PNAS

Supplementary Information for

Life cycle and functional genomics of the unicellular red alga *Galdieria* for elucidating algal and plant evolution and industrial use

Shunsuke Hirooka*, Takeshi Itabashi, Takako M. Ichinose, Ryo Onuma, Takayuki Fujiwara, Shota Yamashita, Lin Wei Jong, Reiko Tomita, Atsuko H. Iwane and Shin-ya Miyagishima*

*Shunsuke Hirooka or Shin-ya Miyagishima
Email: shirooka@nig.ac.jp or smiyagis@nig.ac.jp

This PDF file includes:

SI Material and Methods
Figures S1 to S7
Table S1
Legends for Movies S1 to S7
Legends for Datasets S1 to S2
SI References 1-30

Other supplementary materials for this manuscript include the following:

Movies S1 to S7
Datasets S1 to S2

SI Material and Methods

Flow cytometric analysis and microscopic observation of nuclear DNA

To compare nuclear DNA content between diploid and haploid of *G. partita*, the original diploid, homozygous diploids obtained from N1, and haploid clone N1 were inoculated into 20-mL MA medium at pH 2.0 and pH 1.0, respectively. The cells were cultured statically in 25-cm^2 tissue culture flasks in a CO_2 incubator in the presence of light (50 µmol photons m^{-2} s^{-1}) until OD_{750} reached around 1.0. The cells were harvested by centrifugation at 1,500 × *g* for 5 min and the cell pellet was resuspended in 1-mL ethanol and acetic acid (3: 1) and fixed for 5 min. Then the fixed cells were centrifuged as above and the cell pellet was resuspended in 1 mL of 1 M HCl and incubated at 60˚C for 10 min. The cells were harvested by centrifugation and resuspended in 0.5 mL of 100 mM MES, pH 6.0 supplemented with 0.001% (v/v) Tween 20. The cells were harvested by centrifugation and resuspended in 0.5 mL of 100 mM MES, pH 6.0 supplemented with 0.001% (v/v) Tween 20 and 1 µM SYTOX Green nucleic acid stain (Invitrogen). After 10-min incubation at room temperature, the cells were subjected to flow cytometry (Attune NxT Flow Cytometer, Thermo Fisher Scientific) and observed using fluorescence microscopy (BX-51 and the filter set NIBA, Olympus).

DNA extraction from *G. partita*

For PacBio long-read sequencing, high-molecular-weight genomic DNA was extracted from the haploid clone N1 as follows. The haploid cells were harvested by centrifugation at 3,000 × *g* for 10 min from 160-mL culture (OD_{750} was ~2.0). The cell pellet was suspended in 10-mL SDS-EB buffer [0.5% (w/v) sodium dodecyl sulfate (SDS), 5 mM EDTA, 50 mM Tris-HCl, pH 8.0] supplemented with 25-µg/ml proteinase K (FUJIFILM Wako Pure Chemical Corporation, Japan), and incubated at 50°C for 60 min. Total nucleic acids were extracted from the cell lysate with PCI (phenol: chloroform: isoamyl alcohol = 25∶24∶1) as follows. The cell lysate was mixed with an equal volume of PCI, inverted vigorously ten times, and centrifuged at 10,000 × *g* for five min. The upper aqueous phase was carefully obtained, and nucleic acids were harvested through precipitation with the same volume of isopropanol and 0.1 volume of 3-M sodium acetate. The pellet was dissolved in 1-mL TE (1 mM EDTA, 10 mM Tris-HCl, pH 8.0) supplemented with 0.2-mg/ml RNase A (PureLink RNase A; Invitrogen) and incubated at 37°C for 45 min. The resultant genomic DNA was then purified through PCI extraction, followed by isopropanol precipitation as above, and further purified using Genomic-tip 20/G (QIAGEN) according to the manufacturer's instructions. According to the manufacturer's instructions, genomic DNA was subjected to library construction with 20-kb SMRTbell TPK (Pacific Biosciences of California, Inc., USA) and sequencing using PacBio Sequel.

For Illumina short-read sequencing, genomic DNA was extracted from the original diploid clone, homozygous diploid obtained from N1, heterozygous diploids by crossing two different haploid clones, and haploid clones N1, N2, N3, N4, and, N5 as follows. The cells were harvested by centrifugation at 3,000 × *g* for 10 min from 20-mL respective cultures (OD_{750} was ~1.0), and the cell pellet was suspended in 700-µl SDS-EB buffer. Then, 700-µL PCI and glass beads (acid-washed 425−600-µm; Sigma-Aldrich) were added to the sample and vortexed for ten min at 4°C. Total nucleic acids were extracted from the cell lysate with PCI and precipitated with one volume of isopropanol and 0.1 volume of 3-M sodium acetate as described above. The pellet was dissolved in 1-mL TE supplemented with 0.2-mg/mL RNase A (PureLink RNase A, Invitrogen) and incubated at 37°C for 45 min. The genomic DNA was purified through PCI extraction, followed by isopropanol precipitation.

Genome analyses of *G. partita*

The genomic DNAs from the original diploid clone and haploid clones N1, N2, N3, N4, and N5 were separately subjected to library construction using the TruSeq DNA PCR Free (Illumina, Inc.) and 150-bp paired-end sequencing using NovaSeq 6000. The reads were cleaned up using the Cutadapt ver. 3.1 (1) by trimming low-quality ends (<QV30) and adapter sequences and by removing reads shorter than 50 bp. The trimmed reads were used for subsequent analyses.

For *de novo* assembly of organelle genomes, the Illumina short reads of the haploid clone N1 were assembled using SPAdes ver. 3.15.0 (2). The genes of the chloroplast and

mitochondrial genomes were annotated using GeSeq on Chlorobox (3) with the previously annotated chloroplast and mitochondrial genomes of *G. sulphuraria* 074W (4) as a reference. For the *de novo* nuclear genome assembly, the PacBio long reads of the haploid clone N1 were mapped to the chloroplast and mitochondrial genomes using minimap2 ver. 2.17 (5) and the mapped reads were removed. Then the remaining long reads were assembled *de novo* using Canu ver. 2.0 (6) with the following parameters: genomeSize = 20m, minReadLength = 10,000, minOverlapLength = 5,000. Some Canu output contigs were broken at duplicated regions and connected manually based on BLASTN results. The connected contigs were polished using Racon ver. 1.4.17 (7), followed by SNP and indel error correction using Pilon ver. 1.2.3 (8) with Illumina and PacBio reads that were mapped using Bowtie2 ver. 2.3.4.1 (9) and Minimap2 ver. 2.17 (5), respectively. Some errors, which were not automatically corrected, were further fixed manually. Using Augustus software (10), nuclear genes were predicted by integrating RNA-seq data of the haploid clone N1 and the homozygous 2N obtained from N1 (described later) with additional manual curation. As a result, 7,832 genes were predicted in the nuclear genome of the haploid clone N1. Each translated amino acid sequence was used as a BLASTP query against the Swiss-Prot database (E-value < 1×10^{-3}) through Blast2GO (11) to assign a provisional annotation (Dataset S2). The nuclear genes encoding transcription factors, secretary proteins, and glycosyltransferases were identified using Plant Transcription Factor Database 5.0 (12), SignalP 5.0 (13), and dbCAN2 metaserver (14), respectively.

Colinear blocks of gene duplicates in the nuclear genome were detected using MCScanX (15) inserted in TBtools (16). The duplicated genes were classified as segmental/WGD, tandem, proximal, or dispersed duplicates, based on all-to-all BLASTP (E-value < 1×10^{-10}, top five matches). The sequencing coverage depth was estimated as described in (17) with a window size of 1,000 bp. Genome-wide co-linearity (segmental/WGD label) and coverage depth were plotted as circos graphics using Clico FS (18).

To identify SNPs/indels, the DNA-seq reads were mapped to the *G. partita* reference genome sequence (haploid clone N1) using Bowtie2 ver. 2.3.4.1. After removing reads mapped at multiple sites using SAMtools ver. 1.8, SNPs/indels were called with GATK HaplotypeCaller ver. 3.8 (19). High-confidence SNPs/indels were identified using GATK VariantFiltration ver. 3.8 with the parameters DP > 30 and QUAL > 50. SNPs/indels were visualized using the Integrative Genomics Viewer (IGV) ver. 2.8.2 (20).

RNA extraction from *G. partita*

G. partita haploid clones N1 and N2 and the homozygous diploids obtained from N1 or N2, respectively, were inoculated into 40-mL MA medium at pH 1.0 to give an OD_{750} of 0.2. Then, the cells were cultured in 100-mL conical flasks on a rotary shaker (150 rpm) at 40°C in the presence of light (100 µmol photons m^{-2} s^{-1}) for three days. The growing cells were harvested through centrifugation at 2,000 × *g* for five min, frozen in liquid nitrogen, and stored at −80°C until use. The frozen cells were suspended in 700-µL SDS-EB buffer. Then, 700-µL PCI and glass beads (acid-washed 425−600-µm; Sigma-Aldrich) were added to the sample and vortexed for ten min at 4°C. Total nucleic acids were extracted from the cell lysate with PCI and precipitated with 1 volume of isopropanol and 0.1 volume of 3-M sodium acetate as described above. The pellet was dissolved in DNase buffer [0.8-U/µL RNase Inhibitor (Takara Bio Inc.), 5-mM DTT, 8-mM $MgCl_2$, 40-mM Tris-HCl, pH7.5] supplemented with 0.4-U/µL DNase (Recombinant DNase I; Takara Bio Inc.) and incubated at 37°C for 45 min. The resulting total RNA was purified using PCI, followed by isopropanol precipitation.

Transcriptome analyses of *G. partita*

The RNA samples were subjected to library construction with NEBNext Ultra RNA LP Kit (New England Biolabs) and 150-bp paired-end sequencing by NovaSeq 6000. The reads were cleaned up using Cutadapt ver. 3.1 by trimming low-quality ends (<QV30) and adapter sequences and discarding reads shorter than 50 bp. The trimmed reads were used for subsequent analyses.

One side of the RNA-seq reads was mapped to protein-coding genes using Bowtie2 ver. 2.3.4.1. The count data of respective protein-coding genes were calculated using SAMtools ver. 1.8 (21), BEDtools ver. 2.17.0 (22), and R ver. 3.2.3 (23). The count data were then normalized between the diploid and haploid cultures, and differentially expressed genes (DEGs) were

identified using edgeR ver. 3.12.1 (24), with the following criteria: FDR < 0.01, logCPM>2, logFC > 2 or < –2. Transcripts per millions (TPM) values of respective protein-coding genes were calculated according to ref. (25).

Preparation of linear DNA for the transformation of *G. partita*

The sequences and primers used in this study are listed in Dataset S2. PCR amplification and assembly of DNA sequences were performed using KOD One PCR Master Mix (TOYOBO CO., LTD., Japan) and the In-Fusion Cloning Kit (Takara Bio Inc., Japan), respectively.

To integrate an *mVenus*-fused transgene and *BSD* selectable markers to the intergenic region between g7631.t1 and g7632.t1 loci (IG1) through homologous recombination (Fig. 2A), plasmid pIG1-mVenus-BSD was prepared as follows. Firstly, plasmid pBSD encoding a *BSD* expressing cassette was prepared as follows; *BSD orf* of *Aspergillus terreus* (UniProtKB/Swiss-Prot ID of the amino acid sequence, P0C2P0.1; the nucleotide sequence was codon-optimized to *G. partita* nuclear genome; Dataset S2) was commercially synthesized and amplified by PCR using the primer set Nos. 1/2. *APCC* promoter (*pAPCC*; 500-bp of the upstream flanking sequence of *APCC orf*) and *UBIQUITIN* terminator (*tUBQ*; 300-bp of the downstream flanking sequence of *UBQ orf*) were amplified by PCR using the primer sets Nos. 3/4 and Nos. 5/6, respectively, using *G. partita* haploid N1 genomic DNA as a template. *pAPCC*, *BSD orf*, and *tUBQ* were cloned into pUC19 plasmid (Takara Bio Inc., Japan) by In-Fusion Cloning Kit. Next, plasmid pmVenus-BSD encoding a *mVenus* and *BSD* expressing cassettes was prepared as follows. An *mVenus orf* (the nucleotide sequence was codon-optimized to the *G. partita* nuclear genome; Dataset S2) was commercially synthesized and amplified by PCR using the primer set Nos. 7/8. *ELONGATION FACTOR 1α* promoter (*pEF1α*; 500-bp of the upstream flanking sequence of *EF1α orf*) and *β-TUBULIN* terminator (*tTUBB*; 300-bp of the downstream flanking sequence of *TUBB orf*) were amplified by PCR using the primer sets Nos. 9/10 and Nos. 11/12, respectively, using *G. partita* haploid N1 genomic DNA as a template. *pEF1α*, *mVenus orf*, and *tTUBB* were inserted into pBSD, which was amplified by PCR using the primer set Nos. 13/14, by In-Fusion Cloning Kit. Finally, plasmid pIG1-mVenus-BSD was prepared as follows. The genomic IG1 region was amplified by PCR using the primer set Nos. 15/16 using *G. partita* haploid N1 genomic DNA as a template and cloned into pUC19 plasmid by In-Fusion Cloning Kit, generating pIG1. *mVenus-BSD* was amplified by PCR with the primer set Nos. 17/18 using pmVenus-BSD as a template. The amplified *mVenus-BSD* was inserted into pIG1, which was amplified by PCR with the primer set Nos. 19/20, by In-Fusion Cloning Kit.

To generate *G. partita* transformants expressing mVenus that is targeted into respective organelles (*SI Appendix,* Fig. S3) or labels F-actin (Fig. 5H), the following plasmids were prepared. To generate plasmids pIG1-mTP-mVenus-BSD (targeted into the mitochondrion), pIG1-ERD2-mVenus-BSD (targeted to the Golgi apparatus), and pIG1-calnexin-mVenus-BSD (targeted to the ER), a nucleotide sequence encoding the N terminal mitochondrial transit peptide of elongation factor thermo unstable (EF-Tu), *ERD2 orf* and *CALNEXIN orf* were amplified by PCR using the primer sets Nos. 21/22, Nos. 23/24, and Nos. 25/26, respectively, using *G. partita* haploid N1 cDNA as a template. Respective PCR products were inserted into pIG1-mVenus-BSD, which was amplified by PCR using the primer set Nos. 27/10. To generate plasmids pIG1-NLS-mVenus-NLS-BSD (targeted into the nucleus), pIG1-mVenus-PTS1-BSD (targeted into the peroxisome), and pIG1-Lifeact-mVenus-BSD (binds F-actin), the sequence of nuclear localization signal (NLS)-mVenus-NLS (NLS-encoding sequence was attached to the primers), mVenus-peroxisome targeting signal 1 (PTS1) (PTS1-encoding sequence was attached to the primer), and Lifeact-mVenus (Lifeact and a four-amino-acid-linker-encoding sequence was attached to the primer) were amplified by PCR using the primer sets Nos. 28/29, Nos. 30/31, and Nos. 32/33, respectively, using the commercially synthesized *mVenus orf* as a template. Respective PCR products were inserted into pIG1-mVenus-BSD, which was amplified by PCR using the primer set Nos. 34/10, respectively.

To generate a *G. partita* transformant in which the *BSD* selectable marker and the herpes simplex virus thymidine kinase (*HSVtk*) suicide marker are integrated into the IG1 locus (Fig. 3), plasmid pIG1-tUBQ-HSVtk-BSD-tUBQ was prepared as follows. First, the *HSVtk orf* (UniProtKB/Swiss-Prot ID of the amino acid sequence, P13157; the nucleotide sequence was codon-optimized to *G. partita* nuclear genome; Dataset S2) was commercially synthesized and

amplified by PCR using the primer set Nos. 35/36. The resultant PCR product was inserted into pIG1-mVenus-BSD, which was amplified by PCR using the primer set Nos. 34/10, by In-Fusion Cloning Kit, so that *mVenus orf* was replaced with *HSVtk orf*. The resultant plasmid was amplified with the primer sets Nos. 37/38, and then *tUBQ,* which was amplified by PCR using *G. partita* haploid N1 genomic DNA as a template using the primer set Nos. 5/18, was inserted at the upstream of the *HSVtk* expressing cassette.

To generate a uracil-auxotrophic (Δ*URA1*) BS-resistant *G. partita* transformant (Fig. 4), plasmid pΔURA1-mVenus-BSD was prepared as follows. First, *URA1 orf* flanked with the 1,000-bp upstream and 1,000-bp downstream genomic sequences was amplified by PCR using the primer set Nos. 39/40 using *G. partita* haploid N1 genomic DNA as a template and cloned into pUC19 plasmid by In-Fusion Cloning Kit, generating pURA1. *mVenus* expression cassette and *BSD* selectable marker were amplified by PCR using the primer set Nos. 17/18 using pIG1-mVenus-BSD as a template and inserted into pURA1, which was amplified with the primer set Nos. 41/42, so that *mVenus* and *BSD* were sandwiched between the upstream and downstream sequences of *URA1 orf*.

To knockout *BELL, KNOX, MADS, ACT1*, *ACT2*, *ACT3*, *ACT4*, *MYOSIN* (*MYO*), and *PSY* in *G. partita* (*SI Appendix,* Fig. S5D), plasmids pΔBELL-BSD, pΔKNOX-BSD, pΔMADS-BSD, pΔACT1-BSD, pΔACT2-BSD, pΔACT3-BSD, pΔACT4-BSD, pΔMYO-BSD, and pΔPSY-BSD were prepared as follows. *BELL, KNOX, MADS, ACT1*, *ACT2*, *ACT3*, *ACT4*, *MYO,* and *PSY orf* flanked with the 1,000-bp upstream and 1,000-bp downstream genomic sequences were amplified by PCR using the primer sets Nos. 43/44, Nos. 47/48, Nos. 51/52, Nos. 55/56, Nos. 59/60, Nos. 63/64, Nos. 67/68, Nos. 71/72, and Nos. 75/76, respectively, using *G. partita* haploid N1 genomic DNA as a template and cloned into pUC19 plasmid using In-Fusion Cloning Kit, generating pBELL, pKNOX, pMADS, pACT1, pACT2, pACT3, pACT4, pMYO, and pPSY. *BSD* selectable marker was amplified using the primer sets Nos. 13/18 by PCR using pIG1-mVenus-BSD as a template and inserted into pBELL, pKNOX, pMADS, pACT1, pACT2, pACT3, pACT4, pMYO, or pPSY, which was amplified with the primer sets Nos. 45/46, Nos. 49/50, Nos. 53/54, Nos. 57/58, Nos. 61/62, Nos. 65/66, Nos. 69/70, Nos. 73/74, or Nos. 77/78, so that *BSD* was sandwiched between the upstream and downstream sequences of *BELL, KNOX, MADS, ACT1*, *ACT2*, *ACT3*, *ACT4, MYO,* or *PSY orf*, respectively.

To generate a *CHLD* and *PSY* double knockout *G. partita* line (*SI Appendix,* Fig. S7A-C), *CHLD* was disrupted first using pΔCHLD-HSVtk-BSD, and then *PSY* was disrupted using pΔPSY-BSD. pΔCHLD-HSVtk-BSD was prepared as follows. *CHLD orf* flanked with the 1,000-bp upstream and 1,000-bp downstream genomic sequences was amplified by PCR using the primer sets Nos. 79/80 using *G. partita* haploid N1 genomic DNA as a template and cloned into pUC19 plasmid using In-Fusion Cloning Kit, generating pCHLD. *HSVtk* (*pEF1a-HSVtk-tTUBB*)-*BSD* (*pAPCC-BSD-tUBQ*) was amplified by PCR using the primer sets Nos. 17/18 using pIG1-tUBQ-HSVtk-BSD-tUBQ as a template. *tUBQ* was amplified by PCR using the primer sets Nos. 5/81 using *G. partita* haploid N1 genomic DNA as a template. The amplified *HSVtk-BSD* and *tUBQ* were inserted into pCHLD, which was amplified by PCR using the primer sets Nos. 82/83, using In-Fusion Cloning Kit, so that *tUBQ-HSVtk-BSD-tUBQ* was sandwiched between the upstream and downstream sequences of *CHLD orf* (*SI Appendix,* Fig. S7A).

The linear DNA used for transformation of *G. partita* haploid clone N1 was prepared by PCR using the primer set Nos. 84/85 (designed based on the sequence of pUC19) using respective plasmids as templates, and purified using the QIAquick PCR Purification Kit (QIAGEN). The occurrence of the recombination event in the respective transformant was confirmed by PCR using the genomic DNA as a template with the primer sets indicated in Dataset S2.

Immunoblotting

Haploid cells of the wild-type clone N1 and its mTP-mVENUS expressor were cultivated in 20-mL MA medium at pH 1.0 in 25-cm^2 tissue culture flasks statically in a CO_2 incubator at 42°C under light conditions (50 μmol photons m^{-2} s^{-1}). Cells were harvested using centrifugation at 1,500 × *g* for five min at 4°C. The cell pellets were lysed with the sample buffer (2% SDS, 62-mM Tris-HCl, pH 6.8, 100-mM DTT, 10% glycerol, and 0.01% Bromophenol blue) and then incubated at 96°C for ten min. The cell lysate was centrifuged at 15,000 × *g* for ten min, and the supernatant fraction

was harvested. The protein concentration was determined using XL-Bradford (Integrale, Japan). 5-µg total cellular protein was separated in each lane on 15% SDS-polyacrylamide gels and transferred onto polyvinylidene difluoride membranes (Immobilon; Millipore). The membrane blocking, antibody reactions, and signal detection were conducted as previously described (26). The anti-GFP antibody (JL-8, Takara Bio Inc.; at a concentration of 1-µg/ml) was used to detect mVENUS protein.

Cultivation and pigment analyses of Δ*CHLD*, Δ*PSY,* and Δ*CHLD* Δ*PSY*

Haploid cells of the wild-type clone N1, Δ*CHLD*, Δ*PSY,* and Δ*CHLD* Δ*PSY* were cultivated in 50-mL MA medium supplemented with 100-mM glucose in 200-mL Erlenmeyer flasks on a rotary shaker (140-rpm) in an incubator at 42°C in the dark. The absorption spectrum of each culture was measured in a cuvette with a light path length of 10-mm using a spectrophotometer (UV-2600; Shimazu, Japan) equipped with an integrating sphere (ISR-2600Plus, Shimazu, Japan). To quantify chlorophyll *a* and carotenoids, cell pellets of respective cultures were resuspended in 1.0 mL of *N,N*-dimethylformamide. Cells were removed by centrifugation at 15,000 × *g* for 10 min. Absorbance of the supernatant fractions at 480, 647 and 664 nm was measured with a spectrophotometer (BioSpectrometer basic; Eppendorf, Germany). The chlorophyll *a* and carotenoid concentrations were calculated according to ref. (27). The phycocyanin content was calculated according to the method previously described (28).

For thin layer chromatography (TLC) analysis, cells were harvested using centrifugation at 2,000 g for ten min at 4°C after adding Tween 20 (final concentration of 0.001%) to the 200-mL culture. The cell pellets were resuspended in 20-mL phosphate-buffered saline and centrifuged as above. The cell pellets were immediately frozen in liquid nitrogen and dewatered on a freeze-drier (FDU-1200, EYELA, Japan). The lyophilized samples (10-mg each) were subjected to 50-µL hexane for pigment extraction. Then, 20-µL extract was separated on a Silica gel 60 TLC plate (1.05582.0001, Merck Millipore) using the mobile phase consisting of a mixture of hexane and acetone (7:3 by volume).

To determine the growth rate of the wild-type, Δ*CHLD*, Δ*PSY,* and Δ*CHLD* Δ*PSY* haploid cells in autotrophic, mixotrophic and heterotrophic conditions, the cells grown heterotrophically as above were inoculated into 20-mL MA medium (autotrophic condition) or MA medium supplemented with 100 mM glucose (mixotrophic and heterotrophic conditions), respectively. The cells were cultured in 25-cm^2 tissue culture flasks in a normal (heterotrophic condition) or CO_2 (autotrophic and mixotrophic conditions) incubator at 42°C in the dark on a rotary shaker (140 rpm) at 42°C in the light (50 µmol m^{-2} s^{-1}; autotrophic and mixotrophic conditions) or dark (heterotrophic condition), respectively.

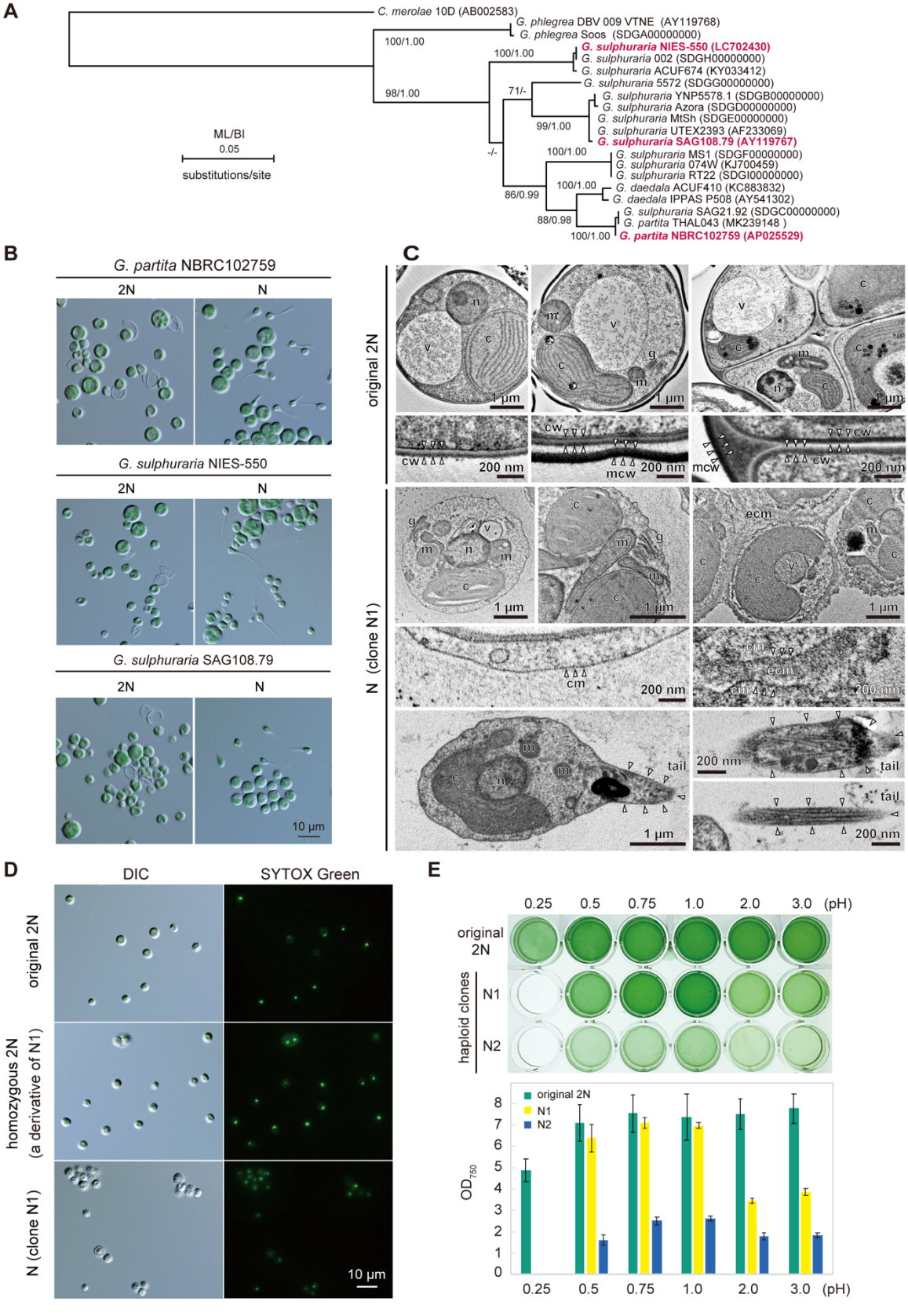
A
C. merolae 10D (AB002583)
G. phlegrea DBV 009 VTNE (AY119768)
G. phlegrea Soos (SDGA00000000)
G. sulphuraria NIES-550 (LC702430)
G. sulphuraria 002 (SDGH00000000)
G. sulphuraria ACUF674 (KY033412)
G. sulphuraria 5572 (SDGG00000000)
G. sulphuraria YNP5578.1 (SDGB00000000)
G. sulphuraria Azora (SDGD00000000)
G. sulphuraria MtSh (SDGE00000000)
G. sulphuraria UTEX2393 (AF233069)
G. sulphuraria SAG108.79 (AY119767)
G. sulphuraria MS1 (SDGF00000000)
G. sulphuraria 074W (KJ700459)
G. sulphuraria RT22 (SDGI00000000)
G. daedala ACUF410 (KC883832)
G. daedala IPPAS P508 (AY541302)
G. sulphuraria SAG21.92 (SDGC00000000)
G. partita THAL043 (MK239148)
G. partita NBRC102759 (AP025529)
100/1.00
100/1.00
98/1.00
71/-
99/1.00
-/-
100/1.00
100/1.00
86/0.99
88/0.98
100/1.00
ML/BI
0.05
substitutions/site
B
G. partita NBRC102759
2N
N
G. sulphuraria NIES-550
2N
N
G. sulphuraria SAG108.79
2N
N
10 µm
C
original 2N
N (clone N1)
1 µm
200 nm
cw
mcw
ecm
cm
tail
D
DIC
SYTOX Green
original 2N
homozygous 2N (a derivative of N1)
N (clone N1)
10 µm
E
0.25
0.5
0.75
1.0
2.0
3.0
(pH)
original 2N
haploid clones
N1
N2
OD750
0
1
2
3
4
5
6
7
8

Fig. S1. Comparison of cellular morphology, nuclear DNA content and growth rate at various pH between diploid and haploid cells of *Galdieria*. (A) Phylogenetic relationship of several *Galdieria* strains. The tree based on nucleotide sequences of the chloroplast-encoded rbcL was generated using the maximum likelihood analysis (RAxML-NG ver. 1.0.3) (29). *C. merolae* was used as an outgroup. Bootstrap values (BP) >50% obtained using ML and posterior probabilities (BI) >0.95 calculated using the Bayesian analysis (MrBayes ver. 3.2.7) (30) are indicated above the branches. The branch lengths reflect the evolutionary distances indicated by the scale bar. The analyses were conducted as described previously (17). The accession numbers of the respective nucleotide sequences are indicated with the species and strain names. **(B)** Micrographs of the diploid (2N; known form and obtained from culture collections, NBRC, NIES and SAG) and the haploid (N; newly generated in this study) cells of the three *Galdieria* strains. **(C)** Transmission electron micrographs of the *G. partit*a original 2N clone and N clone N1 generated from the original 2N clone. c, chloroplast; cm, cell membrane; cw, daughter cell wall; ecm, extracellular matrix; g, Golgi apparatus; m, mitochondrion; mcw, mother cell wall; n, nucleus; tail, tail of a haploid cell, v, vacuole. **(D)** Fluorescent visualization of nuclear DNA in G. partita 2N and N cells with SYTOX Green. Results of flow cytometric analysis are shown in Fig. 1E and S4C. **(E)** The photoautotrophic growth rate of *G. partita* 2N and N cells at various pH. The original 2N clone and two different N clones (N1 and N2) generated from the original 2N clone were inoculated into an inorganic medium (OD_{750} = 0.2) at a series of pH and cultured in the light for 7 d. Then OD_{750} of respective cultures was determined. Data represent mean and standard deviation of three independent cultures.

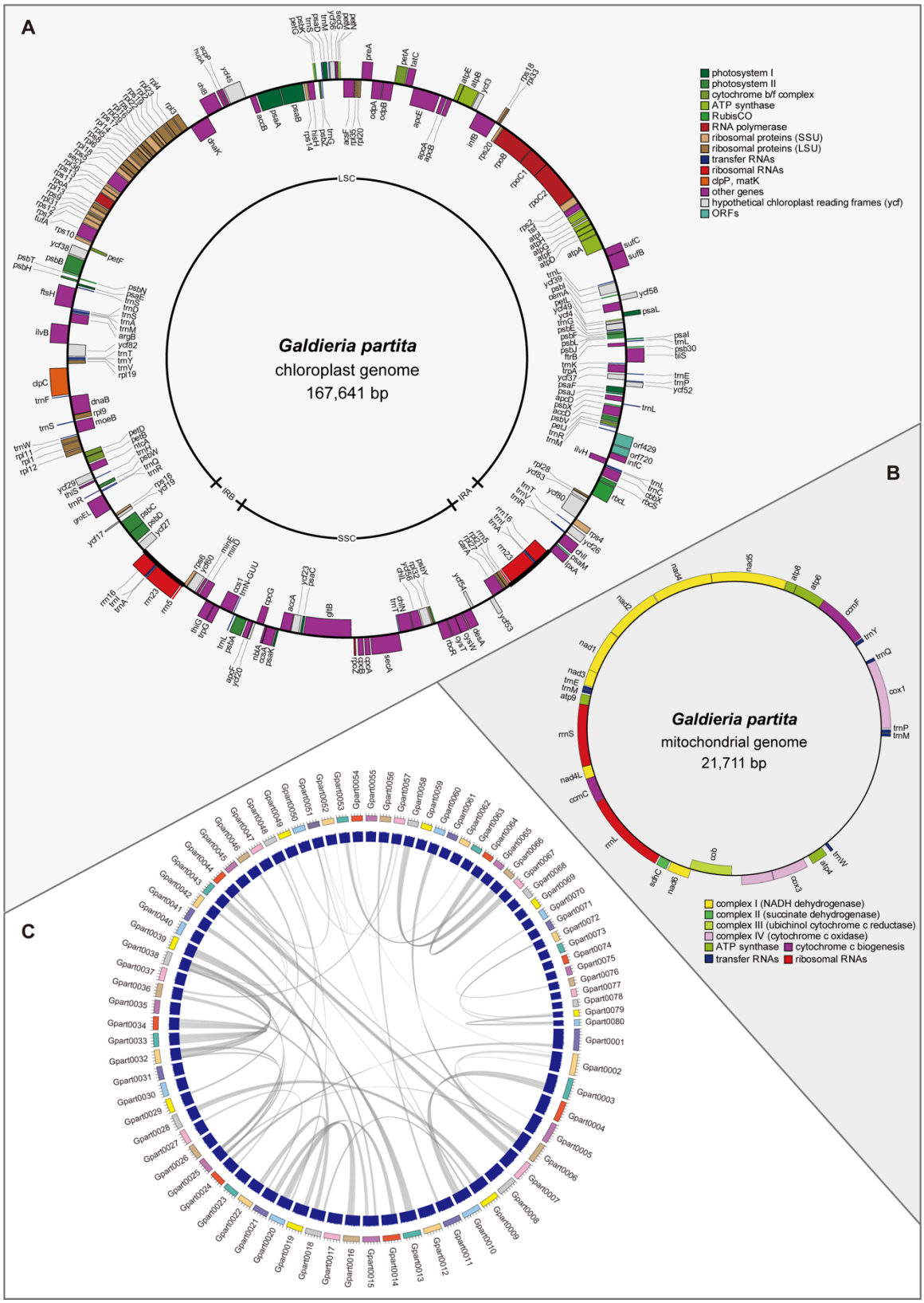
A
Galdieria partita
chloroplast genome
167,641 bp
LSC
IRB
IRA
SSC
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
RubisCO
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
transfer RNAs
ribosomal RNAs
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
ORFs
B
Galdieria partita
mitochondrial genome
21,711 bp
complex I (NADH dehydrogenase)
complex II (succinate dehydrogenase)
complex III (ubichinol cytochrome c reductase)
complex IV (cytochrome c oxidase)
ATP synthase
cytochrome c biogenesis
transfer RNAs
ribosomal RNAs
C

Fig. S2. The structure of the nuclear, chloroplast and mitochondrial genomes of *G. partita*. **(A, B)** The chloroplast **(A)** and mitochondrial **(B)** genome maps of *G. partita*. Annotated genes are colored following functional categories. Genes on the inside are transcribed in the clockwise direction, whereas genes on the outside are transcribed in the counterclockwise direction. **(C)** An overview of duplicated regions in the nuclear genome of *G. partita*. The outer ring shows contigs of the nuclear genome with the scale line at an interval of 50 kb. The inner purple ring illustrates the sequencing depth. The links inside represent duplicated colinear blocks of protein-coding genes identified using MCScanX.

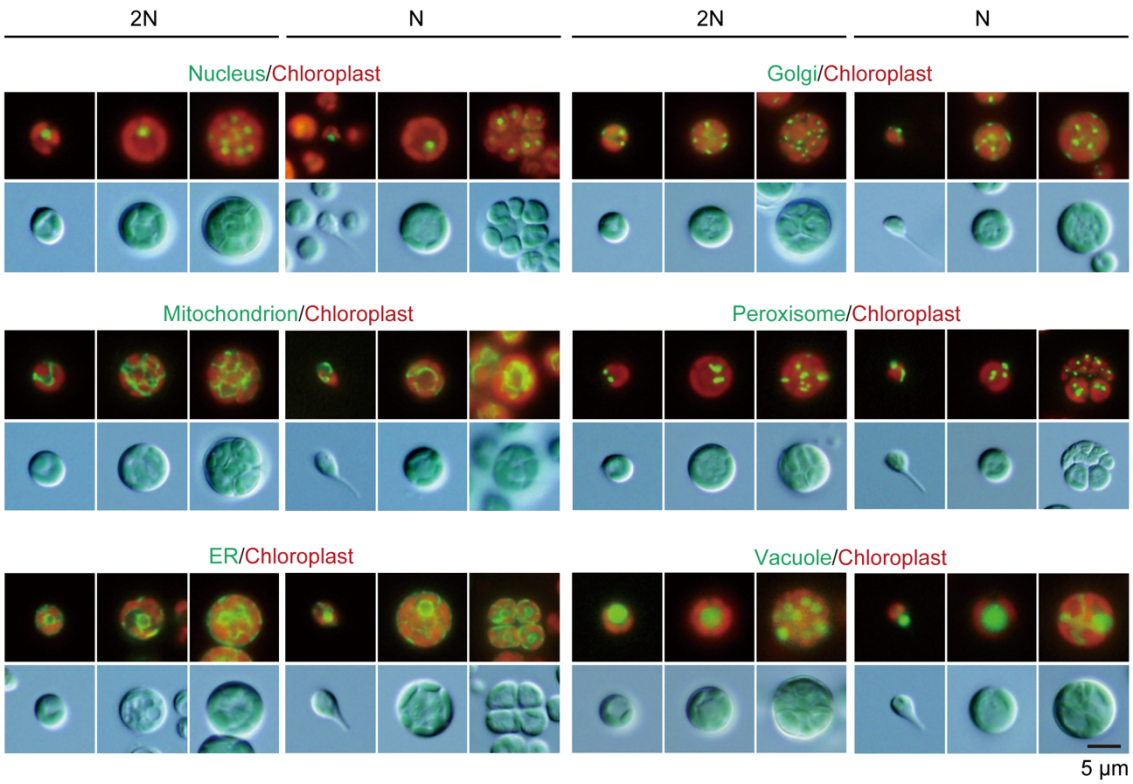

Fig. S3. Fluorescent visualization of membranous organelles in *G. partita* diploid and haploid cells. To visualize respective organelles other than the vacuole using fluorescence microscopy, following transgenes encoding fluorescent proteins were integrated into a chromosomal intergenic region between g7632.t1 and g7631.t1 loci of *G. partita* haploid (N) clone N1 and then transformed into homozygous diploids (2N) by self-diploidization. mVenus with a nucleic localization signal (NLS) on the N and C terminals for the nucleus, mVenus with the mitochondrion-targeted peptide of elongation factor thermo unstable (EF-Tu) on the N terminal for the mitochondrion, Calnexin-mVenus for the ER, ERD2-mVenus for the Golgi apparatus, and mVenus- peroxisome targeting signal 1 (PTS1) for the peroxisome. To visualize vacuole, N clone N1 and the homozygous 2N obtained from N1 were stained using Quinacrine. The green is mVenus/Quinacrine fluorescence, and red is autofluorescence of the chloroplast. Corresponding differential interference contrast images are also shown.

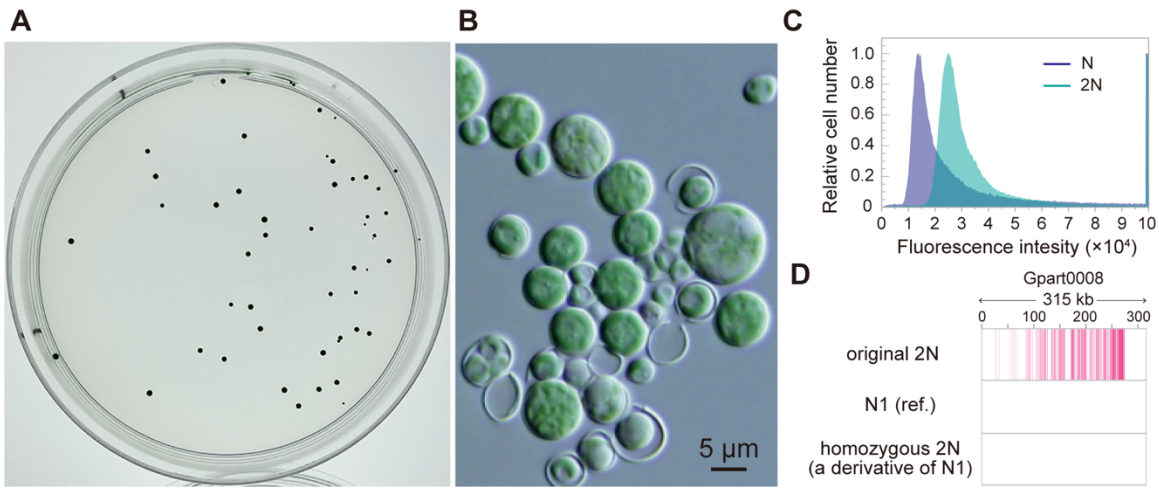

Fig. S4. Generation of homozygous diploids of *G. partita*. To generate a homozygous diploid (2N) from a *G. partita* haploid (N) clone N1 by self-diploidization, N cells growing in an inorganic medium at pH 1.0 were spread on a gellan gum-solidified inorganic medium at pH 2.0 supplemented with 1-mM sodium acetate. Then the cells on the medium plate were statically cultured in the light for 3 weeks until homozygous 2N form visible colonies. **(A)** Photograph of the plate medium on which homozygous 2N form colonies. **(B)** A micrograph of the cells in a colony. **(C)** Flow cytometric analysis of 2N and N cells stained with SYTOX Green. **(D)** Confirmation of homozygosity by mapping SNPs/indels positions of the original heterozygous 2N and homozygous 2N clones. Sites that match the reference genome sequence (N clone N1) are shown in white, heterozygous SNPs/indels are shown in red, and homozygous SNPs/indels are shown in blue.

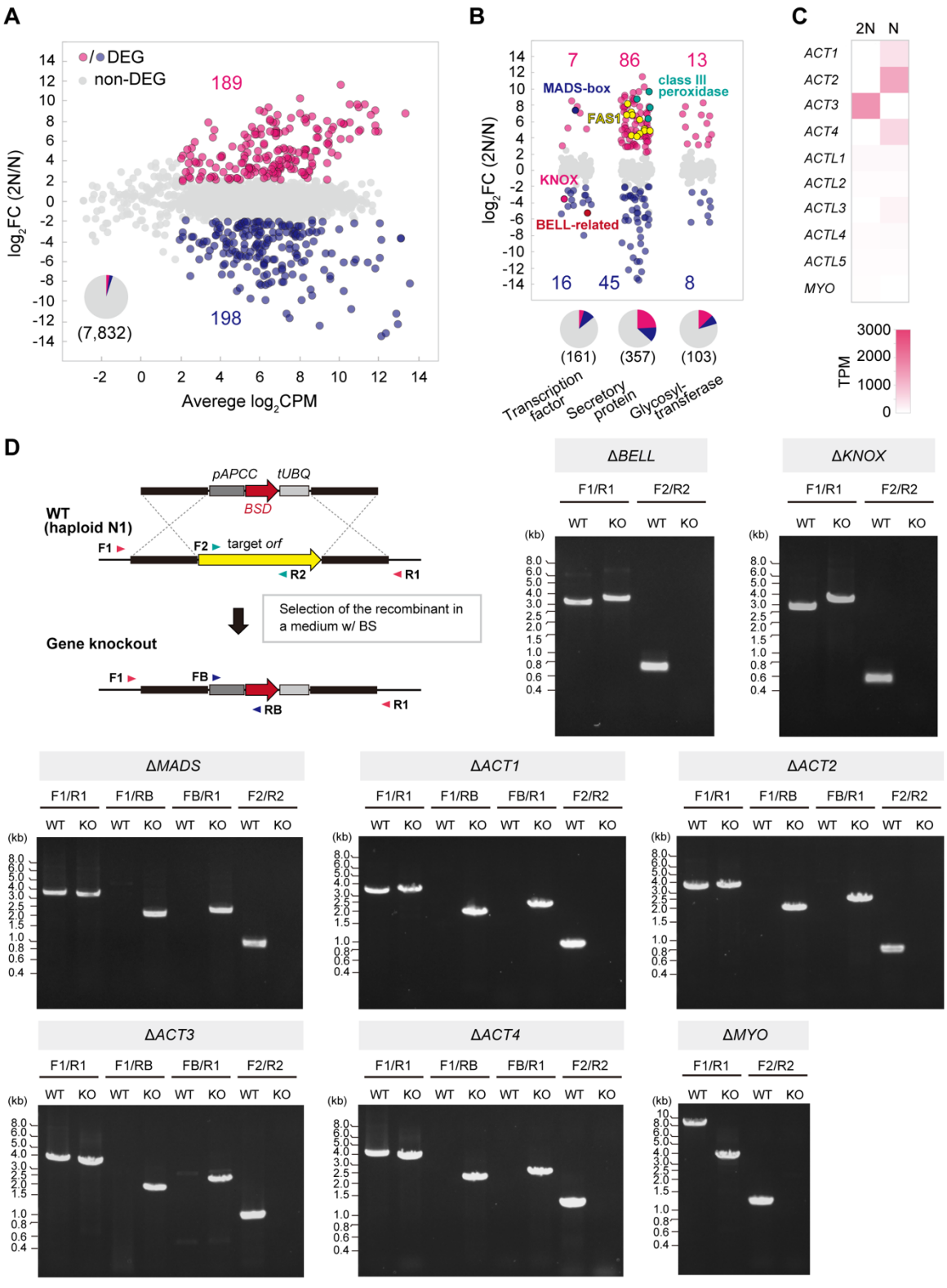
A
DEG
non-DEG
189
198
(7,832)
log2FC (2N/N)
Averege log2CPM
B
7
86
13
MADS-box
class III peroxidase
FAS1
KNOX
BELL-related
16
45
8
(161)
(357)
(103)
Transcription factor
Secretory protein
Glycosyl-transferase
C
2N
N
ACT1
ACT2
ACT3
ACT4
ACTL1
ACTL2
ACTL3
ACTL4
ACTL5
MYO
TPM
3000
2000
1000
0
D
pAPCC
tUBQ
BSD
WT (haploid N1)
F1
F2
target orf
R2
R1
Selection of the recombinant in a medium w/ BS
Gene knockout
FB
RB
ΔBELL
ΔKNOX
ΔMADS
ΔACT1
ΔACT2
ΔACT3
ΔACT4
ΔMYO
F1/R1
F1/RB
FB/R1
F2/R2
WT
KO
(kb)

Fig. S5. Comparison of transcriptome between *G. partita* diploid and haploid and the generation of BELL, KNOX, MADS, actin and myosin gene knockouts. (A) MA plot showing 189 genes upregulated in diploid [homozygous diploid (2N) derived from N2] and 198 genes upregulated in haploid (N; clone N2) among 7,832 nucleus-encoded genes (DEGs, FDR < 0.01, logCPM > 2 and logFC > 2 or < −2 in three independent cultures). **(B)** Ratio of mRNA abundance of genes encoding transcription factors, secretary proteins, and glycosyltransferases between 2N and N cells (clone N2). **(C)** Comparison of mRNA abundance of actin (*ACT*), actin-like (*ACTL*), and myosin (*MYO*) genes between 2N and N cells (clone N2). **(D)** To knockout respective gene in *G. partita*, *BSD* selectable marker with *pAPCC* and *tUBQ* was integrated into the chromosome of the N clone N1 by homologous recombination so that respective *orf* was replaced with *BSD* marker. Replacement of a *BELL*, *KNOX*, *MADS*, *ACT* or *MYO orf* with the *BSD* marker was confirmed using the primers (F1, F2, FB, R1, R2, and RB) indicated by the arrowheads in the illustration. The wild-type N clone N1 (WT) served as a negative control.

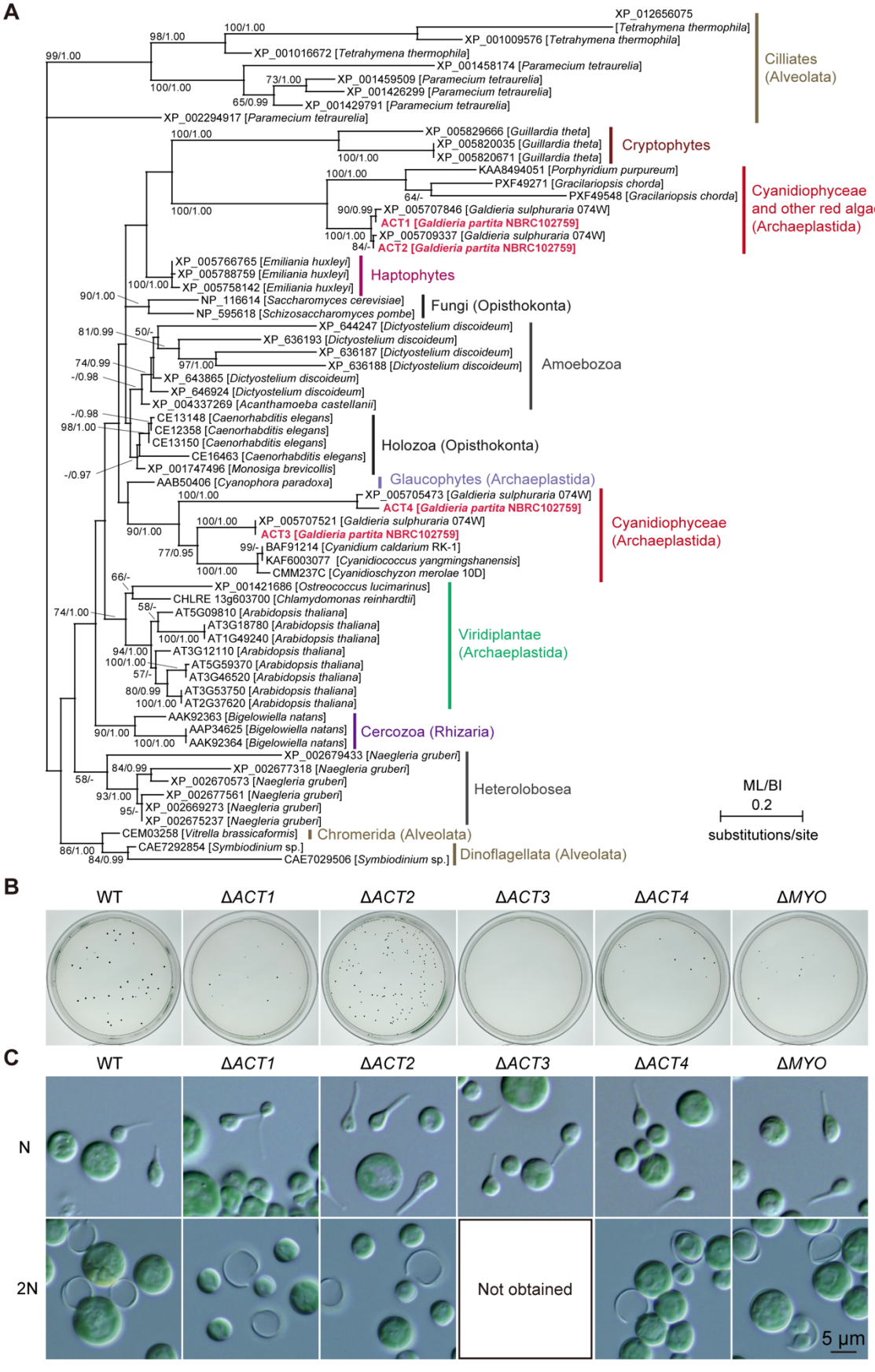
A
XP_012656075 [Tetrahymena thermophila]
XP_001009576 [Tetrahymena thermophila]
XP_001016672 [Tetrahymena thermophila]
XP_001458174 [Paramecium tetraurelia]
XP_001459509 [Paramecium tetraurelia]
XP_001426299 [Paramecium tetraurelia]
XP_001429791 [Paramecium tetraurelia]
XP_002294917 [Paramecium tetraurelia]
Cilliates (Alveolata)
XP_005829666 [Guillardia theta]
XP_005820035 [Guillardia theta]
XP_005820671 [Guillardia theta]
Cryptophytes
KAA8494051 [Porphyridium purpureum]
PXF49271 [Gracilariopsis chorda]
PXF49548 [Gracilariopsis chorda]
XP_005707846 [Galdieria sulphuraria 074W]
ACT1 [Galdieria partita NBRC102759]
XP_005709337 [Galdieria sulphuraria 074W]
ACT2 [Galdieria partita NBRC102759]
Cyanidiophyceae and other red algae (Archaeplastida)
XP_005766765 [Emiliania huxleyi]
XP_005788759 [Emiliania huxleyi]
XP_005758142 [Emiliania huxleyi]
Haptophytes
NP_116614 [Saccharomyces cerevisiae]
NP_595618 [Schizosaccharomyces pombe]
Fungi (Opisthokonta)
XP_644247 [Dictyostelium discoideum]
XP_636193 [Dictyostelium discoideum]
XP_636187 [Dictyostelium discoideum]
XP_636188 [Dictyostelium discoideum]
XP_643865 [Dictyostelium discoideum]
XP_646924 [Dictyostelium discoideum]
XP_004337269 [Acanthamoeba castellanii]
Amoebozoa
CE13148 [Caenorhabditis elegans]
CE12358 [Caenorhabditis elegans]
CE13150 [Caenorhabditis elegans]
CE16463 [Caenorhabditis elegans]
XP_001747496 [Monosiga brevicollis]
Holozoa (Opisthokonta)
AAB50406 [Cyanophora paradoxa]
Glaucophytes (Archaeplastida)
XP_005705473 [Galdieria sulphuraria 074W]
ACT4 [Galdieria partita NBRC102759]
XP_005707521 [Galdieria sulphuraria 074W]
ACT3 [Galdieria partita NBRC102759]
BAF91214 [Cyanidium caldarium RK-1]
KAF6003077 [Cyanidiococcus yangmingshanensis]
CMM237C [Cyanidioschyzon merolae 10D]
Cyanidiophyceae (Archaeplastida)
XP_001421686 [Ostreococcus lucimarinus]
CHLRE 13g603700 [Chlamydomonas reinhardtii]
AT5G09810 [Arabidopsis thaliana]
AT3G18780 [Arabidopsis thaliana]
AT1G49240 [Arabidopsis thaliana]
AT3G12110 [Arabidopsis thaliana]
AT5G59370 [Arabidopsis thaliana]
AT3G46520 [Arabidopsis thaliana]
AT3G53750 [Arabidopsis thaliana]
AT2G37620 [Arabidopsis thaliana]
Viridiplantae (Archaeplastida)
AAK92363 [Bigelowiella natans]
AAP34625 [Bigelowiella natans]
AAK92364 [Bigelowiella natans]
Cercozoa (Rhizaria)
XP_002679433 [Naegleria gruberi]
XP_002677318 [Naegleria gruberi]
XP_002670573 [Naegleria gruberi]
XP_002677561 [Naegleria gruberi]
XP_002669273 [Naegleria gruberi]
XP_002675237 [Naegleria gruberi]
Heterolobosea
CEM03258 [Vitrella brassicaformis]
Chromerida (Alveolata)
CAE7292854 [Symbiodinium sp.]
CAE7029506 [Symbiodinium sp.]
Dinoflagellata (Alveolata)
ML/BI
0.2
substitutions/site
B
WT
ΔACT1
ΔACT2
ΔACT3
ΔACT4
ΔMYO
C
WT
ΔACT1
ΔACT2
ΔACT3
ΔACT4
ΔMYO
N
2N
Not obtained
5 μm

Fig. S6. Phylogenetic relationship of actin proteins and generation of homozygous diploids of actin and myosin gene knockouts of *G. partita.* (A) Phylogenetic relationship of *G. partita* ACT1, ACT2, ACT3, and ACT4 proteins. The tree based on amino acid sequences of actin proteins was generated using the maximum likelihood analysis (RAxML-NG ver. 1.0.3) (29). Bootstrap values (BP) >50% obtained using ML and posterior probabilities (BI) >0.95 calculated using the Bayesian analysis (MrBayes ver. 3.2.7) (30) are indicated above the branches. The branch lengths reflect the evolutionary distances indicated by the scale bar. The analyses were conducted as described previously (17). The accession numbers or locus numbers of the respective amino acid sequences are indicated with the species and strain names. **(B)** To generate a homozygous diploid (2N) from a *G. partita* Δ*ACT1,* Δ*ACT2,* Δ*ACT3,* Δ*ACT4,* and Δ*MYO* haploids (N) by self-diploidization, N cells growing in an inorganic medium at pH 1.0 were spread on a gellan gum-solidified medium at pH 2.0 supplemented with 1-mM sodium acetate. Then the cells on the medium plate were statically cultured in the light for three 3 weeks until homozygous 2N form visible colonies. Photographs of the plate medium on which homozygous 2N form colonies are shown. Efficiency of generation of homozygous 2N colonies is summarized in Fig. 5J. **(C)** Micrographs of N and 2N cells of respective gene knockouts.

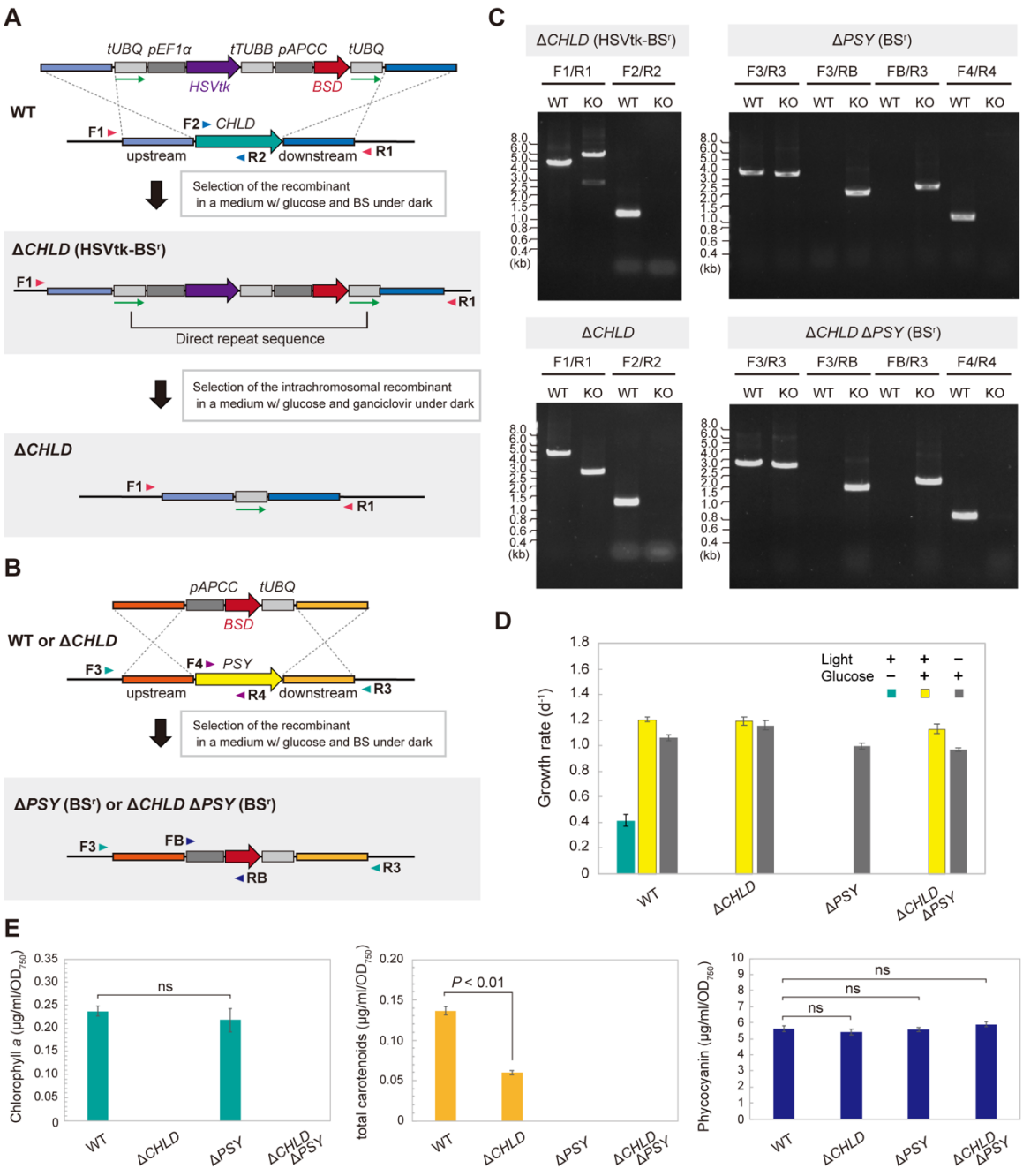
A
tUBQ pEF1α
tTUBB pAPCC
tUBQ
HSVtk
BSD
WT
F1
F2 CHLD
upstream
R2
downstream
R1
Selection of the recombinant in a medium w/ glucose and BS under dark
ΔCHLD (HSVtk-BSr)
F1
R1
Direct repeat sequence
Selection of the intrachromosomal recombinant in a medium w/ glucose and ganciclovir under dark
ΔCHLD
F1
R1
B
pAPCC
tUBQ
BSD
WT or ΔCHLD
F3
F4 PSY
upstream
R4
downstream
R3
Selection of the recombinant in a medium w/ glucose and BS under dark
ΔPSY (BSr) or ΔCHLD ΔPSY (BSr)
F3
FB
RB
R3
C
ΔCHLD (HSVtk-BSr)
ΔPSY (BSr)
F1/R1
F2/R2
F3/R3
F3/RB
FB/R3
F4/R4
WT KO WT KO
WT KO WT KO WT KO WT KO
8.0 6.0 5.0 4.0 3.0 2.5 2.0 1.5 1.0 0.8 0.6 0.4 (kb)
ΔCHLD
ΔCHLD ΔPSY (BSr)
D
Growth rate (d-1)
1.8 1.6 1.4 1.2 1.0 0.8 0.6 0.4 0.2 0
Light + + −
Glucose − + +
WT
ΔCHLD
ΔPSY
ΔCHLD ΔPSY
E
Chlorophyll a (μg/ml/OD750)
0.35 0.30 0.25 0.20 0.15 0.10 0.05 0
ns
total carotenoids (μg/ml/OD750)
0.20 0.15 0.10 0.05 0
P < 0.01
Phycocyanin (μg/ml/OD750)
10 9 8 7 6 5 4 3 2 1 0
ns
ns
ns
WT
ΔCHLD
ΔPSY
ΔCHLD ΔPSY

Fig. S7. The generation, growth rate, and pigment contents of Δ*CHLD*, Δ*PSY* and Δ*CHLD* Δ*PSY G. partita* haploids. (A) To knockout *CHLD* in the haploid (N) clone N1, the *HSVtk* suicide marker, which was connected with *pEF1a* and *tTUBB*, and the *BSD* selectable marker, which was connected with *pAPCC* and *tUBQ*, were integrated into the chromosomal *CHLD* locus by homologous recombination. To remove *HSVtk* and *BSD* markers later, these expression cassettes were sandwiched between a direct repeat of a *tUBQ* (green arrows). The recombinant [Δ*CHLD* (HSVtk-BS^r), photosynthesis-deficient] was selected in a medium supplemented with glucose and BS in the dark. To remove *HSVtk* and *BSD* markers through intrachromosomal recombination between the direct repeats of a *tUBQ,* the intrachromosomal recombinant (Δ*CHLD*, photosynthesis-deficient) was selected in a medium supplemented with glucose and ganciclovir in the dark. **(B)** To knockout *PSY* in the wild-type N clone N1, or Δ*CHLD* N clone prepared as above, *BSD* selectable marker, which was connected to *pAPCC* and *tUBQ*, were integrated into the chromosomal *PSY* locus by homologous recombination. The recombinant [Δ*PSY* (BS^r) or photosynthesis-deficient Δ*CHLD* Δ*PSY* (BS^r)] was selected in a medium supplemented with glucose and BS in the dark. **(C)** The recombination events in respective transformants were confirmed using PCR with the primers (F1, F2, F3, F4, FB, R1, R2, R3, R4, and RB) indicated by the arrowheads in **A** and **B**. The wild-type N clone N1 (WT) served as a negative control. **(D, E)** To determine the growth rate **(D)** and pigment contents **(E)** of the wild-type, Δ*CHLD*, Δ*PSY,* and Δ*CHLD* Δ*PSY* N cells in autotrophic, mixotrophic and heterotrophic conditions, the cells grown heterotrophically were inoculated into an inorganic medium (autotrophic condition) or a medium supplemented with 100 mM glucose (mixotrophic and heterotrophic conditions), respectively. The cells were cultured in the light (50 μmol m^{-2} s^{-1}; autotrophic and mixotrophic conditions) or dark (heterotrophic condition), respectively. The chlorophyll *a*, carotenoid and phycocyanin contents in the heterotrophic culture were estimated by spectrophotometric method according to the method previously described (27, 28). Data represent mean and standard deviation of three independent cultures. Photographs of the cultures are shown in Fig. 6A.

Table S1. Summary of genome analysis of *G. partita* NBRC102759.

Nuclear genome statistics	
Genome size	17.78 Mb
Genomic G+C content	37.6%
Telomere sequences	CCCTAA(A)TAAA
No. of contigs	80 (telomere-to-telomere)
No. of protein coding genes	7,832
Average CDS length	1,351 bp
CDS G+C content	39.7%
Chloroplast genome statistics	
Genome size	167,641 bp
Genomic G+C content	28.4%
No. of protein coding genes	186
No. of tRNA	39
No. of rRNA	6
Mitochondrial genome statistics	
Genome size	21,711 bp
Genomic G+C content	43.4%
No. of protein coding genes	18
No. of tRNA	7
No. of rRNA	2

Movie S1 (separate file). Time-lapse observation of *G. partita* diploid cells. Cells of the original diploid clone obtained from NBRC in a medium supplemented with 100-mM glucose were statically cultured and observed. Time (sec).

Movie S2 (separate file). Time-lapse observation of hatching of eight daughter cells from a *G. partita* diploid mother cell. Cells of the original diploid clone in a medium supplemented with 100-mM glucose were statically cultured and observed. Time (sec).

Movie S3 (separate file). Time-lapse observation of *G. partita* haploid cells. Cells of haploid clone N1 in a medium supplemented with 100-mM glucose were statically cultured and observed. Time (sec).

Movie S4 (separate file). Time-lapse observation of the conversion of a tad-pole-shaped *G. partita* haploid cell to a round cell. Cells of haploid clone N1 in a medium supplemented with 100-mM glucose were statically cultured and observed. Time (sec).

Movie S5 (separate file). Time-lapse observation of three round of successive cell divisions of a *G. partita* haploid cell. Cells of haploid clone N1 in a medium supplemented with 100-mM glucose were statically cultured and observed. Time (min).

Movie S6 (separate file). Time-lapse observation of conversion of round *G. partita* haploid daughter cells after cell division to tad-pole-shaped cells. Cells of haploid clone N1 in a medium supplemented with 100-mM glucose were statically cultured and observed. Time (sec).

Movie S7 (separate file). Time-lapse observation of the *G. partita* wild-type (WT), Δ*ACT*, and Δ*MYO* tad-pole-shaped haploid cells. The wild-type (WT) haploid clone N1 and Δ*ACT*, and Δ*MYO* cells in an inorganic medium were statically cultured and observed. WT (upper left), Δ*ACT1* (upper middle), Δ*ACT2* (upper right), Δ*ACT3* (lower left), Δ*ACT4* (lower middle), and Δ*MYO* (lower right) are shown. Time (sec).

Dataset S1 (separate file). List of *G. partita* genes, edgeR output values, and TPM values of haploid clones N1 and N2 and homozygous diploid clones obtained from N1 and N2, respectively, grown photoautotrophically. N/A indicates genes that could not be annotated by Blast2GO.

Dataset S2 (separate file). List of nucleotide sequences of primers and synthetic DNAs used in this study.

SI References

1. M. Marcel, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**, 10-12 (2011).
2. A. Bankevich *et al.*, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**, 455-477 (2012).
3. M. Tillich *et al.*, GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res* **45**, W6-W11 (2017).
4 . K. Jain *et al.*, Extreme features of the *Galdieria sulphuraria* organellar genomes: a consequence of polyextremophily? *Genome Biol Evol* **7**, 367-380 (2014).
5. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018).
6. S. Koren *et al.*, Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**, 722-736 (2017).
7. R. Vaser, I. Sovic, N. Nagarajan, M. Sikic, Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* **27**, 737-746 (2017).
8. B. J. Walker *et al.*, Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).
9. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
10. M. Stanke, M. Diekhans, R. Baertsch, D. Haussler, Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637-644 (2008).
11. A. Conesa *et al.*, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674-3676 (2005).
12. J. Jin *et al.*, PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res* **45**, D1040-D1045 (2017).
13. J. J. Almagro Armenteros *et al.*, SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* **37**, 420-423 (2019).
14. H. Zhang *et al.*, dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **46**, W95-W101 (2018).
15. Y. Wang *et al.*, MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* **40**, e49 (2012).
16. C. Chen *et al.*, TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant* **13**, 1194-1202 (2020).
17. S. Hirooka *et al.*, Acidophilic green algal genome provides insights into adaptation to an acidic environment. *Proc Natl Acad Sci U S A* **114**, E8304-E8313 (2017).
18. W. H. Cheong, Y. C. Tan, S. J. Yap, K. P. Ng, ClicO FS: an interactive web-based service of Circos. *Bioinformatics* **31**, 3685-3687 (2015).
19. A. McKenna *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303 (2010).
20. J. T. Robinson *et al.*, Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26 (2011).
21. H. Li *et al.*, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
22. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
23. R. Ihaka, A language for data analysis and graphics. *J. Comp. Graph. Stat.* **5**, 299-314 (1996).
24. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
25. B. Li, V. Ruotti, R. M. Stewart, J. A. Thomson, C. N. Dewey, RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics* **26**, 493-500 (2010).
26. T. Fujiwara *et al.*, A nitrogen source-dependent inducible and repressible gene expression system in the red alga *Cyanidioschyzon merolae*. *Front Plant Sci* **6**, 657 (2015).
27. A. R. Wellburn, The spectral determination of chlorophylls *a* and *b*, as well as total carotenoids, using various solvents with spectrophotometers of different resolution. *J. Plant*

Physiol. **144**, 307-313 (1994).
28. D. I. Arnon, B. D. McSwain, H. Y. Tsujimoto, K. Wada, Photochemical activity and components of membrane preparations from blue-green algae. I. Coexistence of two photosystems in relation to chlorophyll *a* and removal of phycocyanin. *Biochim Biophys Acta* **357**, 231-245 (1974).
29. A. Stamatakis, RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690 (2006).
30. F. Ronquist, J. P. Huelsenbeck, MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574 (2003).