OPEN

Genetic tool development in marine protists: emerging model organisms for experimental cell biology

Diverse microbial ecosystems underpin life in the sea. Among these microbes are many unicellular eukaryotes that span the diversity of the eukaryotic tree of life. However, genetic tractability has been limited to a few species, which do not represent eukaryotic diversity or environmentally relevant taxa. Here, we report on the development of genetic tools in a range of protists primarily from marine environments. We present evidence for foreign DNA delivery and expression in 13 species never before transformed and for advancement of tools for eight other species, as well as potential reasons for why transformation of yet another 17 species tested was not achieved. Our resource in genetic manipulation will provide insights into the ancestral eukaryotic lifeforms, general eukaryote cell biology, protein diversification and the evolution of cellular pathways.

The ocean represents the largest continuous planetary ecosystem, hosting an enormous variety of organisms, which include microscopic biota such as unicellular eukaryotes (protists). Despite their small size, protists play key roles in marine biogeochemical cycles and harbor tremendous evolutionary diversity. Notwithstanding their significance for understanding the evolution of life on Earth and their role in marine food webs, as well as driving biogeochemical cycles to maintain habitability, little is known about their cell biology including reproduction, metabolism and signaling. Most of the biological knowledge available is based on comparison of proteins from cultured species to homologs in genetically tractable model taxa. A main impediment to understanding the cell biology of these diverse eukaryotes is that protocols for genetic modification are only available for a small number of species that represent neither the most ecologically relevant protists nor the breadth of eukaryotic diversity.

The development of genetic tools requires reliable information about gene organization and regulation of the emergent model species. Over the last decade, genome1 and transcriptome sequencing initiatives2–6 have resulted in nearly 120 million unigenes being identified in protists6, which facilitates the developments of genetic tools used for model species. Insights from these studies enabled the phylogenetically informed approach for selecting and developing key marine protists into model systems in the Environmental Model Systems (EMS) Project presented herein. Forty-one research groups took part in the EMS Project, a collaborative effort resulting in the development of genetic tools that significantly expand the number of eukaryotic lineages that can be manipulated, and that encompass multiple ecologically important marine protists.

Here, we summarize detailed methodological achievements and analyze results to provide a synthetic ‘transformation roadmap’ for creating new microeukaryotic model systems. Although the organisms reported here are diverse, the paths to overcome difficulties share similarities, highlighting the importance of building a well-connected community to overcome technical challenges and accelerate the development of genetic tools. The 13 emerging model species presented herein, and the collective set of genetic tools from the overall collaborative project, will not only extend our knowledge of marine cell biology, evolution and functional biodiversity, but also serve as platforms to advance protistan biotechnology.

results

overview of taxa in the EMS initiative. Taxa were selected from multiple eukaryotic supergroups7 to maximize the potential of cellular biology and to evaluate the numerous unigenes with unknown functions found in marine protists (Fig. 1). Before the EMS initiative, reproducible transformation of marine protists was limited to only a few species such as *Thalassiosira pseudonana*, *Phaeodactylum tricornutum* and *Ostreococcus tauri* (Supplementary Table 1). The EMS initiative included 39 species, specifically, 6 archaeplasts, 2 haptophytes, 2 rhizarians, 9 stramenopiles, 12 alveolates, 4 discozoans and 4 opisthokonts (Fig. 1). Most of these taxa were isolated from coastal habitats, the focus area of several culture collections. More than 50% of the selected species are considered photoautotrophs, with another 35% divided between heterotrophic osmotrophs and phagotrophs, the remainder being predatory mixotrophs. Almost 20% of the chosen species are symbionts and/or parasites of marine plants or animals, 5% are associated with detritus and several are responsible for harmful algal blooms (Supplementary Table 2).

While some transformation systems for protists have been developed in the past7–11, the challenge for this initiative was to develop genetic tools for species that not only require different cultivation conditions but are also phenotypically diverse. It should be noted that not all main lineages were explored. For example, amoebozoans did not feature in this aquatic-focused initiative, in part because they tend to be most important in soils, at least based on current knowledge, and manipulation systems exist for members of this eukaryotic supergroup, such as *Dictyostelium discoideum*12. The overall EMS initiative outcomes are summarized in Fig. 1 and Table 1. We provide detailed protocols for 13 taxa, for which no transformation systems have been previously reported (category A) and eight taxa, for which existing protocols13–21 were advanced (category B; Figs. 2, 3 and 4, Table 1, Supplementary Tables 1–5 and Methods). We also review an already published EMS transformation protocol22 in one species (category C), and we discuss unsuccessful transformation attempts for 17 additional taxa (Fig. 1 and Methods). Finally, we synthesize our findings in a roadmap for the development of transformation systems in protists (Fig. 5).

*Archeplasts.* Prasinophytes are important marine green algae distributed from polar to tropical regions12. They form a sister group to
chlorophyte algae, and together, these two groups branch adjacent to land plants, collectively comprising the Viridiplantae, which are part of the Archaeplastida2,25 (Fig. 1). Genome sequences are available for the picoprasinophytes (<3 μm cell diameter) tested herein, specifically, Microcmonas commoda, M. pusilla, Ostreococcus lucimarinus and Bathycoccus prasinus. As part of the EMS initiative, we report on genetic tools for Bathycoccus, a scaled, nonmotile genus, and Micromonas, a motile, naked genus with larger genomes than Bathycoccus and Ostreococcus26,27. We also report on genetic tools for Tetrasselmis striata and O. lucimarinus. The latter was transformed based on an adapted homologous recombination system for O. tauri22,25.

O. lucimarinus (RC0802) and B. prasinos (RC4222) were transformed using protocols adapted from O. tauri22,25. Briefly, using electroporation for transfer of exogenous genes, O. lucimarinus was transformed using a DNA fragment encoding the O. lucimarinus high-affinity phosphate transporter (HAPT) gene fused to a luciferase gene and a kanamycin selection marker (Table 1 and Supplementary Table 3), which resulted in transient luciferase expression 24 h after electroporation (Table 1 and Fig. 3a). After 2 weeks of growth in low-melting agarose plates containing G418 (1 mg ml⁻¹), 480 colonies were obtained, picked and grown in artificial seawater with the antibiotic neomycin. Of these, 76 displayed luminescence ≥2.5-fold above background (80 relative luminescence units (RLU)), with widely variable levels (200–3,020 RLU), likely reflecting either variations in the site of integration and/or the number of integrated genes (Fig. 3a, Supplementary Fig. 1 and Methods).

The O. tauri construct did not work in B. prasinos, while the use of the B. prasinos histone H4 and HAP7 sequences in an otherwise identical construct and conditions was successful. Although luciferase expression was not detected 24 h after electroporation, 48 G418-resistant colonies were obtained 2 weeks later, 20 being luminescent when grown in liquid medium. Analysis of 14 resistant transformants revealed that the luciferase sequence was integrated into the genome of five luminescent clones, and one nonluminescent transformant revealed that the luciferase sequence was integrated into the genome of five luminescent clones, and one nonluminescent clone (Fig. 3b and Methods), suggesting that the chromatin context at integration sites in the latter was not favorable to luciferase expression.

Although transformation methods successful for Bathycoccus and Ostreococcus failed in Micromonas, Lonza nucleofection was successful with M. commoda (CCMP2709) (Table 1 and Fig. 3c) using two different codon-optimized plasmids, one encoding the luciferase gene (NanoLuc, Promega) flanked by an exogenous promoter and terminator sequence from the 5′ and 3′ untranslated regions (UTRs) of histone H3 in Micromonas polaris (CCMP2099), and the other encoding an enhanced green fluorescent protein (eGFP) gene flanked by endogenous promoter and terminator sequences from ribosomal protein S9 (Supplementary Table 5). Sensitivities to antibiotics were established (Supplementary Table 3). Constructs did not include a selectable marker, as we aimed to introduce and express foreign DNA while developing conditions suitable for transfection that supported robust growth in this cell wall-lacking protist (Table 1). Transformants revealed a significantly higher level of eGFP fluorescence than wild-type cells, with 1.3% of the population showing fluorescence per cell 45-fold higher than both the nontransformed portion of the culture and the wild-type cells (Fig. 3c and Methods). Additionally, the RLU was 1,500-fold higher than controls when using the luciferase-bearing construct, such that multiple experiments with both plasmids confirmed expression of exogenous genes in M. commoda.

T. striata (KAS-836) was transformed using microprojectile bombardment (Supplementary Fig. 2a). Two selectable marker genes were tested, consisting of a putative promoter and 5′ UTR sequences from the T. striata actin gene and either the coding sequences of the Streptolacticiobes hindustanus bleomycin gene (conferring resistance to zeocin) or the Streptomyces hygroscopicus bar gene (conferring resistance to gusfosinate) (Table 1, Supplementary Fig. 2a and Methods). The terminator sequence was obtained from the T. striata glyceraldehyde-3-phosphate dehydrogenase gene. Linearized plasmids were coated on gold particles and introduced into T. striata cells by using the PDS-1000/He Particle Delivery System (Bio-Rad). Transformants were successfully selected on half-strength f/2 at 50% salinity agar plates containing either 150 μg ml⁻¹ zeocin or 150 μg ml⁻¹ gusfosinate.

Haptophytes (incertae sedis). Haptophytes are a group of photosynthetic protists that are abundant in marine environments and include the principal calcifying lineage, the coccolithophores. Genome sequences are available for Emiliania huxleyi and Chrysochromulina tobinii, and there is one report of nuclear transformation of a calcifying coccolithophore species but transformation of E. huxleyi, the most prominent coccolithophore, has not been achieved yet. Here, as part of the EMS initiative, a stable nuclear transformation system was developed for Ischyrysis galbana, a species that lacks coccoliths, but represents an important feedstock for shellfish aquaculture.

I. galbana (CCMP1323) was transformed by biolistic bombardment with the plgNAT vector, which contains nourseothricin (NTC) N-acetyltransferase (NAT), (for nourseothricin resistance) driven by the promoter and terminator of Hsp70 from E. huxleyi (CCMP1516). Twenty-four hours after bombardment, cells were transferred to liquid f/2 medium at 50% salinity containing 80 μg ml⁻¹ NTC and left to grow for 2–3 weeks to select for transformants (Table 1). The presence of NAT in NTC-resistant cells was verified by PCR and PCR with reverse transcription (RT–PCR) (Fig. 4a, Supplementary Fig. 2b and Methods) and the sequence was verified. To confirm NAT resistance was a stable phenotype, cells were subcultured every 2–4 weeks at progressively higher NTC concentrations (up to 150 μg ml⁻¹) in the above-mentioned media. Cells remained resistant to NTC for approximately 6 months, as confirmed by PCR screening to identify the presence of the NAT gene.

Rhizarians. Rhizarians include diverse nonphotosynthetic protists, as well as the photosynthetic chlorarachniophytes that acquired a plastid via secondary endosymbiosis of a green alga. Uniquely, they represent an intermediate stage of the endosymbiotic process, since their plastids still harbor a relict nucleus (nucleomorph). Here, we report on an advanced transformation protocol for the chlorarachniophyte Amorophochlorella (Lotharella) amoebiformis for which low-efficiency transient transformation has previously been achieved using particle bombardment.

A. amoebiformis (CCMP2058) cells were resuspended in 100 μl of Gene Pulse Electroporation Buffer (Bio-Rad) with 20–50 μg of the reporter plasmid encoding eGFP-Rubisco fusion protein under the control of the native rbcS1 promoter and subjected to electroporation (Table 1). Cells were immediately transferred to fresh ESM medium and incubated for 24 h. Transformation efficiency was estimated by the fraction of cells expressing eGFP, resulting in 0.03–0.1% efficiency, as enumerated by microscopy, showing an efficiency up to 1,000-fold higher than in the previous study (Table 1). Stable transformants were generated by manual isolation using a micropipette, and a transformed line has maintained eGFP fluorescence for at least 10 months without antibiotic selection (Figs. 2 and 4b and Methods).

Stramenopiles. Stramenopiles are a diverse lineage harboring important photoautotrophic, mixotrophic (combining photosynthetic and phagotrophic nutrition) and heterotrophic taxa. As the most studied class in this lineage, diatoms (Bacillariophyceae) were early targets for the development of reverse genetics tools11,29. Diatoms are estimated to contribute approximately 20% of annual carbon fixation and, like several other algal lineages, are used in bioengineering applications and biofuels. Although other cold-adapted eukaryotes have, to our knowledge, yet to be transformed, here we present a protocol for the Antarctic diatom Fragilariopsis cylindrus.
A transformation protocol has also been developed herein for *Pseudo-nitzscha multiseries*, a toxin-producing diatom\(^1\). Here we also present work for nondiatom stramenoples, including a transformation protocol for the eustigmatophyte *Nannochloropsis oceanica*, and an alternative protocol for the labyrinthulomycete *Aurantiochytrium limacinum*\(^{10}\), both of which are used for biotechnological applications. Furthermore, we report on advances for CRISPR/Cas-driven gene knockouts in *Phaeodactylum tricornutum*\(^{11,13}\) and a more efficient bacterial conjugation system for *Thalassiosira pseudonana*\(^{13}\).

Microparticle bombardment was used on *F. cylindrus* (CCMP1102) that was grown, processed and maintained at 4°C in 24h light. Exponential phase cells were harvested onto a 1.2 µm membrane filter that was then placed on an 1.5% agar Aquil plate for bombardment with beads coated with a plasmid containing zeocin resistance and eGFP, both controlled by an endogenous fucoxanthin chlorophyll a/c binding protein (FCP) promoter and terminator (Table 1, Supplementary Table 3 and Methods)\(^3\). Transformation was performed using 0.7 µm tungsten particles and the biolistic particle delivery system PDS-1000/He (Bio-Rad). Rupture disks for
Table 1 | Parameters used for successful transformation as shown in Figs. 2, 3 and 4

| Species             | Transformation method/device | Cell number (input) | Vector amount (µg) | Promoter | Regulatory elements | Drug selection (µg/ml) | Time selection (d) | Efficiency (%) | Transformation status (stable, S; transient, T) | Reporter evidence of transformation | protocols.io link                                                                 |
|---------------------|------------------------------|--------------------|-------------------|-----------|--------------------|-----------------------|----------------------|----------------|-----------------------------------------------|----------------------------------|----------------------------------------------------------------------------------|
| **Archaeoplastids** |                              |                    |                   |           |                    |                       |                      |                |                                               |                                   |                                                                                 |
| *Ostreococcus lucimarinus* (RCC802) | Electroporation Genepulser II | 1–2 × 10^9         | Plasmid PotLuc; linear; 5 | HAPT, Histone H4 | None               | G418 (1,000)          | 10–21               | <0.0001       | S                                             | G418 resist, luminescence, PCR | https://www.protocols.io/view/selection-of-stable-transformants-in-ostreococcus-zg214ge; https://www.protocols.io/view/transient-luciferase-expression-in-ostreococcus-ot-hcib2ue; https://www.protocols.io/view/transient-transformation-of-ostreococcus-species-og86bze |
| *Ostreococcus tauri* | None                         | G418 (1,000)       |                  |           |                    |                       |                      |                |                                               |                                   |                                                                                 |
| *Micromonas commoda* (CCMP2709) | Electroporation Lonza-Nucleofector | 3 × 10^7           | Fusion PCR; pHAPT: pHAPT; linear; 5 | HAPT, Histone H4 | Endogenous, ribosomal protein 59; Endogenous, ribosomal protein 59; | NA                    | 2–6                 | 5.6 ± 1.3 (of post-transformation population) | T                   | eGFP                              | Per cell eGFP fluorescence, fluorescence microscopy | http://doi.org/10.17504/protocols.io.8p9hvr6 |
| *Tetraselmis striata* (KAS-836) | Bio-Rad Biolistics PD5-1000/He biolistics system | 2.0 × 10^7         | pACTpro:Ble; linear; 1.0 | Actin, T. striata | Actin, T. striata | Zeocin (150)          | 21-28               | S                                          | Zeocin resist, PCR             | http://doi.org/10.17504/protocols.io.hjtb4nm |
| **Haptophytes**    |                              |                    |                   |           |                    |                       |                      |                |                                               |                                   |                                                                                 |
| *Isochrysis galbana* (CCMP1323) | Biolistics PDS-1000/He | 1–2 × 10^10        | plgNAT; circular; 1.0 | Hsp70 E. huxleyi | Heterologous         | Nourseothricin (80–150) | 14                  | <0.0001       | S                                             | Nourseothricin resistance, PCR, RT-PCR | https://www.protocols.io/view/biolistic-transformation-of-isochrysis-galbana-2p8qdnw; https://www.protocols.io/view/method-for-electroporation-of-isochrysis-galbana-c-hmab4ze |

Continued
### Table 1 | Parameters used for successful transformation as shown in Figs. 2, 3 and 4 (continued)

| Species | Transformation | Reporter elements | Drug amount | Selection | Time selection | Efficiency (%) | Transformation status (stable, s; transient, t) | Reporter evidence of transformation | protocols.io link |
|---------|---------------|-------------------|-------------|-----------|---------------|---------------|-----------------------------------------------|-----------------------------------|------------------|
| Rhiwariales | Electroporation | GFP, mCherry | 0.5 × 10⁷ | Zeocin (100) | 24 h | 7×10⁶ | S | GFP Fluorescence, vector targeted PCR | http://doi.org/10.17504/protocols.io.h3nb8me |
| Stramenopiles | Bio-Rad | Gene Pulser (165-2076) NEPA21 | 1×10⁸ | Zeocin (100) | 5–7 | 44 per µg of DNA | S | eGFP, shble Zeocin resist, PCR, Southern, fluorescence | http://doi.org/10.17504/protocols.io.h3nb8me |
| Nannochloropsis oceanica | Electroperoration | pMOD, linear/circular; 0.1–1 CMV | 1×10⁹ | None | 0.1–1 | 20 (linear), 1–2 (circular) | T | mT-agBFP2 Fluorescence, PCR, RT–PCR | http://doi.org/10.17504/protocols.io.h3nb8me |
| Phaeodactylum tricornutum | Bacterial conjugation | Pf Cas9-2A-shble PtpBR episome 100 µl E. coli OD600 = 0.9 | 10-16 | Phleomycin (50) | 1.25×10⁻⁵ | ~ 500 c.f.u. | S | shble (Cas9) yfp VENUS Phleomycin resistance, PCR maintained episome, PCR Cas9 target site | http://doi.org/10.17504/protocols.io.4bmgsk6; http://doi.org/10.17504/protocols.io.7gihjue; http://doi.org/10.17504/protocols.io.7gnhjve |
| Aureolochrymm | Bio-Rad Gene Pulser (165-2076) NEPA21 | Glass beads abrasion (425–600 µm) | 5–7×10⁷ | None, other than contained in promoter/term | Manual selection | 24 h | 7×10⁻⁶ ≈ 500 | GFP, mCherry Fluorescence, sequencing, PCR, Western blot | https://www.protocols.io/view/oyster-parasite-perkinsus-marinus-transformation-ugv9bw96; https://www.protocols.io/view/glass-beads-based-transformation-protocol-for-perk-g36byre; https://www.protocols.io/view/fluorescence-activated-cell-sorting-facs-of-perkin-hh2b38e |

Continued
| Species                          | Transformation method/device | Cell number (input) | Vector amount (µg) | Promotor | Regulatory elements | Drug selection (µg/ml) | Time selection (d) | Efficiency (%) | Transformation status (stable, S; transient, T) | Reporter evidence of transformation | Evidence of transformation | protocols.io link                                                                 |
|---------------------------------|------------------------------|---------------------|-------------------|----------|--------------------|------------------------|----------------------|------------------|-----------------------------------------------|---------------------------------|-----------------------------|---------------------------------------------------------------------------------|
| Oxyrrhis marina                 | Electroporation Gene Pulser Xcell; Chemical (CaCl₂) | 1–5×10⁶; 1–10⁶     | Fluorescently labeled DNA (5–25µg) or FITC-dextran; mCherry | NA       | NA, endogenous hsp90 | NA                    | NA                   | 0.5–5%          | T                                           | mCherry                         | Fluorescence                 | https://www.protocols.io/view/electroporation-of-oxyrrhis-marina-vcne2ve; https://www.protocols.io/view/transfection-of-alexa488-labelled-dna-into-oxyrrhis-haB2hwc; https://www.protocols.io/view/electroporation-transformation-of-fitc-dextran-transient-3cmgju6; https://www.protocols.io/view/co-incubation-protocol-for-transforming-heterotrophic-zmb476 |
| Cryptothecodinium cohnii (CCMP 316) | Electroporation LDNZA Nucleofector | 2.5×10⁹            | pAmpAtpBChl; linear; 0.5 | Endogenous | Endogenous Chloramphenicol (20); 3 onward | NA                    | NA                   | <0.001          | T                                           | Fluorescence                 | Fluorescence                 | https://doi.org/10.17504/protocols.io.4r2gv8e |
| Amphidinium carterae (chloroplast) (CCMP3144) | Bio-Rad Biolistics PDS-1000/He biolistics system | 4×10⁶              | Linear-Din3l-nea; linear; 2 | Endogenous | Endogenous Kanamycin (150); 7 | 0.0005               | S         | (3 mon)          | NA                                           | RT-PCR, PCR phenotype          | http://doi.org/10.17504/protocols.io.a4zgv8e |
| Karlodinium verweii (CCMP1975)   | Electroporation               | 4×10⁶              | Linear-Din3l-nea; linear; 2 | Endogenous | Endogenous Kanamycin (150); 7 | 0.0005               | S         | (3 mon)          | NA                                           | RT-PCR, PCR phenotype          | http://doi.org/10.17504/protocols.io.a4zgv8e |
| Discobas (euglenozoans and heteroloboseans) | Bodo saltans (submitted to ATCC) | 1–1.5×10⁷          | Linear-Din3l-nea; linear; 3 | Endogenous | Endogenous G418 (5); 7–9 | 7–9                   | S         | N/A               | NA                                           | GFP                           | Fluorescence, PCR, RT-PCR       | http://doi.org/10.17504/protocols.io.4d4gs4e |
| Diplonema papillatum (ATCC 50162) | Electroporation Lonza-Nucleofector | 5×10⁷              | p57-V5±Neo; linear; 3 | Endogenous | Endogenous G418 (75); 7-14 | -5.5                  | S         | NA               | Western blot (resistance marker), RT-PCR    | Western blot (resistance marker), fluoresce | http://doi.org/10.17504/protocols.io.a4zgv8e |
| Naegleria gruberi               | Electroporation Lonza-Nucleofector | 4×10⁷              | pNAEG-HYG; circular; 4 | Endogenous | Endogenous Hygromycin (300) Neo (700); 15-28 | 80                   | T         | GFP              | Western blot (resistance marker), fluoresce | RT-PCR, PCR phenotype          | http://doi.org/10.17504/protocols.io.a4zgv8e |
| Opisthokonts                    | Acanthamoeba chimera (ATCC PRA-279) | 3×10⁷              | Awhls_H2Bvenus+pUC19; circular; 1-5±4.0 carrier | Endogenous | Endogenous NA | 10–15 10–15 | 1          | T                                           | Venus                          | Fluorescence, RT-PCR            | http://doi.org/10.17504/protocols.io.zez1sfn |
| Salpingoeca rosetta             | Electroporation Lonza-Nucleofector | 4×10⁷              | SrActmCherry-CCTLL+pUC19; circular; 1-10+40 carrier | Endogenous | Endogenous Puromycin (80); 10-12 | S         | mCherry            | Gene expression (Luc, fluorescence)/resistance | https://doi.org/10.17504/protocols.io.68b99w |

For additional information, see protocols.io links, Supplementary Table 5 and Supplementary Note 1. For contacting laboratories working with particular species, see details given in Supplementary Table 6. *May be stable but overgrown by wild-type strain. NA, not applicable; NLS, nuclear localization signal.
used for Schizochytrium\textsuperscript{18}. The highest transformation efficiency was achieved using 1 µg of linearized 18GZG plasmid with two pulses, resulting in a time constant of ~5 ms (Table 1 and Methods). Expression of the fusion protein was confirmed by both the zeocin-resistance phenotype and the detection of eGFP (Fig. 2). Six 18GZG transformants derived from uncut and linearized plasmids were examined in detail. All maintained antibiotic resistance throughout 13 serial transfers, first in selective, then subsequently in non-selective media and then again in selective medium. Integration of the plasmid into the genome was confirmed by PCR as well as by Southern blots using a digoxigenin-labeled ShBle gene probe, showing that four transformants had integrations by single homologous recombination, while in two transformants additional copies of the antibiotic resistance cassette were integrated by nonhomologous recombination elsewhere in the genome (Supplementary Fig. 5).

Electroporation of 	extit{N. oceanica} (CCMP1779) was optimized based on observation of cells treated with fluorescein-conjugated 2,000 kDa dextran and subsequent survival (Table 1 and Methods). A sorbitol concentration of 800 mM and electroporation at between 5 and 9 kV cm\textsuperscript{-1} resulted in highest cell recovery. These conditions were used during introduction of plasmids containing the gene for the blue fluorescent reporter mTagBFP2 under the control of cytomegalovirus (CMV), the cauliflower mosaic virus 35S, or the VCPI promoter previously described from 	extit{Nannochloropsis} sp\textsuperscript{1}. Transient expression of blue fluorescence (compared to cells electroporated simultaneously under the same conditions without plasmid) appeared within 2 h, lasted for at least 24 h and disappeared by 48 h in subsets of cells electroporated with mTagBFP2 under the control of CMV (Supplementary Fig. 6). The transient transformation was more effective when a linearized plasmid was used compared to a circular plasmid (Table 1). VCPI did not induce blue fluorescence with a circular plasmid, while 35S gave inconsistent results with either circularized or linearized plasmids.

For 	extit{P. tricornutum} (CCAP1055/1), we adapted the CRISPR/Cas9 system\textsuperscript{17} for multiplexed targeted mutagenesis. Bacterial conjugation\textsuperscript{1} was used to deliver an epimer that contained a Cas9 cassette and two single-guide RNA (sgRNA) expression cassettes designed to excise a 38 basepair-long domain from the coding region of a nuclear-encoded, chloroplastic glutamate synthase (\textit{Phatr3_J24739}) and introduce an in-frame stop codon after strand ligation (Table 1 and Methods). The GoldenGate assembly was used to clone two expression cassettes carrying sgRNAs into a \textit{P. tricornutum} episome that contained a Cas9–2A-ShBle expression cassette and the centromeric region CenArsHIs (Supplementary Fig. 7). After their addition to a \textit{P. tricornutum} culture, plates were incubated in a growth chamber under standard growth conditions for 2 d and transformed \textit{P. tricornutum} colonies began to appear after 2 weeks. Only colonies maintaining Cas9–2A-ShBle sequence on the delivered episome were able to grow on selection plates because Cas9 and ShBle were transcriptionally fused by the 2A peptide\textsuperscript{18} (Supplementary Fig. 7). Gel electrophoresis migration and sequencing of the genomic target loci confirmed the 38-bp-long excision and premature stop codon (Fig. 4d).

\textbf{Alveolates.} This species-rich and diverse group comprises ciliates, apicomplexans and dinoflagellates (Fig. 1). As a link between apicomplexan parasites and dinoflagellate algae, perkinsids are key for understanding the evolution of parasitism, and also have potential biomedical applications\textsuperscript{17}. Techniques currently exist for transformation of only a small number of ciliates, perkinsids and apicomplexans\textsuperscript{16}. Here, we present a transformation protocol for \textit{Karlodinium veneficum} (CCMP1975), a phagotrophic microtroph that produces fish-killing karlotoxins\textsuperscript{46}. Experiments were also performed on \textit{Oxyrrhis marina} (CCMP 1788/CCMP 1795), a basal-branching phagotroph that lacks photosynthetic plastids and \textit{Cryptecodium collinii} (CCMP 316), a heterotroph used in food supplements. For both of these taxa, evidence of DNA delivery
ResouRce NATure MeTHODs

was synthesized using random hexamers. Out of 16 transforma-
3 d postelectroporation. Fresh seawater with kanamy-

Fig. 8 and Methods). (see Supplementary Fig. 15c for a trace of these
different regions in the cell). Scale bars are as follows: 10 μm for A. amoebiformis, T. pseudonana, A. limacinum, B. saltans, N. gruberi, A. whisleri and S. rosetta; 15 μm for P. marinus; 20 μm for F. cylindrus and 100 μm for P. multiseries.

was achieved (Table 1, Supplementary Results, Supplementary
Fig. 15 and Methods), a goal recently achieved for C. cohnii using

dmined dense cultures after

electroporation19. Additionally, we report on improved transforma-
tion systems for Perkinsus marinus (PRA240) and Amphidinium
carterae (CCMP1314) chloroplast, published recently as part of the
EMS initiative15.

K. veneficum (CCMP1975) was transformed based on electro-
poration and cloning the selectable marker gene aminoglycoside
3′-phosphotransferase (nptII/neo; note that nptII/neo is used synon-
ymously with amino 3′-glycosyl phosphotransferase gene conferring
resistance to kanamycin, neomycin, paromomycin, ribostamycin,
butirosin and gentamicin B) into the backbone of the dinoflagellate-
specific expression vector DinoIII-neo41, which confers resistance
to neomycin and kanamycin (Table 1). In brief, DinoIII-neo was
linearized and electroporated using the Nucleofector optimization
pulse codes, buffer SF/Solution I (Lonza), and 2 μg μl−1 of linearized
DinoIII-neo. Electroporated cells were selected under 150 μg ml−1
kanamycin 3 d postelectroporation. Fresh seawater with kanamy-
cin was added every 2 weeks to the cultures and new subcultures
were inoculated monthly. After 3 months, DNA and RNA were iso-
lated from the resistant cultures as previously reported42 and cDNA
was synthesized using random hexamers. Out of 16 transforma-
tions, two cell lines (CA-137, DS-138) showed stable growth under
kanamycin selection. CA-137 developed dense cultures after
3 months, and the resistance gene was detected in both DNA and RNA
by nested PCR and RT–PCR, respectively (Fig. 4e, Supplementary
Fig. 8 and Methods).

We improved the transformation protocol16,17 of P. marinus, a
pathogen of marine mollusks, fish and amphibians51 (Supplementary
Table 5). We coexpressed two genes and efficiently selected tran-
sient and stable transformants using FACS (Table 1, Figs. 2 and 4f,
Supplementary Fig. 9 and Methods). In addition, we established the
integration profile of ectopic DNA once introduced into the
P. marinus genome. We did not see evidence of integration through
homologous recombination and observed a propensity for plasmid
fragmentation and integration within transposable elements sites.
An optimized alternative protocol for transformation using glass
bead abrasion was also developed. Two versions of the previously
published Moe gene promoter16 were tested. Whereas the 1.0 kb pro-
moter version induced expression after 2 or 3 d, the truncated ver-
sion (0.5 kb) took 7 d for expression to be detected. Resistance genes
to zeocin, blasticidin and puromycin have all been shown to confer
resistance to transformed P. marinus; however, selection regimes
are still relatively slow and inefficient, indicating further room
for improvement17.

We also report a vector for the transformation of the A. carterae
chloroplast, a photosynthetic dinoflagellate. A. carterae, like other
dinoflagellates with a peridinin-containing chloroplast, contains a
fragmented chloroplast genome made up of multiple plasmid-like
minicircles46. The previous transformation protocols made use of
this to introduce two vectors based on the psbA minicircle47.
Here, we show that other minicircles are also suitable for use as vec-
tors. We created an artificial minicircle, using the atpB minicircle
as a backbone, but replacing the atpB gene with a codon-optimized
chloramphenicol acetyltransferase (Table 1 and Methods). This circular vector was introduced by biolistics to *A. carterae* (Supplementary Fig. 10a). Following selection with chloramphenicol, we were able to detect transcription of the chloramphenicol acetyltransferase gene via RT–PCR (Fig. 4g). This result suggests that all 20 or so minicircles in the dinoflagellate chloroplast genome were suitable for use as artificial minicircles, thus providing a large pool of potential vectors.

**Discobans.** This diverse group, recently split into Discoba and Metamonada, includes heterotrophs, photoautotrophs and predatory mixotrophs, as well as parasites. Discobans include parasitic kinetoplastids with clinical significance, such as *Trypanosoma brucei*, *T. cruzi* and *Leishmania* spp., for which efficient transformation protocols are available. However, such protocols are missing for aquatic species. Here, we describe available transformation protocols for the kinetoplastid *Bodo saltans* and the heterolobosean *Naegleria gruberi*. The former was isolated from a lake, but identical 18S rRNA gene sequences have been reported from marine environments. The latter is a freshwater protist that represents a model organism for closely related marine heterolobosean amoebas. Furthermore, we provide advanced methods that build on previous EMS results for the diplonemid *Diplonema papillatum*.

*B. saltans* (ATCC 30904) was transformed with a plasmid containing a cassette designed to fuse an endogenous *EF1-α* gene with eGFP for C-terminal tagging. This cassette includes downstream of eGFP, a *B. saltans* tubulin intergenic region followed by the selectable marker *nptII/neo* gene, conferring resistance to neomycin. *EF1-α* genes exist in tandem repeats. The homologous regions that flank the cassette were chosen as targets for inducing homology-directed repair; however, they target only one copy of the gene. As transcription in *B. saltans* is polycistronic, insertion of the tubulin intergenic region into the plasmid is essential for polyadenylation of the *EF1-α/GFP* fusion and *trans*-splicing of the *nptII/neo* gene (Supplementary Table 5). Selection of transfected cells began with 2 μg ml⁻¹ of neomycin added 24 h after electroporation, and this concentration was gradually increased over 2 weeks to 5 μg ml⁻¹ (Table 1 and Methods). Cells were washed and subcultured into fresh selection medium every 4 d, and neomycin-resistant cells emerged 7–9 d postelectroporation. The eGFP signal was detected 2 d postelectroporation, albeit with low intensity. This may be due to the inefficient translation of eGFP since it has not been codon-optimized for *B. saltans* (Fig. 2). Genotyping analysis 9 months posttransfection confirmed the presence of the *nptII/neo* gene and at least partial plasmid sequence (Fig. 4h and Supplementary Fig. 10b). However, plasmid integration into the *B. saltans* genome through
homologous recombination is still unconfirmed. This suggests either off-target plasmid integration or episomal maintenance.

For *N. gruberi* (ATCC 30224) two plasmids were designed. The first one carried the hygromycin B resistance gene (*hph*) with an actin promoter and terminator, along with an HA-tagged *hph*. The second plasmid carried the hygromycin B resistance gene (*hph*) instead. For each individual circular plasmid, 4 μg of a plasmid was electroporated (Table 1 and Methods). About 48 h after electroporation, dead cells were removed from the suspension and viable cells were washed with PBS. Afterward, 300 μg ml⁻¹ of hygromycin B or 700 μg ml⁻¹ of neomycin was added to the fresh media. One to 4 weeks later, resistant clones were recovered and expression of eGFP and/or hygromycin was confirmed by western blotting (Supplementary Fig. 11). Expression of eGFP was observed by epifluorescence microscopy (Fig. 2 and Supplementary Fig. 11) with ~80% of transformants maintaining hygromycin B or neomycin resistance in addition to expressing eGFP.

*D. papillatum* (ATCC 50162) was transformed by electroporation using 3 μg of a *Swal*-linearized fragment (cut from p57-*V5*+NeoR plasmid) containing the V5-tagged *nptII/neo* gene flanked by partial regulatory sequences derived from the hexokinase gene of theinetoplastid *Blastocrithidia* (strain p57) (Table 1 and Methods) using a published protocol. About 18 h after electroporation, 75 μg ml⁻¹ G418 was added to the medium and after 2 weeks, seven neomycin-resistant clones were recovered. Transcription of *nptII/neo* was verified in four clones by RT–PCR (Supplementary Fig. 12) and the expression of the tagged *nptII/neo* protein was confirmed in two clones by western blotting using the α-V5 antibody (Fig. 4i).

**Opisthokonts.** The opisthokont clade Holozoa includes animals and their closest unicellular relatives choanoflagellates, filastereans, ichthyosporeans and corallochytreans. The establishment of genetic tools in nonmetazoan holozoans promises to help illuminate the cellular and genetic foundations of animal multicellularity. Genomic and transcriptomic data are available for multiple representatives characterized by diverse cell morphologies, some of which can even form multicellular structures. Until recently, only transient transformations had been achieved for some opistokonts such as the filasterean *Capsaspora owczarzak*, the ichthyosporean *Creolimax fragrantissima* and the choanoflagellate *Salpingoeca rosetta*. Through the EMS initiative, we report on evidence for transient transformation of the ichthyosporean *Abeoforma whisleri*, isolated from the digestive tract of mussels, and review a recently published stable transformation protocol for *S. rosetta* achieved by using the selectable puromycin N-acetyl-transferase gene (Fig. 2).
Fig. 5 | 'Transformation roadmap' for the creation of genetically tractable protists. a, Vector design and construction for microeukaryotes of interest and a natural community. b, Transformation approaches. Different symbols represent methods (for example chemical, physical or biological) for introducing DNA/RNA/protein into a living cell. c, Protocol. Key methodological steps for successful transformation are listed in an abbreviated form (for particular examples, see Table 1 and text).
plasmid (empty pUC19) and 1–5 μg of the reporter plasmid (A. whisleri H2B fused to mVenus fluorescent protein, mVFP) (Table 1 and Methods), and subjected to code EN-138 (Lonza). Immediately after the pulse, cells were recovered by adding 80 μl of marine broth (Gibco) before plating in 12-well culture plates previously filled with 1 ml marine broth. After 24h, ~1% of the culture was transformed based on the fraction of cells expressing mVFP in the nucleus (Figs. 2 and 4).

**Microbial eukaryotes in natural planktonic communities.** Model organisms are typically selected based on criteria such as relative ease of isolation and asexual cultivation in the laboratory; however, these attributes may not correlate with the capacity for uptake and expression of the exogenous DNA. We explored whether natural marine planktonic pico- and nanoeukaryote communities would take up DNA in a culture-independent setting. Microplankton from natural seawater was concentrated and electroporated with plasmids containing mTagBFP2 under the control of CMV or 3SS promoters (Supplementary Results and Methods). In most trials, blue fluorescent cells were rare if detected at all (compared to control samples). However, in one natural community tested, a photosynthetic picoeukaryote population exhibited up to 50% of cells with transient expression of blue fluorescence when the CMV promoter was used (Supplementary Fig. 13). This suggests it might be possible to selectively culture eukaryotic microorganisms based on capacity to express exogenous DNA.

**Discussion**

The collaborative effort by the EMS initiative facilitated identification and optimization of the steps required to create new protist model systems, which culminated in the synthetic transformation roadmap (Fig. 5). Our genetic manipulation systems for aquatic (largely marine) protists will enable deeper insights into their cell biology, with potentially valuable outcomes for aquatic sciences, evolutionary studies, nanotechnology, biotechnology, medicine and pharmacology. Successes and failures with selectable markers, transformation conditions and reporters were qualitatively compared across species (Supplementary Tables 3 and 4; Table 1, Figs. 2–4 and Methods).

For some of the selected species, the first step was to identify cultivation conditions for robust growth in the laboratory to either generate high cell densities or large culture volumes for obtaining sufficient biomass required for a variety of molecular biology experiments. Unlike established microbial model species, cultivation of marine protists can be challenging, especially under axenic conditions or for predatory taxa that require cocultivation with their prey. Nevertheless, 13 out of 35 species were rendered axenic before the development of transformation protocols. For the remaining species, we were unable to remove bacteria and therefore had to make sure that transformation signals were coming from the targeted protist rather than contaminants (Supplementary Table 2). Subsequent steps included the identification of suitable antibiotics and their corresponding selectable markers (Table 1 and Supplementary Table 3), conditions for introducing exogenous DNA (Table 1 and Supplementary Table 4) and selection of promoter and terminator sequences for designing transformation vectors (Table 1, Methods, Supplementary Table 5 and Supplementary Notes 1).

As exemplified in the model systems provided herein (Table 1 and Figs. 2–4), a variety of methods were used to test whether exogenous DNA was integrated into the genome or maintained as a plasmid, and whether the introduced genes were expressed. Approaches to show the former included inverse PCR, Southern blotting and whole genome sequencing, whereas approaches to demonstrate the latter included various combinations of PCR, RT–PCR, western blotting, epifluorescence microscopy, FACS, antibody-based methods and/or growth assays in the presence of antibiotics to confirm transcription and translation of introduced selection and reporter genes (for example, eGFP, YFP, mCherry). For fluorescent markers, it was first ensured that the wild-type, or manipulated controls, had no signals conflicting with the marker (Figs. 2 and 3c), an important step because photosynthetic protists contain chlorophyll and other autofluorescent pigments. Overall transformation outcomes for each species were parsed into three groups according to the level of success or lack thereof (A, first transformation protocol for a given species; B, advanced protocol based on previous work and C, published protocol based on the EMS initiative) and are discussed according to their phylogenetic position (Fig. 1).

Our studies did not result in a universally applicable protocol because transformability and a range of other key conditions varied greatly across taxa and approaches, such as intrinsic features of the genome and differences in cellular structure and morphology. In general, electroporation proved to be the most common method for introducing exogenous DNA stably into cells. This approach was used for naked cells and protoplasts, yet frequently also worked, albeit with lower efficiency, on cells protected by cell walls. Linearized plasmids were most effective for delivery, and 5’ and 3’ UTR-containing promoters of highly expressed endogenous genes provided the strongest expression of selective reporters and markers. If successful, teams usually continued with fluorescence-based methods. Furthermore, large amounts of carrier DNA usually facilitated successful initial transformations (for example, M. commoda, A. whisleri) or improved existing protocols (S. rosetta). We also provide the contact details of all coauthors who are assigned to particular species (Supplementary Table 6).

Some lineages were difficult to transform, especially dinoflagellates and coccolithophores. Here, even if DNA appeared to be delivered (Supplementary Table 5), expression of the transformed genes could not be confirmed. Examples include the dinoflagellates C. cohnii, Symbiodinium microadriaticum and the coccolithophore E. huxleyi. Thus, at least these three species need concerted future efforts.

The combination of results presented herein together with previously published protocols from the EMS initiative significantly expands the segment of extant eukaryotic diversity amenable to reverse genetics approaches. Out of the 39 microbial eukaryotes selected for the initiative, exogenous DNA was delivered and expressed in more than 50% of them. The transformation systems enable us to shed light on the function of species-specific genes, which likely reflect key adaptations to specific niches in dynamic ocean habitats.

**Online content**

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41592-020-0796-x.

Received: 26 June 2019; Accepted: 2 March 2020; Published online: 6 April 2020

**References**

1. Worden, A. Z. et al. Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science* **347**, 1257594 (2015).
2. de Vargas, C. et al. Eukaryotic plankton diversity in the sunlit global ocean. *Science* **348**, 1261605 (2015).
3. Collier, J. L. & Rest, J. S. Swimming, gliding, and rolling toward the mainstream: cell biology of marine protists. *Mol. Biol. Cell.* **30**, 1245–1248 (2019).
4. Curtis, B. A. et al. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* **492**, 59–65 (2012).
5. Armbrust, E. V. et al. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**, 79–86 (2004).
6. Read, B. A. et al. Pan genome of the phytoplankton *Emiliania* underpins its global distribution. *Nature* **499**, 209–213 (2013).
7. Keeling, P. J. et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol. 12, e1001889 (2014).

8. Nymark, M., Sharma, A. K., Sparstad, T., Bones, A. M. & Winge, P. A. CRISPR/Cas9 system adapted for gene editing in marine algae. Sci. Rep. 6, 24951 (2016).

9. Hopes, A., Nebrasov, V., Kamoun, S. & Mock, T. Editing of the urease gene by CRISPR-Cas in the diatom Thalassiosira pseudonana. Plant Methods 12, 49 (2016).

10. Carradec, Q. et al. A global atlas of eukaryotic genes. Nat. Commun. 9, 373 (2018).

11. Apt, K. E., Kroth-Pancic, P. G. & Grossman, A. R. Stable nuclear transformation of the diatom Phaeodactylum tricornutum. Mol. Genet. Genomics. 282, 579–596 (2014).

12. Paschke, P. et al. Genetic engineering of Dictyostelium discoideum based on selection and growth on bacteria. J. Vis. Exp. 143, e58981 (2019).

13. Karas, B. J. et al. Designer diatom episomes delivered by bacterial conjugation. Nat. Commun. 6, 6925 (2015).

14. Hirakawa, Y., Kofuji, R. & Ishida, K. Transient transformation of a chloroplast-nanoalgae chimeric protoplast. N. J. Gen. Genomics. 35, 45–53 (2008).

15. Sakamoto, H. et al. Puromycin selection for stable transfectants of the oyster-infecting parasite Perkinsus marinus. Parasitol. Int. 69, 13–16 (2018).

16. Kaur, B. et al. Transformation of the picoeukaryote Isochrysis galbana e4074 (2012).

17. Diao, J., Song, X., Zhang, X., Chen, L. & Zhang, W. Genetic engineering of Cryptothecodiun cohnii to increase growth and lipid accumulation. Front. Microbiol. 9, 492 (2018).

18. Sakaguchi, K. et al. Versatile transformation system that is applicable to both multiple transgene expression and gene targeting for thraustochytrids. Appl. Environ. Microbiol. 78, 3193–3202 (2012).

19. Booth, D., Middleton, H. & King, N. Chlorellaeflagellate transformation illuminates their cell biology and the ancestry of animal sepsis. Mol. Biol. Cell. 29, 3026–3038 (2018).

20. Wetzel, L. A. et al. Predicted glycosyltransferases promote development and prevent spurious cell clumping in the chlorellaeflagellate S. rouxii. eLife 7, e41482 (2018).

21. van Baren, M. J. et al. Efficient gene targeting and removal of foreign DNA by homologous recombination in the picoeukaryote Ostreococcus tauri. BMC Genomics 1–22 (2016).

22. Demers, I., Carbonell, P., Pino, S. & Farías, A. Variation of fatty acids in Isochrysis galbana (T-Iso) and Tetraselmis suecica, cultured under different nitrate availabilities. J. Fish. Aquacult. 5, 1–3 (2014).

23. Dunahay, T. G., Jarvis, E. E. & Roessler, P. G. Genetic transformation of the picoeukaryote Ostreococcus tauri. J. Vis. Exp. 65, e4074 (2012).

24. Howde, B. T. et al. Genome sequence and transcriptome analyses of the bloom-forming diatom Fragilaropsis cylindrus. Nature 541, 536–540 (2017).

25. Princeton, T. et al. Evolutionary genomics of the cold-adapted diatom Fragilaropsis cylindrus. Nature 541, 536–540 (2017).

26. Bronson, J. K. et al. Biosynthesis of the neurototoxic domoic acid in a bloom-forming diatom. Science 361, 1356–1358 (2018).

27. Kroth, P. G. Genetic transformation: a tool to study protein targeting in diatoms. Methods Mol. Biol. 1090, 257–267 (2014).

28. Sabatino, V. et al. Establishment of genetic transformation in the sexually reproducing diatoms Pseudo-nitzschia multiseries and Pseudo-nitzschia arenisculus and inheritance of the transgene. Mar. Biotechn. 17, 452–462 (2015).

29. Oro, N., Aki, T. & Kawai, S. Method for introducing a gene into a marine alga. US patent 7,888,123 (2013).

30. Kilian, O., Benemann, C. S., Niyogi, K. K. & Vick, B. High-efficiency homologous recombination in the oil-producing alga Nanochloropsis sp. Proc. Natl Acad. Sci. USA 108, 21265–21269 (2011).

31. Duda, K. et al. High-efficiency genome editing via 2A-coupled co-expression of fluorescent proteins and zinc finger nucleases of CRISPR/Cas9 nicsa pairs. Nucl. Acids Res. 42, e84 (2014).

32. Donald, R. G. & Roos, D. S. Stable molecular transformation of Toxoplasma gondii: a selectable dihydrofolate reductase-thymidylate synthase marker based on drug-resistance mutations in malaria. Proc. Natl Acad. Sci. USA 90, 11703–11707 (1993).

33. Barbrook, A. C., Howe, C. J. & Nisbet, R. E. R. Breaking up is hard to do: the complexity of the diatom flagellate chloroplast genome. Perspect. Photol. 6, 31–37 (2019).

34. Sprecher, B. N., Zhang, H. & Lin, S. Nuclear gene transformation in the diatom Oxyrrhis marina. Microorganisms 8, 126 (2020).

35. Zhang, H. et al. Signal recognition particle RNA in diatom flagellates and the Perkinsus marinus. Pratit 164, 74–761 (2013).

36. Chambouvet, A. et al. Cryptic infection of a broad taxonomic and geographic diversity of tadpoles by Perkinsnea protists. Proc. Natl Acad. Sci. USA 112, E4743–E4751 (2015).

37. Adl, S. M. et al. Revision to the classification, nomenclature and diversity of eukaryotes. J. Euk. Microbiol. 66, 4–119 (2019).

38. Matthews, K. R. 25 years of African trypanosome research: from description and 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 and your intended use is not 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) 2020

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 is 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) 2020

Drahomíra Faktorová1,58, R. Ellen R. Nisbet2,54,58, José A. Fernández Robledo3,58, Elena Casacuberta4,58, Lisa Sudek5,58, Andrew E. Allen6,7, Manuel Ares Jr.8, Cristina Aresté4, Cecilia Balestrieri9, Adrian C. Barbrook2, Patrick Beardslee10, Sara Bender11, David S. Booth12, François-Yves Bouget13, Chris Bowler14, Susana A. Breglia15, Colin Bronzee9, Gertraud Burger16, Heriberto Cerutti17, Rachele Cesarioni17, Miguel A. Chiuriollo18, Thomas Clemente10, Duncan B. Coles3, Jackie L. Collier19, Elizabeth C. Cooney20, Kathryn Coyne21, Roberto Docampo18,
Methods

Studied species and used transformation methods. For the full list of vector sequences and maps see Supplementary Notes 1 and for detailed description of Figs. 3 and 4 see Supplementary Note 2. Antibiotic concentrations effective for selection of transformants can be found in Supplementary Table 3, the details of the transformation methods applied to this study in Supplementary Table 4 and contact details for individual laboratories in Supplementary Table 6. Full list of protists (including details of culture collection) and links to the complete step-by-step transformation protocols and published vector sequences are listed in Supplementary Table 5. The protocols.io links listed in Table 1 and Supplementary Table 5 are summarized in Supplementary Tables 7 and 8.

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

Data availability

The data that support the findings of this study are available from the corresponding authors as well as the other authors upon request (for the contacts see Supplementary Table 6). Source data for Figs. 3 and 4 and Supplementary Figs. 9b,c, 11a and 12b,c are available online.

Acknowledgements

We thank M. Salisbury and D. Lacono, C. Poirier, M. Hamilton, C. Eckmann, H. Igel, C. Yung and K. Hoadley for assistance; V.K. Nagarajan, M. Acerbi and P.J. Green who carried out Agrobacterium studies in Heterosigma akashiwo, and N. Kraeva, C. Bianchi and V. Yurchenko for the help with designing the p57-V5+NeoR construct. We are also grateful to the protocols.io team (L. Teytelman and A. Broellochs) for their support. This collaborative effort was supported by the Gordon and Betty Moore Foundation EMS Program of the Marine Microbiology Initiative (grant nos. GBMF4972 and 4972.01 to F.Y.B.; GBMF4970 and 4970.01 to M.A. and A.Z.W.; GBMF3788 to A.Z.W.; GBMF4978 to A.Z.W.; GBMF4968 and 4968.01 to H.C.; GBMF4984 to V.H.; GBMF4974 and 4974.01 to C. Brownlee; GBMF4964 to Y. Hirakawa; GBMF4961 to T. Mock; GBMF4958 to P.S.; GBMF4957 to A.T.; GBMF4960 to G.J.S.; GBMF4979 to K.C.; GBMF4982 and 4982.01 to J.L.C.; GBMF4964 to P.J.K.; GBMF4981 to P.v.D.; GBMF5006 to A.E.A.; GBMF4986 to C.M.; GBMF4962 to J.A.F.R.; GBMF4980 and 4980.01 to S.L.; GBMF4977 and 4977.01 to R.F.W.; GBMF4962.01 to C.H.S.; GBMF4985 to J.M.; GBMF4976 and 4976.01 to C.H.; GBMF4963 and 4963.01 to V.E.; GBMF5007 to C.L.D.; GBMF4983 and 4983.01 to J.L.; GBMF4975 and 4975.01 to A.D.T.; GBMF4973 and 4973.01 to I.R.-T. and GBMF4965 to N.K.), by The Leverhulme Trust (RPG-2017-364) to T. Mock and A. Hopes, and by ERD funds (16_019/0000759) from the Czech Ministry of Education to J.L.

Author contributions

The project was conceived and designed by A.C.J., J.Z.K., S.B., D.F., J.L., R.E.R.N., J.A.F.R., E.C., L.S., A.Z.W., T. Mock, A.E.A., F.-Y.B, C. Brownlee, C. Bowler, H.C., T.C., J.L.C., K.C., C.L.D., V.E., V.H., Y. Hirakawa, C.J.H., P.J.K., N.K., S.L., C.M., J.M., I.R.-T., P.A.S., C.H.S., G.J.S., A.D.T., P.v.D., A.T. and R.F.W. Data analysis was carried out by M.A., C.A., C. Balestrieri, A.C.B., P.B., D.S.B., S.A.B., G.B., R.C., M.A.C., D.B.C., E.C.C., R.D., E.E., P.A.E., E.F., V.E.-R., N.J.E., K.F., P.A.G., P.R.G., F.C., S.G.G., J.G., Y. Hanawa, E.R.H.-C., E.H., A. Highfield, A. Hopes, I.H. J.L., N.A.T.I., Y.L., N.E.J., A.K., K.E.-K., E.K., I.A.K., N.L., I.L., Z.L., J.-C.L., F.L., S.M., T. Matute, M.M., S.R.N., D.N., I.C.N., L.N., A.M.G.N.V., M.N., N., A. Pain, A. Piersanti, S.P., J.S.R., M.R., A.R., M.A.S., E.C.S., B.N.S., R.S., T.v.d.H., L.T., J.T., M.V., V.V., L.W., X.W., G.W., A.W. and H.Z. The manuscript was written by D.F., R.E.R.N., J.A.F.R., E.C., L.S., T. Mock, A.Z.W. and J.L. with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41592-020-0796-x.

Correspondence and requests for materials should be addressed to D.F., T.M., A.Z.W. or J.L.

Peer review information Lei Tang was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  - Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection
FACS softwares noted in our methods section are FloJo v.10.6.1, WinList 3D v7.0 (Verity Software House) and BD FACS Software v1.2.0.142.

Data analysis
FACS softwares noted in our methods section that were used for data analysis after collection on the FACS instrument are FloJo v.10.6.1 and WinList 3D v7.0 (Verity Software House).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available from the corresponding authors as well as the other authors upon request (for the contacts see Suppl. Table 5).
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | No sample size calculation was performed, rather standard practices in the field of genetics were used which involved replicating experiments the sample sizes and replication are the same or more than in most publications presenting developments in genetic techniques. |
| Data exclusions | No data was excluded from the analyses. |
| Replication | For each experiment, all attempts at replication were successful. |
| Randomization | The experiments were designed to determine whether a specific protocol rendered transformation in cells. Since clonal cell lines and the same starting material were used under pulse conditions/plasmids and negative controls and because these are not survey experiments randomization is not an appropriate aspect of the design |
| Blinding | Blinding was not appropriate to our study as all measurement were performed using the same clonal cell lines and analyzed with identical methodologies (within each experiment). Our experiments did not involve testing placebos nor were they what is termed randomized controlled trials (RCT) as appropriate in e.g. drug clinical trials. |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

| Materials & experimental systems | Methods |
|---------------------------------|---------|
| n/a | n/a |
| ☐ ☒ Antibodies | ☐ ☒ ChiP-seq |
| ☒ ☒ Eukaryotic cell lines | ☒ ☒ Flow cytometry |
| ☐ ☒ Palaeontology | ☒ ☒ MRI-based neuroimaging |
| ☐ ☒ Animals and other organisms | ☐ ☒ |
| ☒ ☒ Human research participants | ☐ ☒ |
| ☒ ☒ Clinical data | ☐ ☒ |

Antibodies

Commercial antibodies:
- Anti-V5 tag Monoclonal Antibody (2F11F7)
  - Supplier: Invitrogen
  - Catalog number: 37-7500
  - LOT number: 1468908A
  - Dilution: 1:1,000

- Monoclonal Anti-α-Tubulin clone DM1A produced in mouse, ascites fluid
  - Supplier: Sigma Aldrich
  - Catalog Number: T9026
  - Lot #: N/A
  - Dilution: 1:2,000

- Anti-GFP Living Colors® A.v. Monoclonal Antibody
  - Supplier: Takara
  - Cat. #: 632380
  - Clone name: JL-8
  - Lot #: N/A
  - Dilution: 1:1,000
| Antibody                              | Supplier                     | Catalog Number | Lot | Dilution |
|---------------------------------------|------------------------------|----------------|-----|----------|
| Anti-GFP polyclonal antibody, Invitrogen | Cat.no. A11122              |                |     | 1:1000   |
| Anti-Histone H3 antibody, Abcam        | Cat.no. ab1791               |                |     | 1:1000   |
| Anti-mCherry antibody, Abcam           | Cat.no. ab167453             |                |     | 1:1000   |
| H9658- Monoclonal Anti-HA antibody     | Supplier: Sigma Aldrich      | A9044          |     | 1:2,000  |
| Anti-Mouse IgG, HRP-Linked Whole Ab    | Supplier: GE Healthcare      | NA931-100UL    |     | 1:10,000 |
| Goat Anti-rabbit HRP, ImmunoReagents   | Cat. no. GtxRb-003-DHRPX     |                |     | 1:10,000 |

**Validation**

Validation of anti-V5 tag Monoclonal Antibody (2F11F7): The manufacturer states that the antibody works on a wide range of species (human, bovine, amphibian and mouse cells or tissues as well as with yeast or fungi). https://www.thermofisher.com/antibody/product/V5-Tag-Antibody-clone-2F11F7-Monoclonal/37-7500

We tested it on D. papillatum and T. brucei.

Validation of Monoclonal Anti-α-Tubulin: Manufacturer’s validation: Monoclonal Anti-α-Tubulin is immunospecific for tubulin as determined by indirect immunofluorescent staining and immunoblotting procedures. The manufacturer states that the antibody works on a wide range of species (human, bovine, amphibian and mouse cells or tissues as well as with yeast or fungi). https://www.sigmaaldrich.com/catalog/product/sigma/t9026?lang=en&region=CZ

We tested it on D. papillatum and T. brucei.

Validation of Anti-GFP Living Colors® A.v. Monoclonal Antibody: Manufacturer’s validation: The quality and performance of Living Colors A.v. Monoclonal Antibody (JL-8) was tested by Western blot analysis using lysate made from a HEK 293 cell line stably expressing AcGFP1.

Validation of Anti-GFP polyclonal antibody (Invitrogen): The manufacturer states that the antibody works on a wide range of species (human, bovine, amphibian and mouse cells or tissues as well as with yeast or fungi). https://www.thermofisher.com/antibody/product/GFP-Tag-Antibody-clone-2F11F7-Monoclonal/37-7500

We tested it on D. papillatum and T. brucei.
species, we tested it on T. gondii and Perkinsus in our lab.

Validation of Anti-Histone H3 antibody (Abcam): validated to work with a wide range of species according to the manufacturer, including other alveolate species such as T. gondii and dinoflagellates as tested by our lab.

Validation of Anti-mCherry antibody (Abcam): The manufacturer states this antibody to be species independent. This antibody works on alveolate organisms P. falciparum and T. gondii.

Validation of HYGROMYCIN PHOSPHOTRANSFERASE ANTIBODY (10-1421) (anti-hygromycin antibody, Fitzgerald Industries): Recombinant hygromycin phosphotransferase from E.coli

Validation of H9658- Monoclonal Anti-HA antibody produced in mouse (Sigma):
1. The quality was tested by western blot analysis in E.coli expressing HA-Tag with tiger 1:40,000 dilution
2. In whole extract of human HEK-293T cells overexpressing N-terminal HA tagged fusion protein was separated on SDS-PAGE and probed with Monoclonal Anti-HA Clone: HA7 (Cat. No. H9658).

For each primary antibody, these antibodies have been used in several independent experiments with consistent results.

### Eukaryotic cell lines

**Policy information about cell lines**

| Cell line source(s) | Protist culture collections: Ostreococcus - RCC802; Bathycoccus - RCC4222; Micromonas - CCMP2709 and CCMP 1545; Tetraselmins -KAS-836; Pyramimonas - SCCAP K-0007; Isochrysis - CCMP1323; Emiliana - CCMP1516; Amorphochlora - CCMP2058; Bigelowia - CCMP 2755; Fragilariosis - CCMP1102; Thalasssiosira - CCMP1335; Seminavis - DCG 0498 and DCG 0514; Pseudo-nitzschia - 15091C3; Heterosigma - CCMP 2393; Aurantiocytherium - ATCC MYA-1381; Nannochloropsis - CCMP 1779; Phaeodactylum - CCAP1055/1; Chromera - CCMP 2878; Perkinsus - ATCC PRA240; Oxyrrhis - CCMP1788 and CCMP1795; Fugucium - CCMP 2468; Alexandrium - CCMP BF-5; Breviolum - NIES-4271; Crypthecodinium - CCMP316; Amphidinium - CCMP1314; Karlodinium - CCMP1975; Diplonema - ATCC50162; Eutreptiella - SCCAP K-0333; Naegleria - ATCC30224; Pirum - ATCC PR-280; Sphaeroforma - ATCC PRA-297; Abeoforma - ATCC PRA-279; Salpingoeca - ATCC PRA-390. |
| Authentication | Genome sequencing, sequencing of phylogenetic marker genes (e.g. 18S) |
| Mycoplasma contamination | The cell lines were not tested for mycoplasma contamination. |
| Commonly misidentified lines (See ICLAC register) | No commonly misidentified cell lines were used. |

### Flow Cytometry

**Plots**

Confirm that:
- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

**Methodology**

**Sample preparation**
Sample preparation is fully described in the Supplementary Materials. For experiments on the stramenopile (Nannochloropsis) electroporation occurred with or without fluorescein dextran and with or without a plasmid containing the mTagBFP2 gene under different promoters. No other stains were added after electroporation prior to cytometric analysis. For experiments with the prasinophyte (Micromonas) no stains were used at any point and experiments were performed with an eGFP plasmid with promoters from Micromonas, on cells that were either electroporated or not pulsed.

**Instrument**
BD InFlux cell sorter

**Software**
BD FACS™ Software for data collection
FlowJo for data analysis of stramenopiles, WinList for analysis of prasinophytes and WinList for all display panels. (please see above for version information).

**Cell population abundance**
For experiments demonstrating transfection of Nannochloropsis and Micromonas in pure culture cell population abundance was determined using an in-line flow meter on the sample line and by weighing the samples before and after data collection. This information on volume run was combined with counts acquired during each run to determine the abundance per ml for each population observed (non-transfected, transfected). Because of the small sample volumes and low abundance of target cells in samples of natural communities, it was not possible to re-test purity of sorted samples. Instead, immediately before sorting of the sample, the sorting efficiency was confirmed to >95% with 3 um UR calibration beads. With the same InFlux instrument and sorting protocols, we have previously determined that
sorted pico-cyanobacterial populations from natural samples are >99% pure, by re-running sorted samples using trigger based on the FSC channel and with voltage and threshold set to be able to detect even non-fluorescent particles with approx. 2x lower FSC than Prochlorococcus.

### Gating strategy

For Micromonas: The trigger channel was Forward Scatter (FSC), to allow detection of pigmented cells (both Micromonas and Nannochloropsis are algae with natural chlorophyll fluorescence), dead cells, and cell detritus. No gates were applied in the data collection mode, only for post analysis. All cells (with plasmid in treatments either pulsed, or not pulsed) were visualized in a two parameter histogram of FSC vs Chlorophyll (692/40 nm bp). For means, the population of non-detrital particles (living cells) was gated into a second two parameter histogram of FALS vs GFP (520/35 nm bp), as well as GFP vs Chlorophyll. From there cells with eGFP fluorescence were analyzed as were cells with baseline eGFP fluorescence and the geometric mean of these fluorescences were compared, as were cell abundance in each population.

For Nannochloropsis: The trigger channel was Forward Scatter (FSC), to allow detection of pigmented cells, dead cells, and cell detritus. A first gate was drawn based on FSC and Side Scatter to include particles with the optical scattering characteristics (related approximately to size and cell complexity) of Nannochloropsis cells. Living Nannochloropsis cells maintain high red fluorescence from chlorophyll, so a second gate was drawn based on chlorophyll fluorescence (692 nm, excited by the 488 nm laser) and FSC. In samples electroporated either with (treatment) or without plasmid DNA (sham controls), the proportion of cell-like particles (based on FSC and SSC) which had chlorophyll fluorescence remained high (>95%) during the first two hours but had dropped to 72-85% at 24 hours, while in non-electroporated samples the proportion remained >95% for the first two hours. In this way, blue fluorescence related to mTagBFP2 was assessed only on cells which maintained the optical characteristics of healthy Nannochloropsis.

For natural samples: The trigger channel was FSC, with voltage and threshold set to include all particles which displayed high yellow autofluorescence (at 580 nm) and high red autofluorescence (at 692 nm), which correspond to Synechococcus-like cells with yellow fluorescent phycoerythrin. In control (not electroporated) and sham-control (electroporated without plasmid), gates were drawn based on red fluorescence (due to chlorophyll, 692 nm excited by the 640 nm laser) and FSC as shown in Supplementary Fig. 6. Spherical fluorescent calibration particles (3 um Spherotech UR) were used as guides for setting FSC parameters to define smaller picoplankton versus cells that represent larger picoplankton or nanoplancton. Three clear phytoplankton “populations” (groups of particles with tightly similar optical characteristics) were visually identified and gates were drawn around them. The gate with lowest FSC was composed of 100% or nearly 100% cells exhibiting high yellow fluorescence due to phycoerythrin, so were assigned as "Cyanobacteria". The other two populations were therefore assumed to correspond to a type of small picophytoplankton (with FSC and pulse width well below that of 3 um calibration spheres) and larger picophytoplankton. These three populations together showed a clear relationship between chlorophyll fluorescence and FSC. Cells/particles with chlorophyll fluorescence below this relationship did not form clear groups and are assumed to represent non-phytoplankton, some of which can exhibit low red fluorescence due to phytoplankton prey in food vacuoles. For each time point, the gate to define blue fluorescent (putative BFP-expressing) cells was drawn to exclude over 95% of cells in sham-controls which were electroporated with dextran green (to track electroporation) but without plasmid DNA. This gate was drawn diagonally on 460 nm (blue fluorescence) versus 530 nm (green fluorescence) to exclude any signal that could come from bleedover of the green fluorescence of fluorescein dextran.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.