the allelic frequencies observed at each polymorphic locus displayed a bell shape distribution with a mean of 50% (i.e. 1:1 allelic ratio), a result consistent with a diploid structure (Fig. 4a). However, the distribution observed for chromosome XVII was bimodal with means of 33 and 66%, as predicted for a trisomic state (Fig. 4b). The latter interpretation is supported by the finding that the coverage depth estimated for chromosome XVII was 1.5-fold greater relative to other chromosomes (Fig. 4d). With regards to
chromosome IV, the allelic frequency distribution appeared to be a composite of the two patterns just described (Fig. 4c), and strikingly, the polymorphic loci strongly departing from the 1:1 ratio are tightly clustered in a region covering 420 kb (Fig. 5). Note that shorter segments on other chromosomes also revealed aberrant allelic ratios (Supplementary Fig. 3).

**Cellular copy number of Chloropicon organelle genomes.** No polymorphisms were detected by mapping the sequencing reads generated in the course of this study against the 64.3-kb plastome and 37.9-kb mitogenome reference sequences available for *C. primus*.\(^{20,31}\) Comparison of the coverage depth of PacBio subreads (449x and 253x for the plastome and mitogenome, respectively) to the mean PacBio sequencing depth obtained for the 19 chromosomes present in two copies (73×) enabled us to estimate that each *Chloropicon* cell, which contains a single chloroplast and a single mitochondrion, houses about 12 plastid and seven mitochondrial DNA molecules. These values are respectively 7- and 9-fold lower compared to those reported for *C. reinhardtii* (over 80 copies of the 205-kb plastome and about 130 copies of the 15.8-kb mitogenome as estimated by Gallaher et al.\(^{33}\)); however, they are higher compared to the estimates provided for the Mamiellphyceae (4–6 copies for cell)\(^{22}\).

**Genes involved in meiosis and production of flagella.** As reported for most of the chlorophytes previously investigated, the meiosis-specific genes encoding SPO11–2, DCM1, Rad51c, MSH4, and MSH5 are present in *Chloropicon* and there is no indication that these genes are pseudogenes. We detected nil or very low levels of expression of these genes under vegetative growth (Supplementary Data 9).

Although vegetative cells of *Chloropicon* lack flagella, they possess an almost full complement of protein genes known to be involved in flagellar structure and maintenance (Supplementary Data 10). Other non-motile chlorophytes such as *Ostreococcus*, *Bathykokoccus*, and *Chlorella* have also retained protein genes of this nature, but unlike *Chloropicon*, they are missing numerous genes involved in intraflagellar transport and radial-spoke formation.
Fig. 3 Variation in GC content and distribution of polymorphic loci and repeats throughout individual Chloropicon chromosomes. The curves shown in blue were obtained by plotting the numbers of SNPs and small indels present in sliding windows of 10,000 nt with a 5000-nt step; dispersed repeats were masked before the identification of these polymorphisms. The orange curves depict the percentages of nucleotides that are part of repeats in the same windows, while the gray curves depict the percentages of nucleotides with guanine and cytosine bases. The locations of indels ≥100 bp are denoted by arrows pointing up (deletions) and down (insertions), and the loci of putative MID genes by magenta diamonds. Chromosomes are labeled with Roman numbers and their lengths are represented to scale on the corresponding x-axis. The scale bar represents 200 kb.
Furthermore, although no scales were distinguished on the three-layer wall of *Chloropicon*, our analyses unveiled a large number of N- and O-glycan modifying enzymes that have been implicated in scale formation based on their specific occurrence and/or enrichment in the scale-bearing prasinophytes *Bathyccoccus* and *Dolichomastix*. Notably, 59 genes encoding sialyltransferases and nine encoding sialidases were identified, the great majority of which are silent or transcribed at very low levels under the growth conditions used (Supplementary Data 4).

**Search for a candidate chromosome carrying the SDR.** The primarily diploid and heterozygous structure of the *Chloropicon* genome suggests that fusion between haploid gametes of different genotypes yielded a diploid individual that reproduced subsequently by mitosis. If gametes of distinct mating types (*mt* and *mt−*) were engaged in this mating as is the case for heterothallic volvocine algae, one would expect the *Chloropicon* SDR (mating-type locus) to be heterozygous in the same way as vegetative diploids of *Chlamydomonas*. The SDR region usually exhibits genomic rearrangements and a high density of repeated sequences. Our genome analyses revealed nine *Chloropicon* genes encoding putative RWP-RK transcription factors, but none was part of a large indel region or a region exceptionally rich in repeats or polymorphisms (Fig. 3), and sequence alignment of motifs 1 and 2 of these proteins with those present in the Mid proteins of volvocine algae revealed no obvious ortholog of Mid (Supplementary Fig. 4).

**Genes for light-harvesting complex proteins.** The chlorophyll- and carotenoid-binding proteins making up the light-harvesting complexes associated with photosystem II (LHCCI) and photosystem I underwent considerable changes during the evolution of the Chloroplastida. The LHCCI proteins (Lhcb) are especially diverse. Besides the invariably conserved minor Lhcb4 (CP29) and Lhcb5 (CP26) proteins, mamiellophycean feature at least five prasinophyte-specific LHCCI proteins (Lhcbp) whose evolutionary origin predates the divergence of streptophytes and chlorophytes. In contrast, all six *Chloropicon* Lhcb proteins are typical of core chlorophytes (Supplementary Data 11).

**Carotenoid biosynthesis genes.** The Chloropiconophyceae contain the basic set of pigments found in the Chlorophyceae (neoxanthin, violaxanthin, lutein, zeaxanthin, antheraxanthin, and β-β carotene) as well as loroxanthin. In addition, like *Haematococcus pluvialis* and *Chromochloris zofingiensis*, most members of the Chloropiconophyceae, including *C. primus*, produce astaxanthin. On the other hand, carotenoids characteristic of the prasino-3 group (prasinoxanthin, dihydrolutein, urilide, micromonal, and micromonol) are found in the Mamiellophyceae.

We identified the complete set of *Chloropicon* genes responsible for carotenoid biosynthesis (Supplementary Data 12 and Supplementary Fig. 5). As in land plant and core chlorophytes, two desaturases (Pds and Zds) and two *cis*-trans isomerases (Z-Isol and CrTso) catalyze the four steps of dehydrogenation that are required for converting 15-*cis* phytoene into all-*cis* lycopene. In addition, there is a gene encoding a bacterial/fungal-type phytoene desaturase (CrtI), an enzyme performing multiple dehydrogenation steps. All members of the Mamiellophyceae are missing a number of genes in the pathway typically found in both chlorophytes and streptophytes.
The proteins known to participate directly in astaxanthin synthesis, namely beta-carotene ketolase (Bkt) and beta-carotene hydroxylase (Chyb), are encoded by single-copy genes in Chloropicon. The order in which they function in H. pluvialis and C. zoofingiensis is still uncertain. Considering that only chyb has been retained in the Mamiellophyceae, we speculate that its protein product is required to synthesize zeaxanthin, a precursor of both violaxanthin and neoxanthin; hence, assuming that the Chloropicon Chyb has the same substrate specificity, we propose that subsequent ketolation of zeaxanthin by Bkt yields astaxanthin (Supplementary Fig. 5).

Thiamine biosynthesis genes. We inferred that Chloropicon has the ability to synthesize thiamine (vitamin B1) (Supplementary Data 13 and Supplementary Fig. 6). Considering that the availability of exogenous vitamins is limited in open oceans, the retention of this biosynthesis pathway may help explain its ubiquitous presence of Chloropicon across the globe. In the form of thiamine diphosphate (TPP), vitamin B1 is an essential cofactor for energy generation and general metabolism that is produced via the coupling of pyrimidine and thiazole precursors, namely hydroxymethylpyrimidine pyrophosphate (HMP-PP) and hydroxyethyl-thiazole phosphate (HET-P) or 5-(2-hydroxyethyl)-4-methyl-1,3-thiazole-2-carboxylic acid (cHET-P). Catalyzed by the bifunctional enzyme TH1, this reaction results in thiamine monophosphate, a precursor that is subsequently converted to TPP by the kinase TPK. The pyrimidine moiety is produced in reactions catalyzed by ThiC and TH1, whereas the thiazole group is synthesized by Thi4.

In Ostreococcus and Micromonas, thiC, thi4 and thi1 are pseudogenes. To survive, these picoeukaryotes are dependent on exogenous vitamin B1 or related micronutrients supplied by B1-synthesizing marine bacteria or eukaryotic phytoplankton. It has been recently shown that the thiazole precursor cHET is a key micronutrient in the Mamiellophyceae and that its utilization is dependent upon the activity of the kinase ThiM. A Chloropicon homolog of this protein could not be detected by BlastP searches.

Catabolism of branched-chain amino acids. In plants, leucine, isoleucine, and valine are known to play a crucial role as alternative sources of energy under carbohydrate starvation or prolonged darkness. (see Supplementary Fig. 7 for the branched-chain amino acids, BCAA, degradation pathways). This auxiliary
supply of energy is generated in the mitochondrion. The Chloropicon genome potentially encodes all the proteins necessary to degrade the three BCAAs and similarly, complete BCAA catabolic pathways were inferred for the trebouxiiophytes Helicosporidium, Auxenochlorella, and Coccomyxa (Supplementary Fig. 7 and Supplementary Data 2). In the case of the four other examined core chlorophytes, only one enzyme (enoyl-CoA hydratase, K07511) in the valine and isoleucine catabolic pathways was not detected using GhostKOALA. In contrast, numerous proteins in the degradation pathways of the three BCAAs were not retrieved in the Mamiellophyceae.

A metabolic pathway unique to Chloropicon. Among the proteins unique to Chloropicon, we uncovered the complete set of enzymes performing the degradation of propionate via the 2-methylcitrate cycle (2-MCC)\(^{17}\), namely ACS, PrpC, PrpD, AcnB, and PrpB (Fig. 6 and Supplementary Data 15). The gene encoding PrpF was also identified, albeit this enzyme is not likely to function in an alternative branch of the 2-MCC due to the absence of the upstream gene product (AcnD). To ensure that the 2-MCC was not previously overlooked in the Chloroplastida, we examined all green algal genomes sequenced to date using GhostKOALA and did not identify PrpC, PrpB, and PrpF. Although PrpD was retrieved in the Ostreococcus and Micromonas genera, the latter protein sequences were highly divergent from that of Chloropicon (E-value ≥ 1.0E-10), an unanticipated result considering the very strong level of similarity (E-value = 0) observed for the Chloropicon and mamiellalean orthologs of the bifunctional AcnB enzyme (aconitate hydratase in the TCA cycle and 2-methylisocitrate dehydratase in the 2-MCC).

Congruent with BlastP searches, phylogenetic trees inferred from the Chloropicon PrpB, PrpC, PrpD, and PrpF proteins identified different bacteria and/or non-photosynthetic euukaryotes as their closest relatives (Supplementary Fig. 8), highlighting complex evolutionary histories for the genes encoding these proteins which likely involved horizontal transfer events from bacteria and multiple losses in euukaryotic lineages. Reminiscent of the gene organization found in bacteria\(^{48}\), prpB, prpD, and prpC are closely linked on chromosome V of the Chloropicon genome (Supplementary Data 2).

Discussion

Like mamiellaean genomes, the Chloropicon genome has a small size and is densely packed with coding sequences, with a minimum of gene redundancy; however, extensive differences are seen at the gene content and gene arrangement levels. Chloropicon exhibits genes/pathways that are present in core chlorophytes but not in mamiellaeans, and conversely, mamiellaeans exhibits genes/pathways that are not present in Chloropicon, suggesting that the two groups of picoalgae adopted different strategies to adapt to their oceanic environments. Key genes belonging to the former category include those encoding pigment-binding proteins of light-harvesting complexes, proteins required for synthesis of carotenoid and thiamine, and enzymes participating in degradation of BCAAs.

The common ancestor of the Chloropiconphycceae and Mamiellophyceae emerged about 950 Mya\(^{49}\) and like the algae belonging to the sister class of the Mamiellophyceae (Pyramimonadophyceae), was presumably a scaly flagellate with a large cell body and a complex internal morphology that enabled a

\[ \text{Propionate} \xrightarrow{\text{ACS (K01895)}} \text{Propionyl-CoA} \]

\[ \text{Propionyl-CoA} \xrightarrow{\text{ECH (K07511)}} \text{Acryloyl-CoA} \]

\[ \text{Acryloyl-CoA} \xrightarrow{\text{HADH (K09605)}} \text{2-Methylcitrate} \]

\[ \text{2-Methylcitrate} \xrightarrow{\text{PrpD (K01720)}} \text{3-Hydroxy-propionyl-CoA} \]

\[ \text{3-Hydroxy-propionyl-CoA} \xrightarrow{\text{AcnB (K01682)}} \text{Acetyl-CoA} \]

\[ \text{Acetyl-CoA} \xrightarrow{\text{HDH (K00020)}} \text{Malonate semialdehyde} \]

\[ \text{Malonate semialdehyde} \xrightarrow{\text{MMSDH (K00140)}} \text{Succinate + Pyruvate} \]

\[ \text{Succinate + Pyruvate} \xrightarrow{\text{PrpB (K03417)}} \text{PrpF (K09788)}} \text{2-Methyl-trans-aconitate} \]

\[ \text{2-Methyltrans-aconitate} \xrightarrow{\text{PrpC (K01659)}} \text{2-Methyl-cis-aconitate} \]

\[ \text{2-Methyl-cis-aconitate} \xrightarrow{\text{AcnB (K01682)}} \text{2-Methylcis-aconitate} \]

\[ \text{2-Methylcis-aconitate} \xrightarrow{\text{PrpD (K01720)}} \text{2-Methylacetate} \]

\[ \text{2-Methylacetate} \xrightarrow{\text{ACS (K01895)}} \text{Propionate} \]

\[ \text{Propionate} \xrightarrow{\text{ACS (K01895)}} \text{Propionyl-CoA} \]

\[ \text{Propionyl-CoA} \xrightarrow{\text{ECH (K07511)}} \text{Acryloyl-CoA} \]

\[ \text{Acryloyl-CoA} \xrightarrow{\text{HADH (K09605)}} \text{2-Methylcitrate} \]

\[ \text{2-Methylcitrate} \xrightarrow{\text{PrpD (K01720)}} \text{3-Hydroxy-propionyl-CoA} \]

\[ \text{3-Hydroxy-propionyl-CoA} \xrightarrow{\text{AcnB (K01682)}} \text{Acetyl-CoA} \]

\[ \text{Acetyl-CoA} \xrightarrow{\text{HDH (K00020)}} \text{Malonate semialdehyde} \]

\[ \text{Malonate semialdehyde} \xrightarrow{\text{MMSDH (K00140)}} \text{Succinate + Pyruvate} \]

\[ \text{Succinate + Pyruvate} \xrightarrow{\text{PrpB (K03417)}} \text{PrpF (K09788)}} \text{2-Methyl-trans-aconitate} \]

\[ \text{2-Methyl-trans-aconitate} \xrightarrow{\text{PrpC (K01659)}} \text{2-Methyl-cis-aconitate} \]

\[ \text{2-Methyl-cis-aconitate} \xrightarrow{\text{AcnB (K01682)}} \text{2-Methylcis-aconitate} \]

\[ \text{2-Methylcis-aconitate} \xrightarrow{\text{PrpD (K01720)}} \text{2-Methylacetate} \]

\[ \text{2-Methylacetate} \xrightarrow{\text{ACS (K01895)}} \text{Propionate} \]

Fig. 6 Chloropicon proteins predicted to participate in the degradation of propionate via the modified β-oxidation and 2-MCC pathways. Supplementary Data 15 report the loci of the encoding genes, the presence/absence of homologs in other green algae and additional information, including expression data and predicted subcellular localization. PrpC, 2-methylcitrate synthase [EC:2.3.3.5]; PrpD, 2-methylcitrate dehydratase [EC:4.2.1.79]; AcnB, aconitate hydratase [EC:4.2.1.99]; PrpB, methylisocitrate lyase [EC:4.1.3.30]; PrpF, 2-methylacetoacetate cis-trans-isomerase [EC:5.3.3.1]; ACOX, acetyl-CoA dehydrogenase [EC:1.3.99.5]; ECH, enoyl-CoA hydratase [EC:4.2.1.17]; HADH, 3-hydroxyisobutryl-CoA hydrolase [EC:3.1.2.4]; HDH, 3-hydroxyisobutyrate dehydrogenase/3-hydroxypropionate dehydrogenase [EC:1.1.1.31]; MMSDH, malonate-semialdehyde dehydrogenase (acetylating)/methylmalonate-semialdehyde dehydrogenase [EC:1.2.1.18 1.2.1.27]; ACS, acetyl-CoA synthetase [EC:6.2.1.1]
phagomixotrophic mode of nutrition. A draft genome of the phagophotrophic *Cymbomonas tetratomaformis* highlighted a considerably greater genome size (850–1200 Mb) and gene repertoire (37,366 gene models) compared to those of *Chloropicon* and the Mamiellales. Our study thus reinforces the view that several independent events of genome minimizations, entailing differential gene losses, occurred during prasinophyte diversification.

Analyses of allelic ratios at polymorphic sites and of chromosome coverage depths revealed that all 20 *Chloropicon* chromosomes, except the trisomic chromosome XVII, are present in two copies. Further supporting this conclusion is the finding that the DNA content previously estimated for *Chloropicon* cells by flow cytometry predicted a genome size of about 45 Mb, which is twice larger than our genome assembly. While diploid strains of *C. reinhardtii* that reproduce asexually can be generated in the laboratory, a predominantly diploid phase of growth (i.e. diploplontic life cycle) has not been observed among natural isolates of unicellular green algae, but it is prevalent in other groups of unicellular algae such as the diatoms. In green algae, diploid vegetative cells have been documented solely for ulvophytes with a haplo-diploplontic life cycle.

Despite its higher metabolic cost, diploidy (i.e. a life cycle with a predominantly diploid growth phase) has been traditionally viewed as evolutionary favored over haploidy. By allowing the masking of nearly all deleterious mutations arising in the genome, diploidy has been proposed to confer higher organismal fitness and a larger pool of mutant alleles that may help adaptation to new environments. However, given that more mutations arise in haploids and that selection is less efficient, theoretical studies have shown that a dominant diploid phase is disadvantageous under certain circumstances, e.g. when the environment is static, and is tolerated as long as deleterious mutations are masked and as long as there is enough sex and recombination in the population. By contrast, newly arising deleterious mutations are more efficiently eliminated in haploids and beneficial mutations spread more easily, increasing fitness at equilibrium. Furthermore, haploid cells may have a nutritional advantage over diploid cells because they are often smaller than haploids and thus have a higher surface area to volume ratio; in the yeast *Saccharomyces cerevisiae*, it has been experimentally shown that haploids are more suitable than diploids for some conditions of nutrient limitations. In this context, it is noteworthy that different isolates of *Chloropicon roscoffensis* were shown to differ in DNA content and that two strains from the low genome size (20 Mb) group also display a 25% smaller average cell diameter compared to strains from the high genome size (42 Mb) group.

Only a few green algae, including the prasinophyte *Nephrospelms olivaceus*, have been observed to engage in sex in nature or laboratory conditions. In line with the idea that sexual reproduction is nearly universal in eukaryotes, genomes and/or population genetics studies algae sequenced have revealed indirect evidence for the capacity of sexual reproduction in most of the green so far, including *Chloropicon*. The conservation of genes related to meiosis/sex and flagellar structure and maintenance (flagella play a role in the recognition of mating partners) suggests that *Chloropicon* had a past history of sex or has retained the ability to produce haploid gametes whose fusion gives rise to new diploid genotypes with purged deleterious alleles. Our data on the expression of sialyltransferase genes also raise the possibility that there exists a life-cycle phase in which cells exhibit sialylated scales or scale-like structures. This phase might be linked to sexual reproduction, given that a sialyltransferase activity believed to be important for cell-cell interaction has been identified on the external surface of gametes in *Chlamydomonas moewusii*.

Assuming that *Chloropicon* is heterothallic (matting occurring between gametes with genetically determined mating-types), one would expect that the SDR would be diploid in the same way as vegetative diploids of *Chlamydomonas* (with one copy of each mt+ and mt− loci) and that, as documented for heterothallic volvocine algae, the dominant sex-determining gene MID gene would be located within a genomic region enriched in indels and SNPs. Our failure to identify a region with these characteristics may be interpreted as indicating that the *Chloropicon* mt+ and mt− loci are short and highly similar to one another. However, such an exceptionally high level of similarity seems unlikely considering our current knowledge about SDR architecture and for this reason, the hypothesis that *Chloropicon* is homothallic (self-fertile) cannot be ruled out. In volvocine lineages, homothallism repeatedly evolved from heterothallic ancestors and for a homothallic species of Volvox, it has been shown that epigenetic control of expression of the dominant MID is probably involved in the formation of male and female gametes within a single clonal culture, although no information is currently available on the genomic region corresponding to the *mt* locus of heterothallic relatives. Aneuploidy is a widespread phenomenon in unicells such as *S. cerevisiae* and the primarily asexual fungus *Candida albicans* (an heterozygous diploid like *C. primus*) that imposes a burden on the cell but provides a fitness advantage under certain growth or stressful conditions. *Chloropicon* might have gained a third copy of chromosome XVII during the period of asexual growth following isolation of the original *Chloropicon* CCMP1205 strain in 1965. Additional *C. primus* isolates will need to be investigated to determine the time of emergence of this trisomy.

The significant deviations from the 1:1 allelic ratio observed on chromosome IV and other chromosomal regions might be the result of biased gene conversion events (LOH events) arising from mitotic recombination, followed by segregation of distinct genotypes during further growth of *Chloropicon* CCMP1205. In asexual diploid eukaryotes, LOH is thought to provide a mechanism for removing deleterious mutations and thus compensate for a lack of sex.

Among the proteins unique to *Chloropicon*, we uncovered the complete set of enzymes performing the degradation of propionyl-CoA via the 2-MCC. Propionate, in the form of propionate, is a toxic metabolite that is produced by several pathways: breakdown of several amino acids (methionine, threonine, and the BCAAs isoleucine and valine), β-oxidation of odd-chain fatty acids, and catabolism of branched fatty acids, such as phytanic acid which is derived from the degradation of chlorophyll. Propionyl-CoA has been shown to interfere with enzymes of the central metabolism and to inhibit polyketide synthesis. In many prokaryotes and some eukaryotes, propionate also serves as a carbon source. Organisms have evolved a variety of mechanisms to convert propionyl-CoA into products that can be further metabolized via the TCA cycle: a propionate carboxylation pathway dependent on vitamin B12 that produces succinyl-CoA, a modified β-oxidation pathway leading to acetyl-CoA and CO₂, and the 2-MCC pathway converting propionate into succinate and pyruvate.

It remains unclear whether the 2-MCC is the only metabolic route available to *Chloropicon* for propionyl-CoA detoxification. We identified candidate genes that potentially encode all the proteins catalyzing the reactions of the modified β-oxidation pathway (Fig. 6 and Supplementary Data 15); however, all are also required for BCAA catabolism (Supplementary Data 14 and Supplementary Fig. 7) and contrary to the model plant *Arabidopsis*, *PredAlgo* and TargetP analyses predicted that the putative enoyl-CoA hydratase (ECH, K07511) responsible for the production of the key metabolite β-hydroxypropionpropionate is located in the mitochondrion instead of the peroxisome.
Considering the scarcity of nutrients available in open oceans compared to coastal waters, the ability to recycle propionate for other metabolic purposes such as a source of carbon and energy could confer a clear benefit to Chloropicon. To our knowledge, the colorless obligate-heterotrophic trebouxiophyte Prototheca zopfi is the sole green alga that has been reported to utilize propionate as carbon source, and in this case, this organic acid appears to be metabolized by the modified β-oxidation pathway. In contrast, no propionate catabolic pathway appears to be present in the Mamielophyceae; given their inability to catabolize BCAAs using the pathways typically found in eukaryotes, these prasinophytes can perhaps dispense with propionyl-CoA detoxification.

As reported for plants, BCAA degradation might contribute to cell fitness in the Chloropicon and core chlorophytes under carbohydrate starvation or prolonged darkness. The apparent absence of the typical BCAA catabolic pathways that are not available to any other eukaryotes, as the Ehrlich pathway used by S. cerevisiae during anaerobic fermentation metabolism, alternatively, mamielophyceans might be less dependent than Chloropicon and core chlorophytes upon BCAA catabolism under nutrient- or light-limiting conditions if they are more efficient at producing reserves of energy in the forms of carbohydrates or lipids or at harvesting energy from alternative metabolites. In this context, it is worth reporting that our genome analyses predict that Chloropicon utilizes a more limited range of intermediates than mamielophyceans during anaerobic fermentation metabolism. The nitrogenase of M. myxa subellipsoidea is the sole green alga that has been reported to utilize propionate as carbon source, and in this case, this organic acid appears to be metabolized by the modified β-oxidation pathway. In contrast, no propionate catabolic pathway appears to be present in the Mamielophyceae; given their inability to catabolize BCAAs using the pathways typically found in eukaryotes, these prasinophytes can perhaps dispense with propionyl-CoA detoxification.

DNA and RNA sequencing. The Chloropicon genome was sequenced using a combination of high-throughput short and long reads generated on the Illumina (San Diego, CA, USA) and Pacific Biosciences (Menlo Park, CA, USA) platforms, respectively. Paired-end (average insert size of 322 bp) and mate-pair (average insert size of 1757 bp) libraries were prepared using the Illumina TruSeq DNA kit and the mate-pair protocol of Bird et al. and MiSeq sequencing on the Illumina HiSEQ platform. Genomes were sequenced using four SMRT cells, P6-C4 chemistry at the University of Michigan DNA Sequencing Core (Ann Arbor, MI), providing 1.3 GB of reads, with polymerserase and subread N50s of 7.99 and 4.89 kb, respectively (Supplementary Fig. 9).

A normalized RNAseq library was prepared by the Plafteforme d’Analyses Génomiques of Université Laval from a mixture containing equal amounts of RNAs isolated from the L2, L8, D2, and D8 synchronized cells and exponential and stationary cultures grown in continuous light. The library was constructed using the Illumina TruSeq RNA kit and the normalization protocol of Zhulidov et al. and sequenced on the Illumina Hiseq 2000 platform (300 cycles) at the McGill University and Genome Quebec Innovation Centre (Montréal, Canada), yielding 22,223,345 paired-end reads.

Genome assembly. Independent de novo assemblies were generated from short and long Illumina and PacBio reads and then merged to produce a high-quality consensus draft using overlap-consensus-approaches followed by manual curation/inspection of discrepancies. Low-quality bases in the Illumina reads were removed with PRINSEQ 0.20.4 and adapter sequences were discarded using FASTX-CLIPPER from the FASTX-Toolkit 0.14.14. Pre-assembly of overlapping paired-end reads was performed with FLASH. Assembly was performed using Ray 2.3.1 and a kmer value of 61 on the paired-end and mate-pair datasets, with reads overlapped with FLASH used as singletons. PacBio assemblies were performed using MHAP 1.5 and HGAP3 as implemented in the Celera assembler v. 8.3RC1 and DNA Assembly Portal v. 2.3.0, for the Celera assembly. FASTQ reads were generated from the bax.h5 files with DEXTRACTOR, converted to Celera format with the fastToCA tool from the Celera package, quality-trimmed by overlap using runCA, and then assembled with MHAP 1.5.

Illumina Ray and PacBio HGAP3 assemblies were merged using the alignment tool for large DNA fragments implemented in Sequencher 5.4.1 (Gene Codes Corporation, Ann Arbor, MI, USA). Discrepancies in the consensus sequence were resolved manually using the editing tool of Sequencher and the basecalling of the Illumina contigs, and when indels were encountered, sequences of the long alleles were integrated into the consensus sequence. For the Celera assembly, FASTQ reads were generated from the bax.h5 files with DEXTRACTOR, converted to Celera format with the fastToCA tool from the Celera package, quality-trimmed by overlap using runCA, and then assembled with MHAP 1.5.

The absence of contaminants in the assembled contigs was confirmed by taxonomized BLAST homology searches against the NCBI nr/taxonomy databases using runTaxonomizedBLAST.pl.

Genome annotations. Ribosomal and transfer RNA genes were identified using RNAmer 1.2 and rRNAscan-SE 1.31, respectively. Protein-coding genes were predicted using the MAKER 2.32 and BRAKER 1.8 pipelines. For BRAKER 1.8, alignments of Illumina RNAs were generated using BLAST against the assembled genome generated in BAM format using HISAT2 2.0.1 and default parameters. To detect protein-coding genes improperly predicted or missed in previous analyses, RNAseq data were mapped independently on the assembled genome with PAS 2.23 using the default mapping seed (111111011111011) with a seed step of 3, the homopolymer flag turned off (-f 1c), an identity of 98%, a filtering length threshold of 30 nucleotides, and a gap mode set to 2. In addition, protein datasets derived from Auxenochlorella protothecoides, Bathycoccus prasinos, Chlorella variabilis, C. reinhardtii, Coccolithus abyssalis, Polytomella geyleri, Helicosporidium sp., Klebsormidium flaccidum, Micromonas pusilla, Micromonas commoda, Ostreococcus tauri, Ostreococcus lucimarinus, and Volvox carteri were curated against the Chlorophyta genomes using TBLASTN searches. Results from gene predictions, RNA mapping and homology searches were loaded in a local Web-Apollo 2.0-2 RC3 server using built-in and custom Perl scripts, and the annotations curated using the sum of all information and Web-Apollo built-in tools.

Curated GFF3 annotations were exported using Web-Apollo built-in tools, converted to EMBL format with WebApolloGF3toEMBL.pl and further curated manually using Artemis 16.0.0, wherein locus tags were automatically added using

**Methods**

**Cell culture and DNA/RNA extraction.** Chloropicon primus strain CCMP 1205 (isolated in 1965 from the Atlantic Ocean) was obtained from the National Center for Marine Algae and Microbiota (East Boothbay, ME, USA) where it is maintained as a cryopreserved, genetically pure culture after it was re-cloned in October 1987. For DNA extraction, cells were grown synchronously in L1 medium at 18°C under 12-hour light/dark cycles and harvested after 2-3 weeks. For RNA extraction, independent cultures were grown in the same medium under synchronized and continuous light conditions; synchronized cells were harvested 2 and 8 h after the onset of the light (L2 and L8) and dark (D2 and D8) cycles, whereas cells grown in continuous light were harvested at the exponential and stationary growth phases. Total DNA and RNA were extracted using the HP Plant DNA Mini and Total RNA II kits (Omega Bio-Tek, Norcross, GA, USA), respectively. RNA integrity was monitored using a Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).
the feature Edt > Automatically Create Gene Names. Amino acid sequences from predicted proteins were exported with EMBLto PROT.pl and putative functions were inferred independently using InterProScan 5 and BLASTP homology searches against the UniProt Swiss-Prot/TrEMBL databases and the Chlamydomonas and Coccomyxa protein datasets. Curated EMBL annotations were converted to TBL format using the curated product list and EMBLto TBL.pl. Final accessions were generated and validated with TBL2ASN.

Gene expression analysis. RNAseq Illumina reads were mapped against the Chloropicon genome with PASS 2.23 using the parameters described in the previous section. The resulting alignments in SAM format were converted to BAM format, then sorted using Samtools 1.3.1. RNAseq sequencing depths for individual nucleotide positions were inferred with the depth -a command line option of Samtools. Gene expression levels were inferred by averaging the sequencing depths of all nucleotides from the exons of the corresponding genes in the Chloropicon genome annotation with genes_expressed.pl.

BUSCO analyses. BUSCO 3 was used to compare the quality and completeness of the DNA assemblies and gene annotations of Chloropicon with those of 12 other green algae. The annotated proteins of these algae were queried against the OrthoDB v10 database, which contains the 2168 near-universal single-copy orthologs found in the Chlorophyta.

Variant identification and GC-content analyses. Variants were identified by mapping genomic reads onto the chromosome assemblies that were masked for repetitive sequences using RepeatModeler 1.0.11 and RepeatMasker 4.0.7 with the -no IS and -nolow options. The Illumina paired-end reads were aligned using minimap2 using the -x sr preset and both SNPs and small indels were called with VarScan2 2.4.3 using a minimum variant frequency (–min-reads2) set to 50, as implemented in get_SNPs.pl. The same methodology was used to calculate the PacBio sequencing depth of the chloroplast and mitochondrial genomes (GenBank accession KJ746601 and MK085998, respectively). Sequencing depths were plotted with Circos using 10,000-nl window and 500-nt slide parameters with GC_content_to_Circos.pl.

Determination of sequencing depth. Illumina reads (paired-ends, mate pairs, paired-ends + mate pairs) and PacBio subreads were aligned independently on the Chloropicon unmasked nuclear genome using minimap2. Sequencing depth of each chromosome was determined from the BAM alignment files with the Samtools depth function as implemented in get_SNPs.pl. The same methodology was used to calculate the PacBio sequencing depth of the chloroplast and mitochondrial genomes (GenBank accession KJ746601 and MK085998, respectively). Sequencing depths were plotted with Circos using 10,000-nt window and 500-nt slide parameters with Coverage_to_Circos.pl.

Comparative analyses of gene content and gene order. Putative orthologs of the Chloropicon gene products in the 13 green algae mentioned in the section Genome annotations were searched by homology searches using BLASTP (E-value cutoff = 1.0E-10) and TBLASTN (E-value cutoff = 1.0E-10). BLASTP and TBLASTN hits were parsed at the desired cutoff using shared_proteins.pl. Maps of metabolic pathway were retrieved from the KEGG online repository with KEGG.sh, and KEGG orthologs (KOs) were assigned for each algal genome with GhostKOALA. The identified KOs were concatenated into distinct KEGG pathway matrices with KO_to_matrices.pl and plotted using the Biocomparator ComplexHeatmap package as implemented in R 3.5.0 with MatrixR_plotter.pl. Gene order data were extracted from the GFF annotation files of the compared algal genomes with gff_to_synteny.pl using regular expressions tailored to account for the variations between GFF files. Orthologous proteins encoded by shared gene pairs were identified by BLASTP searches with get_synteny.pl, iteratively allowing for 0, 1, 5, 10, and 50 genes interspersed between potential pairs. In this analysis, the genes in each potential shared pair had to be in the same relative orientation in the compared genomes, present on a single contig/chromosome in each genome (highly fragmented genomes will inherently produce fewer gene pairs with this assumption), and distant by no more than the number of genes allowed in-between (genes were allowed in-between to account for indels and for inaccurate/spurious predictions resulting in low homology or extraneously predicted genes). Gene colocalization on similar chromosomes or contigs between genomes was investigated with chromosome_explorer.pl and the output was converted to matrices with get_matrix.pl. Gene colocalization data were plotted using the Biocomparator ComplexHeatmap package as implemented in R 3.4.0 with script contatenate_matrices.pl.

Phylogenetic analyses. To determine the phylogenetic relationships of Chloropicon with the 12 algae selected for comparative genome analyses, a maximum likelihood tree was inferred from the proteins encoded by the orthologous single-copy genes present in all species. These genes were identified using Orthofinder 2.3.1; individual protein products were aligned with MAFFT 7.407 using the L-INS-i method; unambiguously aligned regions of each alignment were removed using BMGE 1.12 and default parameters; and finally a supermatrix was created with create_supermatrix.pl. The resulting dataset was analyzed using IQ-TREE 1.6.1 and the GTR+2 + R4 model. Separate phylogenies were also inferred from the Chloropicon PrPβ, PrPc, PrPd and PrPf proteins and their orthologs from the NCBI nr database. Protein alignments were carried out with MAFFT 7.313 and the–auto option, and after filtration of each alignment using TrimAl 1.3 and the block = 6, gt = 0.7, st = 0.005 and sw = 3 options, trees were inferred with IQ-TREE 1.6.1 using the automatically selected best-fit model for all trees. For trees, confidence of branch points was estimated by ultrafast bootstrap analysis with 1000 replicates.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data supporting the findings of this work are available within the paper and its Supplementary Information files. A reporting summary for this Article is available as a Supplementary Information file. The datasets generated and analyzed during the current study are available from the corresponding author upon request. The annotated Chloropicon genome sequence has been deposited in the NCBI Genbank database (accessions CP031034-CP031053). All raw DNA sequencing (SRR8185492-SRR8185497) and RNAseq (SRR8992761) data have been submitted to the NCBI Sequence Read Archive. All these data are accessible under the NCBI Bioproject PRJNA316521 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA316521). The source data underlying Figs. 1d, 4a–d and 5a–c are provided as a Source Data file.

Code availability

All custom Perl scripts used in this study are available on the Pombert Lab GitHub page [https://github.com/PombertLab].

Received: 19 September 2018 Accepted: 16 August 2019
Published online: 06 September 2019

References

1. Leliaert, F. et al. Phylogeny and molecular evolution of the green algae. Crit. Rev. Plant Sci. 31, 1–46 (2012).
2. Leliaert, F. et al. Chloroplast phylogenomic analyses reveal the deepest-branched lineage of the Chlorophyta, Pal mothophyceae class. nov. Sci. Rep. 6, 23567 (2016).
3. Lemieux, C., Otis, C. & Turmel, M. A clade uniting the green algae Mesostigma viride and Chlorokybus aphthicosus represents the deepest branch of the Streptophyta in chloroplast genome-based phylogenies. BMC Biol. 5, 2 (2007).
4. Delure, E. et al. Genome analysis of the smallest free-living eukaryote Ostreococcus tauri unveils many unique features. Proc. Natl Acad. Sci. USA 103, 11647–11652 (2006).
5. Moreau, H. et al. Gene functionalities and genome structure in Bathycoccus prasinus reflect cellular specializations at the base of the green lineage. Genome Biol. 13, R74 (2012).
6. Worden, A. Z. et al. Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes Micromonas. Science 324, 268–272 (2009).
7. Palenik, B. et al. The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation. Proc. Natl Acad. Sci. USA 104, 7789–7790 (2007).
8. Hanschen, E. R. et al. The Gonium pectorale genome demonstrates co-option of cell cycle regulation during the evolution of multicellularity. Nat. Commun. 7, 11370 (2016).
9. Merchant, S. S. et al. The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science 318, 245–250 (2007).
10. Prochnik, S. E. et al. Genomic analysis of organismal complexity in the multicellular green alga Volvox carteri. Science 329, 223–226 (2010).
11. Featherston, J. et al. The 4-Celled Marrella tetrabasmatida supermatrix. Mol. Biol. Evol. 35, 855–870 (2018).
12. Blanc, G. et al. The genome of the polar eukaryotic microalga Coccomyxa subellipsoidea reveals traits of cold adaptation. Genome Biol. 13, R39 (2012).
13. Blanc, G. et al. The Chlorella variabilis NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. Plant Cell 22, 2943–2955 (2010).
14. Hirooka, S. et al. Acidophilic green algal genome provides insights into adaptation to an acidic environment. Proc. Natl Acad. Sci. USA 114, E2304–E2313 (2017).
15. Horii, K. et al. Klebsielliformium flaccum genus reveals primary factors for plant terrestrial adaptation. Nat. Commun. 5, 3978 (2014).
16. Bogen, C. et al. Reconstruction of the lipid metabolism for the microalga Monoraphidium neglectum from its genome sequence reveals characteristics suitable for biooil production. BMC Genomics 14, 926 (2013).
17. Roth, M. S. et al. Chromosome-level genome assembly and transcriptome of the green alga Chromochloris zofingiensis illuminates astaxanthin production. Proc. Natl Acad. Sci. USA 114, E4296–E4305 (2017).
18. Arriola, M. B. et al. Genome sequences of Microcystis aeruginosa and M. wesenbergii. Science 356, 920 (2017).
19. Hori, K. et al. The reason for high photosynthesis in the green alga Prasinocladus: new perspectives and prospects. New Phytol. 216, 62–68 (2017).
20. Liu, J. et al. Utilization of cane molasses towards cost-saving astaxanthin production by a Chlorella zofingiensis mutant. J. Appl. Phycol. 25, 1447–1456 (2013).
21. Jeffrey, S. W., Wright, S. W. & Zapata, M. Microalgal classes and their signature pigments. In Phytoplankton Pigments: Characterization, Chemotaxonomy and Applications in Oceanography (eds Llewellyn, C. A., Egeland, E. S., Johnsen, G. & Roy, S.) (Cambridge Univ. Press, 2011).
22. Helliwell, K. E. The role of B vitamins in phytoplankton nutrition: new perspectives and prospects. New Phytol. 216, 62–68 (2017).
23. Paerl, R. W. et al. Carboxylthiazole is a key microbial nutrient currency and critical component of thiamin biosynthesis. Sci. Rep. 8, 5940 (2018).
24. Helliwell, K. E. et al. Use of plant-derived vitamin B1 precursors, especially thiazole-related precursor, by key marine picoeukaryotic phytoplankton. Proc. Natl Acad. Sci. USA 114, 753–765 (2016).
25. Peng, C., Uygun, S., Shiu, S.-H. & Lot, R. L. The impact of the branch-chain ketocarboxylic dehydrogenase complex on amino acid homeostasis in Arabidopsis. Plant Physiol. 169, 1807–1820 (2015).
26. Watson, E. et al. Metabolic network rewiring of propionate flux compensates vitamin B12 deficiency in C. elegans. elife 5, e17670 (2016).
27. Dolan, S. K. et al. Loving the poisoning: the methylthioimidazole cycle and bacterial pathogenesis. Microbiol. 164, 251–259 (2018).
28. Leliaert, F. et al. Chloroplast phylogenomic analyses reveal the deepest-branching lineage of the Chlorophyta, Palnophytaclade. class. Nov. Sci. Rep. 6, 23567 (2016).
29. Helliwell, K. E., Paasch, A., Narechana, A. & Kin, E. Comparative genomics of a bacteriovorous green alga reveals evolutionary causalities and consequences of phago-mixo-trophic mode of nutrition. Genome Biol. Evol. 7, 3047–3061 (2015).
30. von Dassow, P. & Montresor, M. Unveiling the mysteries of phytoplankton life cycles: patterns and opportunities behind complexity. Plankton Res. 33, 312–312 (2011).
31. Otto, S. P. & Gerstein, A. C. The evolution of haploidy and diploidy. Curr. Biol. 18, R1121–R1124 (2008).
32. Mable, B. K. & Otto, S. P. The evolution of life cycles with haploid and diploid phases. Bioessays 20, 453–462 (1998).
33. Suda, S., Watanabe Makoto, M. & Inouye, I. Evidence for sexual reproduction in the primitive green alga Nephroselmis olivacea (Prasinoxiphiae). J. Phycol. 5, 596–600 (1989).
34. Grimsley, N., Pequin, B., Bachy, C., Moreau, H. & Pignane, G. Cryptic sex in the smallest eukaryotic marine green alga. Mol. Biol. Evol. 27, 47–54 (2010).
35. Colombo, L. F., Bosmann, H. B. & McLean, R. J. Cell surface localization of the sialyltransferase ectoenzyme system during the Chlamydomonas mating reaction. Exp. Cell Res. 112, 25–30 (1978).
36. Hanschen, E. R., Herron, M. D., Wiens, J. J., Nozaki, H. & Michod, R. E. Repeated evolution and reversibility of self-fertilization in the volvocine green algae. Evolution 72, 386–398 (2018).
37. Yamamoto, K. et al. Molecular evolutionary analysis of a gender-limited MBD ortholog from the homoeosporous species Volvox africansus with male and monocerosoid spheroids. PLoS ONE 12, e1003317 (2017).
38. Yona, A. H. et al. Chromosomal duplication is a transient evolutionary solution to stress. Proc. Natl Acad. Sci. USA 109, 21010–21015 (2012).
39. Ene, I. V. et al. Global analysis of mutations driving microevolution of a heterozygous diploid fungal pathogen. Proc. Natl Acad. Sci. USA 115, E6968–E6967 (2018).
40. Dale, A. L. et al. Mitotic recombination and rapid genome evolution in the invasive forest pathogen Phytophthora ramorum. mBio 10, e02452–e02418 (2019).
41. Magwene, P. M. et al. Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in Suickaryocymes cerasi.a. Proc. Natl Acad. Sci. USA 108, 1987–1992 (2011).
42. Mandegar, M. A. & Otto, S. P. Mitotic recombination counteracts the benefits of genetic segregation. Proc. Biol. Sci. 274, 1301–1307 (2007).
43. Otzen, C., Bardl, B., Jacobsen, I. D., Nett, M. & Brock, M. Candida albicans utilizes a modified beta-oxidation pathway for the degradation of toxic propionyl-CoA. J. Biol. Chem. 289, 8151–8169 (2014).
44. Zhang, Y. Q. & Keller, N. P. Blockage of methylcarboxylic cycle inhibits polyketide production in Aspergillus nidulans. Mol. Microbiol. 52, 541–550 (2004).
45. Helliwell, K. E., Wheeler, G. L., Leptos, K. C., Goldstein, R. E. & Smith, A. G. Insights into the evolution of vitamin B12 autotrophy from sequenced algal genomes. Mol. Biol. Evol. 28, 2921–2933 (2011).
46. Lucas, K. A., Filley, J. R., Erh, J. M., Graybill, G. R. & Hawes, J. W. Peroxisomal metabolism of propionic acid and isobutyric acid in plants. J. Biol. Chem. 282, 2498–24989 (2007).
47. Schertl, P., Danne, L. & Braun, H.-P. 3-Hydroxyisobutyrate dehydrogenase is involved in both, valine and isoleucine degradation in Arabidopsis thaliana. Plant Physiol. 175, 51–61 (2017).
48. Callely, A. G. & Lloyd, D. The metabolism of propionate in the colourless alga Prototheca zopfii. Biochem. J. 92, 338–345 (1964).
Acknowledgements

We thank Robert R. Butler III (Illinois Institute of Technology) and Christina McHenry (University of Michigan) for help with PacBio sequencing. We are also grateful to Brian Boyle (Université Laval) for his assistance in planning the methodological aspects of this sequencing project and to Wenche Eikrem and Thibaut Pollina (University of Oslo) for kindly providing an electron microscopic image of Chloropicon primus. This study was supported by grants from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2007-2830 to M.T. and C.L. and RGPIN-2017-04506 to C.L.) and by startup funds from the Illinois Institute of Technology to J.-F.P.

Author contributions

C.L. and M.T. conceived the study. C.L., C.O., and J.-F.P. produced the genome assemblies, and J.-F.P. carried out the genome annotations. J.-F.P. and C.L. performed the bioinformatics analyses. C.L. and M.T. analyzed the genomic data. M.T., J.-F.P., and C.L. wrote the manuscript and generated the figures. All authors read and approved the final manuscript.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-12014-x.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Peer review information: Nature Communications thanks the anonymous reviewers for their contribution to the peer review of this work.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019