Single cell genomics reveals plastid-lacking Picozoa are close relatives of red algae

Max E. Schön1,2, Vasily V. Zlatogursky1,10, Rohan P. Singh3, Camille Poirier4,5,11, Susanne Wilken5,12, Varsha Mathur6, Jürgen F. H. Strassert1,13, Jarone Pinhassi7, Alexandra Z. Worden4,5, Patrick J. Keeling6, Thijs J. G. Ettema8, Jeremy G. Wideman3 & Fabien Burki1,9✉

The endosymbiotic origin of plastids from cyanobacteria gave eukaryotes photosynthetic capabilities and launched the diversification of countless forms of algae. These primary plastids are found in members of the eukaryotic supergroup Archaeplastida. All known archaeplastids still retain some form of primary plastids, which are widely assumed to have a single origin. Here, we use single-cell genomics from natural samples combined with phylogenomics to infer the evolutionary origin of the phylum Picozoa, a globally distributed but seemingly rare group of marine microbial heterotrophic eukaryotes. Strikingly, the analysis of 43 single-cell genomes shows that Picozoa belong to Archaeplastida, specifically related to red algae and the phagotrophic rhodelphids. These picozoan genomes support the hypothesis that Picozoa lack a plastid, and further reveal no evidence of an early cryptic endosymbiosis with cyanobacteria. These findings change our understanding of plastid evolution as they either represent the first complete plastid loss in a free-living taxon, or indicate that red algae and rhodelphids obtained their plastids independently of other archaeplastids.
The origin of plastids by endosymbiosis between a eukaryotic host and cyanobacteria was a fundamental transition in eukaryotic evolution, giving rise to the first photosynthetic eukaryotes. These ancient primary plastids, estimated to have originated >1.8 billion years ago, are found in Rhodophyta (red algae), Chloroplastida (green algae, including land plants), and Glaucophyta (glaucophytes)—together forming the eukaryotic supergroup Archaeplastida. Unravelling the sequence of events leading to the establishment of the cyanobacterial endosymbiont in Archaeplastida is complicated by antiquity, and by the current lack of modern descendants of early-diverging relatives of the main archaeplastidan groups in culture collections or sequence databases. Indeed, the only other known example of primary endosymbiosis are the chromophores in one unrelated genus of amoeba (Paulinella), which originated about a billion years later. Recently, two newly described phyla (Prasinodermophyta and Rhodelphidia) were found to branch as sister to green and red algae, respectively. Most transformative was the discovery that rhodelphids are obligate phagotrophs that maintain non-photosynthetic plastids, implying that the ancestor of red algae was likely mixotrophic, a finding that greatly alters our perspectives on early archaeplastid evolution.

While there is substantial evidence that Archaeplastida is a group descended from a photosynthetic ancestor, non-photosynthetic and plastid-lacking lineages have been found to branch near the base or even within archaeoplastid groups in phylogenomic trees. For example, Cryptista (which includes plastid-lacking and secondary plastid-containing species) have been inferred to be sister to either green algae and glaucophytes, or red algae, although other phylogenomic analyses have recovered the monophyly of Archaeplastida to the exclusion of the cryptists. Another non-photosynthetic group that recently showed affinities to red algae based on phylogenomics is Picozoa. But as for cryptists, the position of Picozoa has lacked consistent support, mostly because there is no member of Picozoa available in continuous culture, and genomic data are currently restricted to a few, incomplete, single amplified genomes (SAGs). Thus, the origin of Picozoa remains unclear.

Picozoa (previously known as picobiliphytes) were first described in 2007 in marine environmental clone libraries of the 18S ribosomal RNA (rRNA) gene and observed by epifluorescence microscopy in temperate waters. Based on orange autofluorescence reminiscent of the photosynthetic pigment phycobiliprotein and emanating from an organelle-like structure, picozoans were initially described as likely containing a plastid. Orange fluorescence was also observed in association with these uncultured cells in subtropical waters. However, the hypothesis that the cells were photosynthetic was challenged by the characterisation of SAG data from three picozoan cells isolated by fluorescence-activated cell sorting (FACS). The analysis of these SAGs revealed neither plastid DNA nor nuclear-encoded plastid-targeted proteins, but the scope of these conclusions is limited due to the small number of analysed cells and the highly fragmented and incomplete obtained data. Most interestingly, a transient culture was later established, enabling the formal description of the first (and so far only) picozoan species—Picomonas judraskeda—as well as ultrastructural observations with electron microscopy. These observations revealed an unusual structural feature in two body parts, a feeding strategy by endocytosis of nano-sized colloid particles, and confirmed the absence of plastids. Only the 18S rRNA gene sequence of P. judraskeda is available as the transient culture was lost before genomic data could be generated.

Here, we present an analysis of genomic data from 43 picozoan single-cell genomes sorted with FACS from the Pacific Ocean off the Californian coast and from the Baltic Sea. Using a gene and taxon-rich phylogenomic dataset, these data allowed us to robustly infer Picozoa as a lineage of archaeoplastids, branching with red algae and rhodelphids. With this expanded genomic dataset, we confirm Picozoa as the first archaeplastid lineage lacking a plastid. We discuss the important implications that these results have on our understanding of the origin of plastids.

Results
Single-cell assembled genomes representative of Picozoa diversity. We isolated 43 picozoan cells (40 from the eastern North Pacific off the coast of California, 3 from the Baltic Sea) using FACS and performed whole genome amplification by multiple displacement amplification (MDA). The taxonomic affiliation of the SAGs was determined either by PCR with Picozoa-specific primers or 18S rRNA gene sequencing using general eukaryotic primers, followed by Illumina sequencing of the MDA products (see ’Methods’). The sequencing reads were assembled into genomic contigs, with a total assembly size ranging from 350 Mbp to 66 Mbp (Fig. 1a and Supplementary Data 1). From these contigs, the 18S rRNA gene was found in 37 out of the 43 SAGs, which we used to build a phylogenetic tree with reference sequences from the protist ribosomal reference PR2 database (Supplementary Fig. 1). Based on this tree, we identified 6 groups representing 32 SAGs that possessed nearly identical 18S rRNA gene sequences within each group. These SAGs with identical ribotype were reassembled by pooling all reads in order to obtain longer, more complete co-assemblies (CO-SAGs). The genome size of the CO-SAGs ranged from 32 to 109 Mbp (Fig. 1a and Supplementary Data 1), an increase of 5–45% over individual SAGs. The genome completeness of the SAGs and CO-SAGs was estimated based on two datasets: (i) a set of 255 eukaryotic marker genes available in BUSCO, and (ii) a set of 317 conserved marker genes derived from a previous pan-eukaryote phylogenomic dataset that we used here as starting point in downstream analyses (Fig. 1b). These comparisons showed that while most SAGs were highly incomplete (Fig. 1a, b), the CO-SAGs were generally more complete (up to 60%). When taken together, 90% of the BUSCO markers and 88% of the phylogenomic markers were present in at least one assembly, suggesting that while the single-cell genome assemblies are fragmentary, they together represent a much more complete Picozoa meta-assembly.

The final 17 assemblies (11 SAGs and 6 CO-SAGs) were mainly placed within the three proposed groups of Picozoa BP1-3 (Fig. 1c), sensu Cuvelier et al., but SAG11 was placed outside of these groups. The deep-branching picozoan lineages identified by Moreira and López-García, as well as other possibly early-diverging lineages were not represented in our data (Fig. 1c). Interestingly, one CO-SAG (COSAG03) was closely related to the only described species, Picomonas judraskeda, for which no genomic data are available (18S rRNA gene 100% identical). Using our assemblies and reference sequences from PR2 as queries, we identified by sequence identity 362 OTUs related to Picozoa (≥90 %) in the data provided by the Tara Ocean Project. Picozoa were found in all major oceanic regions, but had generally low relative abundance in V9 18S rRNA gene amplicon data (less than 1% of the eukaryotic fractions in most cases, Supplementary Fig. 2). An exception was the Southern Ocean between South America and Antarctica, where the Picozoa-related OTUs in one sample represented up to 30% of the V9 18S rRNA gene amplicons. Thus, Picozoa seems widespread in the oceans but generally low in abundance based on available sampling, although they can reach higher relative abundances in at least circumpolar waters.
Phylogenomic dataset construction. To infer the evolutionary origin of Picozoa, we expanded on a phylogenomic dataset that contains a broad sampling of eukaryotes and a large number of genes that was recently used to study deep nodes in the eukaryotic tree\(^1\). Homologues from the SAGs and CO-SAGs as well as a number of newly sequenced key eukaryotes were added to each single gene (see Supplementary Table 1 for a list of taxa). After careful examination of the single genes for contamination and orthology based on individual phylogenies (see ‘Methods’), we retained all six CO-SAGs and four individual SAGs together with the available SAG MS584-11 from a previous study\(^1\). The rest of the SAGs were excluded due to poor data coverage (less than five markers present) and, in one case (SAG33), because it was heavily contaminated with sequences from a cryptophyte (see ‘Data availability’ for access to the gene trees). In total, our phylogenomic dataset contained 794 taxa and 317 protein-coding genes, with orthologues from Picozoa included in 279 genes (88%) (Fig. 1b). This represents an increase in gene coverage from 18 to 88% compared to the previously available genomic data for Picozoa. The most complete assembly was COSAG01, from which we identified orthologues for 163 (51%) of the markers.

Picozoa group with Rhodophyta and Rhodelphidia. Concatenated protein alignments of the curated 317 genes were used to infer the phylogenetic placement of Picozoa in the eukaryotic Tree of Life. Initially, a maximum likelihood (ML) tree was reconstructed from the complete 794-taxa dataset using the site-homogeneous model LG+F+G and ultrafast bootstrap support with 100 replicates (Supplementary Fig. 3). This analysis placed Picozoa together with a clade comprising red algae and rhodelphids with strong support (100% UFBoot2), but the monophyly of Archaeplastida was not recovered due to the internal placement of cryptists. To further investigate the position of Picozoa, we applied better-fitting site-heterogeneous models to a reduced dataset of 67 taxa, since these models are computationally much more demanding. The process of taxon reduction was driven by the requirement of maintaining representation from all major groups, while focusing sampling on the part of the tree where Picozoa most likely belong to, i.e. Archaeplastida, TSAR, Haptista and Cryptista. We also merged several closely related lineages into OTUs based on the initial ML tree in order to reduce missing data (Supplementary Data 2). Both ML and Bayesian analyses produced highly similar trees, and received maximal support for the majority of relationships, including deep divergences (Fig. 2). Most interestingly, both analyses recovered the monophyly of Archaeplastida (BS = 93%; PP = 1), with cryptists as sister lineage (BS = 100%; PP = 1). Consistent with the initial ML tree (Supplementary Fig. 3), red algae and rhodelphids branched together (BS = 95%; PP = 1), with Picozoa as their sister with full support (BS = 100%; PP = 1). This grouping was robust to fast-evolving sites removal analysis (Supplementary Fig. 4), trimming of the 25 and 50% compositionally most biased
sites (Supplementary Fig. 5), and was also recovered in a super-tree method (ASTRAL-III) consistent with the multi-species coalescent model (Supplementary Fig. 6). Although this group is robust, we observed one variation in the branching order between Picozoa, rhodelphids and red algae when trimming the 50% most heterogeneous sites (Supplementary Fig. 7) and after removing genes with less than two picozoan sequences (Supplementary Fig. 8). In these analyses, Picozoa and red algae were most closely related, although this relationship was never significantly supported. An approximately unbiased (AU) test rejected all tested topologies except in the two cases where Picozoa branched as the closest sister to red algae ($p = 0.237$) and the topology of Fig. 2 ($p = 0.822$; Supplementary Table 2). Finally, we identified in Picozoa and rhodelphids a two amino acids replacement signature in the eukaryotic translation elongation factor 2 protein (SA instead of the ancestral GS residues, see Supplementary Data 3) that was previously shown to unite red and green algae (and land plants), haptophytes and some cryptists. The presence of SA in Picozoa supports their affiliation with red algae and rhodelphids.

Picozoa SAGs show no evidence of a plastid. Since there have been conflicting conclusions about the occurrence of plastids in picozoans, we extensively searched our genomic data for evidence

Fig. 2 Maximum likelihood tree of eukaryotic species showing the position of Picozoa. The tree is based on the concatenated alignment of 317 marker genes and was reconstructed using the site-heterogeneous model LG + C60 + F + G-PMSF. Support values correspond to 100 non-parametric bootstrap replicates/posterior probability values estimated using PhyloBayes CAT-GTR + G. Black circles denote full support ($=100/1.0$). Insert shows the only other topology not rejected in an AU topology test, which was also recovered when trimming the 50% most heterogeneous sites of the alignment.
of cryptic plastids. First, we searched the SAG and CO-SAG assemblies for plastidial contigs as evidence of a plastid genome. While there were some contigs that initially showed similarities to reference plastid genomes, these were all rejected as bacterial (non-cyanobacterial) contamination upon closer inspection. In contrast, mitochondrial contigs were readily identified in 26 of 43 SAGs (Supplementary Data 4). Although mitochondrial contigs remained fragmented in most SAGs, four complete or near-complete mitochondrial genomes were recovered with coding content near-identical to the published mitochondrial genome for picozoan MS5584-11 (Supplementary Fig. 9). The ability to assemble complete mitochondrial genomes from the SAGs suggests that the partial nature of the data does not specifically hinder organelle genome recovery if present, at least in the case of mitochondria.

Second, we investigated the possibility that the plastid genome was lost while the organelle itself has been retained—as is the case for Rhodelphis. For this, we reconstructed phylogenetic trees for several essential nuclear-encoded biochemical plastid pathways derived by endosymbiotic gene transfer (EGT) that were shown to be at least partially retained even in cryptic plastids. These included genes involved in the biosynthesis of isoprenoids (ispD,E,F,G,H, ddx, dxs), fatty acids (fabD,F,G,H,I,Z, ACC), heme (hemB,D,E,F,H,Y, ALAS), and iron-sulfur clusters (subB,C,D,E,S, NirU, iscA; see also Supplementary Data 55). In all cases, the picozoan homologues grouped either with bacteria—but not cyanobacteria, suggesting contamination—or the mitochondrial/nuclear copies of host origin. Furthermore, none of the picozoan homologues contained predicted N-terminal plastid transport peptides. We also searched for picozoan homologues of all additional proteins (n = 62) that were predicted to be targeted to the cryptic plastid in rhodelphids. This search resulted in one protein (Arogenate dehydrogenase, OG0000831) with picozoan homologues that were closely related to red algae and belonged to a larger clade with host-derived plastid-targeted plant sequences, but neither the picozoan nor the red algal sequences displayed predicted transit peptides. Finally, to eliminate the possibility of missing sequences because of errors during the assembly and gene prediction, we additionally searched the raw reads sequences for the same plastid-targeted or plastid transport machinery genes, which revealed no obvious candidates. In contrast, we readily identified mitochondrial genes (e.g. homologues of the mitochondrial import machinery from the TIM17/TIM22 family), which further strengthened our inference that the single-cell data are in principle adequate to identify organellar components, when they are present.

The lack of cryptic plastids in diverse modern-day picozoans does not preclude photosynthetic ancestry if the plastid was lost early in the evolution of the group. To assess this possibility, we searched more widely for evidence of a cyanobacterial footprint on the nuclear genome that would rise above a background of horizontal gene transfers for proteins functioning in cellular compartments other than the plastids. The presence of a significant number of such proteins may be evidence for a plastid-bearing ancestor. We clustered proteins from 419 genomes, including all major eukaryotic groups as well as a selection of bacteria into orthogonal groups (OGs) (Supplementary Data 6). We built phylogenies for the OGs that contained at least cyanobacterial and algal sequences, as well as a sequence from one of 33 focal taxa, including Picozoa, a range of photosynthetic taxa, but also non-photosynthetic plastid-containing, and plastid-lacking taxa to be used as controls. Putative gene transfers from cyanobacteria (EGT) were identified as a group of plastid-bearing eukaryotes that included sequences from the focal taxa and branched sister to a clade of cyanobacteria. We allowed up to 10% of sequences from groups with no plastid ancestry.

This approach identified 16 putative EGTs for Picozoa where at least 2 different SAGs/CO-SAGs grouped together, compared to between 89 and 313 EGTs for photosynthetic species, and up to 59 EGTs for species with non-photosynthetic plastids (Fig. 3a). At the other end of the spectrum for species with non-photosynthetic plastids, we observed that the number of inferred cyanobacterial genes for e.g. rhodelphids (14) or Paraphysomonas (12) was comparable to Picozoa (16) or other, plastid-lacking taxa such as Telonema (15) or Gonionomas (18). In order to differentiate these putative endosymbiotic transfers from a background of bacterial transfers (or bacterial contamination), we next attempted to normalise the EGT signal by estimating an extended bacterial signal (indicative of putative HGT: horizontal gene transfers) using the same tree sorting procedure (Supplementary Fig. 10). When comparing the number of inferred EGT with that of inferred HGT, we found a marked difference between plastid-containing (including non-photosynthetic) and plastid-lacking lineages. While all plastid-containing taxa—with the notable exception of Rhodelphis—showed a ratio of EGT to HGT above 1, all species without plastid ancestry and Hematodinium, one of the few taxa with reported plastid loss, as well as Rhodelphis and Picozoa showed a much higher number of inferred HGT than EGT.

Discussion

The 17 SAGs and CO-SAGs of Picozoa obtained in this study provide robust data for phylogenomic analyses of this important phylum of eukaryotes. With this data, we are able to firmly place Picozoa within the supergroup Archaeplastida, most likely as a sister lineage to red algae and rhodelphids. Archaeplastids contain all known lineages with primary plastids (with the exception of Paulinella), which are widely viewed to be derived from a single primary endosymbiosis with a cyanobacterium. This notion of a common origin of primary plastids is supported by cellular and genomic data (see refs. 23,24 and references therein for review), as well as plastid phylogenetics.25,26 The phylogenetic support for Archaeplastida based on host (nuclear) data has been less certain, but our analysis is consistent with recent reports that have also recovered a monophyletic origin—here including Picozoa—when using gene and taxon-rich phylogenomic datasets.19,10 This position has important implications for our understanding of plastid origins because, in contrast to all other archaeoplastids known to date, our results indicate that Picozoa lack plastids and plastid-associated EGTs. The lack of plastid in Picozoa was also inferred based on smaller initial SAG data11 as well as ultrastructural observation of P. judraskeda.14 Two main possible hypotheses exist to explain the lack of plastids in Picozoa: that this group was never photosynthetic, or complete plastid loss occurred early in their evolution.

To suggest that Picozoa was never photosynthetic requires that the current distribution of primary plastids is due to multiple independent endosymbioses, specifically that red algae (and possibly Rhodelphis) arose from one or two separate primary endosymbioses from that leading to green algae and glaucophytes. This scenario would have involved the endosymbioses of closely related cyanobacterial lineages in closely related hosts to explain the many similarities between primary plastids.24 Although this may sound unlikely, there is accumulating evidence that similar plastids were derived independently from similar endosymbions in closely related hosts in dinoflagellates with tertiary plastids, and has been argued before for primary plastids.31–34 However, the current bulk of cell and molecular evidence suggests that multiple independent origins of primary plastids are unlikely, including several features of plastid biology that are not present in cyanobacteria (e.g., protein targeting...
alveolates: in Cryptosporidium plastid loss all come from parasitic lineages (all in myzozoan because to date, the only known unambiguous cases of total loss in a free-living lineage like Picozoa would be unprecedented; Picozoa entirely lost its primary plastid. The possibility of plastid components shared between all archaeplastid lineages. than HGT. No ratio could be calculated for integration. EGT has occurred in all algae, although its impact on nuclear genomes can vary and the inference of EGT to the plastids, are recognised as a hallmark of organelle transfer of genes from endosymbiont to host nucleus via EGT, genome that would result from an ancestral endosymbiosis. The partial nature of eukaryotic SAGs makes it possible that EGTs are absent from our data, even with >90% of inferred genomic completeness. Additionally, the possibility exists that the number of EGT might have always been low during the evolution of the group, even if a plastid was once present. Recent endosymbioses where EGT can be pinpointed with precision showed a relatively low frequency. For example, they represent at most a few percent of the chromatophore proteome in Galdieria sulphuraria (168), suggesting that the bulk of endosymbiotic transfers in red algae may have happen after their divergence from rhodelphids.

In this study, we used single-cell genomics to demonstrate that Picozoa are a plastid lacking major lineage of archaeplastids. To our knowledge, this is the first example of an archaeplastid lineage without plastids, which can be interpreted as either plastid loss, or evidence of independent endosymbiosis in the ancestor of red algae and rhodelphids. Under the most widely accepted scenario of a single plastid origin in Archaeplastida, Picozoa would represent the first known case of plastid loss in this group, but also more generally in any free-living species. In order to discriminate plastid loss from multiple plastid gains in the early archaeplastid evolution, and more generally during the evolution

| Taxon                                      | Plastid status                        | Photosynthetic plastid | Non-photosynthetic plastid | Confirmed plastid loss | Picozoa | No plastid ancestry |
|--------------------------------------------|--------------------------------------|------------------------|-----------------------------|------------------------|---------|---------------------|
| Rattus norvegicus                         | (0)                                  | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Neurospora crassa                         | (1)                                  | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Dictyostelium discoideum                  | (1)                                  | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Tetrahymena thermophila                   | (1)                                  | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Thecamonas tafraei                         | (1)                                  | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Cryptosporidium minus                      | (1)                                  | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Phytophthora capsici                        | (6)                                  | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Paraphysomonas bandoensis                  | (12)                                 | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Rhodelphis                                 | (14)                                 | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Telonema subtile                           | (15)                                 | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Toxoplasma gondii                          | (17)                                 | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Goniomonas pacifica                        | (18)                                 | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Pedospmella elongata                       | (19)                                 | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Spumella bureschi JBL14                    | (31)                                 | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Cryptomonas paramecium                     | (34)                                 | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Polytomella parva                         | (58)                                 | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Helicosporidium sp ATCC50302               | (59)                                 | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Mallomonas sp CCMP3275                     | (85)                                 | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Ochromonadales sp CCMP2298                 | (110)                                | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Dinobryon sp UTEXLB2267                    | (122)                                | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Virella brassiciforms                      | (130)                                | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Guillardia theta                           | (143)                                | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Leptocylindrus danicus                     | (151)                                | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Paulinella chromatophora                   | (160)                                | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Bigelowiella natans                        | (161)                                | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Galdieria sulphuraria                      | (168)                                | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Emiliania hueleyi                          | (177)                                | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Alexandrium tamarense                      | (181)                                | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Chloropica primus                          | (225)                                | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Cyanophora paradoxa                        | (228)                                | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |
| Arabidopsis thaliana                       | (313)                                | photosynthetic plastid | non-photosynthetic plastid | confirmed plastid loss | Picozoa | no plastid ancestry |

The lack of a genomic baseline to assess plastid loss in Picozoa is further complicated by limitations of our data and methods. The partial nature of eukaryotic SAGs makes it possible that EGTs are absent from our data, even with >90% of inferred genomic completeness. Additionally, the possibility exists that the number of EGT might have always been low during the evolution of the group, even if a plastid was once present. Recent endosymbioses where EGT can be pinpointed with precision showed a relatively low frequency. For example, they represent at most a few percent of the chromatophore proteome in Paulinella, or as few as nine genes in tertiary endosymbiosis in dinoflagellates. Thus, it is possible that the much higher number of EGT inferred in red algae (e.g. 168 in Galdieria) occurred after the divergence of Picozoa, and that Picozoa quickly lost its plastid before more EGT occurred. An observation that supports this hypothesis is the low number of putative EGTs found in Rhodelphis (14), suggesting that the bulk of endosymbiotic transfers in red algae may have happen after their divergence from rhodelphids.

In this study, we used single-cell genomics to demonstrate that Picozoa are a plastid lacking major lineage of archaeplastids. To our knowledge, this is the first example of an archaeplastid lineage without plastids, which can be interpreted as either plastid loss, or evidence of independent endosymbiosis in the ancestor of red algae and rhodelphids. Under the most widely accepted scenario of a single plastid origin in Archaeplastida, Picozoa would represent the first known case of plastid loss in this group, but also more generally in any free-living species. In order to discriminate plastid loss from multiple plastid gains in the early archaeplastid evolution, and more generally during the evolution
of secondary or tertiary plastids, a better understanding of the early steps of plastid integration is required. In the recently evolved primary plastid-like chromatophores of Paulinella, the transfer of endosymbiotic genes at the onset of the integration was shown to be minimal. Similar examples of integrated plastid endosymbionts but with apparently very few EGTs are known in dinoflagellates. Therefore, new important clues to decipher the origin of plastids will likely come from a better understanding of the role of the host in driving these endosymbioses, and crucially the establishment of a more complete framework for archaeplastid evolution with the search and characterisation of novel diversity of lineages without plastids. The fact that this lineage has never been successfully maintained in culture, with just one study achieving transient culture, might indicate a lifestyle involving close association with other organisms (such as symbioses) and further underscores the enigma of picozoan biology, the lack of information on which hinder our interpretation of their evolution.

Methods

Cell isolation, identification, and genome amplification

Bottle Sea. Surface (depth: up to 2 m) marine water was collected from the Linnaeus micro station (56°50.7′N, 17°41.6′E) at 06:35 on 13 May 2013 (11°C and 9.9 psu salinity). The samples were transported to the laboratory and filter-fractionised. The size fractions larger than 2 µm were discarded whereas the fraction collected on 0.2 µm filters were resuspended in 2 mL of the filtrate. The obtained samples were used for fluorescence-activated cell sorting (FACS). Aliquots of 4 µL of 1 mM Mitotracker Green FM (ThermoFisher) stock solution were added to the samples and were kept in the dark at 15°C for 15–20 min. Then the cells were sorted into empty 96-well plates using MoFlo Astrios EQ cell sorter (Beckman Coulter). Gates were set mainly based on Mitotracker intensity and the dye was detected by a 488 nm and 640 nm laser for excitation, 100 µm nozzle, sheath pressure of 25 psi and 0.1 µm sterile filtered 1× PBS as sheath fluid. The region with the highest green fluorescence and forward scatter contained the target group and was thereafter used alongside with exclusion of red autofluorescence (Summit v 6.3.1).

The SAGs were generated in each well with REPLI-g® Single Cell Kit (QIagen) following the manufacturer’s recommendations but scaled down to 5 µL reactions. Since the cells were sorted in dry plates, 400 nL of 1× PBS was added prior to 300 nL of lysis buffer D2 for 10 min at 65°C and 10 min on ice, followed by 300 nL stop solution. The PBS, reagent D2, stop solution, water, and reagent tubes were UV-treated at 2 joules before use. A final concentration of 0.5 µM SYTO 13 (Invitrogen) was added to the MDA mastermix. The reaction was run at 30°C for 6 h followed by incubation at 65°C for 5 min and was monitored by detection of SYTO13 fluorescence every 15 min using a FLUOstar® Omega plate reader (BMG Labtech, Germany). The single amplified genome (SAG) DNA was stored at −20°C until PCR amplification. The obtained products were PCR screened using Pico-PCR approach (primers PICOBI1F, 5′-CCAGTTTGGTAGC3C-3′, and P0ITIS1R, 5′-CATCTCAATGTTCACGTGG-3′), as described in ref. 18 and the wells showing signal for Picozoa were selected for sequencing.

Sequencing

Sequencing libraries were prepared from 100 ng DNA using the TruSeq Nano DNA sample preparation kit (cat# 2011596/4, Illumina Inc.) targeting an insert size of 350 bp. For six samples, less than 12.5 ng of DNA was used due to DNA input between 8.7 and 17 ng). The library preparation was performed by SNP&SEQ Technology Platform at Uppsala University according to the manufacturers’ instructions. All samples were then multiplexed on one lane of an Illumina HiSeq instrument with 150 cycles paired-end sequencing using the v2.5 sequencing chemistry, producing between 10,000 and 30,000,000 read pairs.

For each of these 17 assemblies we estimated the amount of prokaryotic/viral contamination by comparing the predicted proteins against the NCBI nr database using DIAMOND in blastp mode60. If at least 60% of all proteins from a contig produced significant hits only to sequences annotated as prokaryotic or viral, we considered that contig to be a putative contamination. In general only a small fraction of each assembly was found to be such a contamination (Supplementary Fig. 11 for a tree with extended taxon sampling). Additionally, we estimated the average nucleotide identity (ANI) for all pairs of SAGs using fastANI v1.28 (Supplementary Fig. 12). Based on the 18S rRNA gene tree and the ANI value, groups of closely related SAGs with almost identical 18S rRNA gene sequences (sequence similarity above 99%) were identified as putative co-assembly. Co-assemblies were then generated in the same way as described above, pooling sequencing libraries from closely related single cells. ORFs and rRNA genes were similarly extracted from the co-assemblies. The completeness of the SAGs and CO-SAGs was then assessed using BUSCO v4.1.37 with 255 markers for eukaryotic genomes and transcriptomes (Supplementary Fig. 13) as well as using the 320 marker phylogenomic database as described below. General genome characteristics were computed using QUAST v5.0.2. Alignment reconstructions were performed for the 18S rRNA genes from the co-assemblies and those SAGs that were not included in any CO-SAG together with PR2 references for cryptists and katablepharids (the closest groups to Picozoa in 18S rRNA phylogenies) in the same way as described above. The tree was reconstructed using GTR + R + F after model selection and support was assessed with 100 non-parametric bootstrap replicates. The six CO-SAGs and the 11 individual SAGs were used in all subsequent analyses.

Phylogenomics

Existing untrimmed alignments for 320 genes and 763 taxa from ref. 7 were used to create HMMP profiles in HMMER v3.2.17, which were then used to reconstruct the homologous sequence for each of the assemblies used in the Picozoa (or co-assemblies) as well as in 20 additional, recently sequenced eukaryotic genomes and transcriptomes (Supplementary Table 1). Each
single gene dataset was filtered using PREQUAL v1.026 to remove non-
homologous residues prior to alignment, aligned using MAFFT E-INS-i, and fil-
tered with TrimAl v1.4rev16. Alignment of gene trees with IQ-TREE (--mset LG, LG4X; 1000 ultrafast bootstraps with the BNNI
optimisation). All trees were manually scrutinised to identify contamination and
paralogues. These steps were repeated at least two times, until no further con-
taminations or paralogues could be detected. We excluded three genes that showed
ambiguous groupings to rhodelphids or red algae as well as for other groupings (Supplementary Fig. 4).

We also performed trimming of the 25 and 50% most heterogeneous sites based on
relative abundance on a world map.

Mitochondrial contig identification and annotation. Using the published
picozoan mitochondrial genome (Picozoa sp. MSS84-11: MG202007.1 from ref. 19),
we then prepared a reduced dataset with a more focused taxon sampling of 67 taxa
and BMGE (-g 0.2 -b 10 -m BLOSUM75, v1.12) and concatenated into two
dataset were re-aligned using MAFFT E-INS-i, filtered using both Divvier -partial
BMGE (-g 0.2 -b 10 -m BLOSUM75, v1.12) and concatenated into two
supermatrix. Model selection of mixture models was performed using
Modeller v2.4.073. For Picozoa and a selection of 32 photosynthetic or heterotrophic lineages
( Supplementary Table 3), we inferred trees for 2626 clusters that contained the species
under consideration, at least one cyanobacterial sequence, and at least one archaeplastid
sequence of red algae, green algae or plants. Alignments for these clusters were
reconstructed using IQ-TREE (-m LG4X; 1000 ultrafast bootstraps with the BNNI
optimisation). We then identified trees where the target species grouped with other
plastid-bearing lineages (allowing up to 10% non-plastid sequences) and sister at least
two cyanobacterial sequences. For Picozoa, we added the condition that sequences from
at least two SAGs/OSAG assemblies must be monophyletic. For species with no known
plastid ancestry such as Rattus or Phyllodictyon, putative EGTs can be interpreted as
false positives due to contamination, poor tree resolution or other mechanisms, since we
expect no EGTs from cyanobacteria to be present at all in these species. This rough
estimate of the expected false-positive rate for this approach can give us a baseline of
false positives that can be expected for picozoa as well.

To put the number of putative EGTs into relation to the overall amount of gene
transfers, we applied a very similar approach to the one described above for
detecting putative HGT events. We prepared additional trees (in the same way as
described for the detection of EGTs) for clusters that contained the taxon of
interest and non-cyanobacterial bacteria and identified clades of the taxon under
consideration (including a larger taxonomic group, e.g. Streptophyta for
Arabidopsis or Metaezoa for Rattus) that branched sister to a bacterial clade.

Distribution of Picozoa in Tara Oceans. We screened available OTUs that were
taken from V9 18S rRNA gene eukaryotic ampiclon data generated by Tara
Oceans11 for sequences related to Picozoa. Using the V9 region of the 18S rRNA
gene sequences from the 17 Picozoa assemblies as well as from the picozoan PR2
sequences used to reconstruct the 18S rRNA gene tree described above, we applied
VSEARCH v2.15.14 (--usearch_global -id 0.90) to find all OTUs with at least
90% similar V9 regions to any of these reference picozoan sequences. Using
the relative abundance information available for each Tara Oceans sampling
location, we then computed the sum for all identified Picozoa OTUs per station
and plotted the relative abundance on a world map.

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

Data availability
All data used for the analyses as well as results files such as contigs and single gene
trees are available at figshare (https://doi.org/10.6084/m9.figshare.c.5388176).
A sequenceServer BLAST server was set up for the SAG assemblies: http://eucocolbio.com/SAGdb/burki/. Raw
sequencing reads were deposited in the Sequence Read Archive (SRA) at NCBI under accession PRJNA747736. Source data are provided with this paper.

Code availability
All code used in this study is available at https://github.com/maxeill/picozoa-scripts (https://doi.org/10.5281/zenodo.5561108) under an MIT license.

Received: 8 May 2021; Accepted: 29 October 2021;
Published online: 17 November 2021

References
1. Strasser, J. F. H., Irisarri, I., Williams, T. A. & Burki, F. A molecular timescale
for eukaryote evolution with implications for the origin of red algal-derived
plastids. Nat. Commun. 12, 1879 (2021).
2. Burki, F., Roger, A. J., Brown, M. W. & Simpson, A. G. B. The new tree of
eukaryotes. Trends Ecol. Evol. 35, 43–55 (2019).
3. Marín, B., Nowack, E. C. M. & Melkonian, M. A plastid in the making: evidence for a second primary endosymbiosis. *Protist* **156**, 425–432 (2005).

4. Nowack, E. C. M. & Weber, A. P. M. Genomics-informed insights into endosymbiotic organelle evolution in photosynthetic eukaryotes. *Annu. Rev. Plant Biol.* **69**, 1–34 (2018).

5. Gawryluk, R. M. et al. Non-photosynthetic predators are sister to red algae. *Phil. Trans. R. Soc. B Biol. Sci.* **369**, 20130144 (2014).

6. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *Eur. Biophys. J.* **41**, 149–161 (2012).

7. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for aligning genomic and protein sequences. *Bioinformatics* **21**, 363–364 (2005).

8. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

9. Hehenberger, E., Burki, F., Kolisko, M. & Keeling, P. J. Functional relationship between a dinoflagellate host and its diatom endosymbiont. *Genome Biol. Evol.* **3**, 2376–2390 (2011).

10. Needham, D. M. et al. A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators. *Proc. Natl Acad. Sci. USA* **116**, 20574–20583 (2019).

11. Stiller, J. W., Roel, D. C. & Johnson, J. C. A single origin of plastids revisited: convergent evolution in organellar genome content. *J. Phycol.* **39**, 95–105 (2003).

12. Larkum, A. W. D. Lockhart, P. J. & Howe, C. J. Shopping for plastids. *Trends Plant Sci.* **12**, 189–195 (2007).

13. Singh, A. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

14. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. PRINQUAL: detecting non-homologous clusters of high confidence. *PLoS Comput. Biol.* **7**, e1001915 (2011).

15. Whelan, S., Irisarri, I. & Burki, F. PREQUAL: detecting non-homologous clusters of high confidence. *PLoS Comput. Biol.* **7**, e1001915 (2011).

16. Pinney, J., Barbrook, A. C., Nisbet, R. E. R., Lockhart, P. J. & Larkum, A. W. D. The origin of plastids. *Philos. Trans. R. Soc. B Biol. Sci.* **363**, 2675–2685 (2008).

17. Stiller, J. W. Toward an empirical framework for interpreting plastid evolution. *J. Phycol.* **49**, 497–517 (2013).

18. Bayrakhadyarova, A., Archibald, J. M., Weber, A. P. M. & Reyes-Prieto, A. How do endosymbioses become organelles? Understanding early events in plastid evolution. *Bioessays* **29**, 1229–1246 (2007).

19. Kim, E. & Maruyama, S. A contemplation on the secondary origin of green and plant plastids. *Acta Soc. Bot. Pol.* **83**, 331–336 (2014).

20. Zha, G., Marchewka, M. J. & Keihly, J. S. Cryptosporidium parvum appears to lack a plastid genome. *Microbiology* **146**, 315–321 (2000).

21. Janoušková, J., Issler, M. A., Huang, C. Y. & Martin, W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomal states. *Nat. Rev. Genet.* **5**, 123–135 (2004).

22. Archibald, J. M. Genomic perspectives on the birth and spread of plastids. *Proc. Natl Acad. Sci. USA* **112**, 10147–10153 (2015).

23. Burki, F. et al. Re-evaluating the green versus red signal in eukaryotes with secondary plastid of red algal origin. *Genome Biol. Evol.* **4**, 626–635 (2012).

24. Deschamps, P. & Moreira, D. Reevaluating the green contribution to diatom genomes. *Genome Biol. Evol.* **4**, 683–688 (2012).

25. Quo, H., Yoon, H. S. & Bhattacharya, D. Algal endosymbiosms as vectors of horizontal gene transfer in photosynthetic eukaryotes. *Front. Plant Sci.* **4**, 1–8 (2013).

26. Morozov, A. A. & Galachyants, Y. P. Diatom genes originating from red and green algae: implications for the secondary endosymbiosis models. *Mar. Genom.* **45**, 72–78 (2019).

27. Sibbald, S. J. & Archibald, J. M. Genomic insights into plastid evolution. *Microbiol. Biol.* **12**, 165 (2020).

28. Singer, A. et al. Massive protein import into the early-evolutionary-stage photosynthetic organelle of the Amoeba Paulinella chromatophora. *Curr. Biol.* **30**, 2763–2773.e5 (2017).

29. Burki, F. et al. Endosymbiotic gene transfer in tertiary plastid-containing dinoflagellates. *Eukaryot. Cell* **13**, 246–255 (2014).

30. Hohenberg, E., Burki, F., Kolisko, M. & Keeling, P. J. Functional relationship between a dinoflagellate host and its diatom endosymbiont. *Mol. Biol. Evol.* **33**, 2376–2390 (2016).

31. Needham, D. M. et al. A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators. *Proc. Natl Acad. Sci. USA* **116**, 20574–20583 (2019).

32. Mori, A. A. & Galachyants, Y. P. Diatom genes originating from red and green algae: implications for the secondary endosymbiosis models. *Mar. Genom.* **45**, 72–78 (2019).

33. Buick, F. et al. Principles of plastid reductive evolution illuminated by single-cell genomics. *Genome Biol. Evol.* **12**, 246 (2020).

34. Zha, G., Marchewka, M. J. & Keihly, J. S. Cryptosporidium parvum appears to lack a plastid genome. *Microbiology* **146**, 315–321 (2000).

35. Bhattacharya, D., Archibald, J. M., Weber, A. P. M. & Reyes-Prieto, A. How do endosymbioses become organelles? Understanding early events in plastid evolution. *Bioessays* **29**, 1229–1246 (2007).

36. Bayrakhadyarova, A., Archibald, J. M., Weber, A. P. M. & Reyes-Prieto, A. How do endosymbioses become organelles? Understanding early events in plastid evolution. *Bioessays* **29**, 1229–1246 (2007).
Acknowledgements

This work was supported by a grant from Science for Life Laboratory available to F.B. and a scholarship from Carl Tryggers Stiftelse to V.V.Z. (F.B.). T.J.G.E. thanks the European Research Council (ERC consolidator grant 817834); the Dutch Research Council (NWO-VICI grant V.C.192.016); Moore–Simons Project on the Origin of the Eukaryotic Cell (Simons Foundation 735925LPI, https://doi.org/10.46714/735925LPI); and the Marie Skłodowska-Curie ITN project SINGEK (H2020-MSCA-ITN-2015-615, 599 which provided funding for M.E.S. P.J.K. and V.Y. were funded by an Investigator Grant from the Gordon and Betty Moore Foundation (https://doi.org/10.37807/GBMF9201). The Pacific Ocean work was supported by GBMF3788 to A.Z.W. Sampling at the LMO station in the Baltic Sea was carried out by support from the Swedish Research Council VR and the marine strategic research programme EcoChange to J.P. Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala, part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. Cell sorting and whole genome amplification was performed at the Microbial Single Cell Genomics Facility (MSCG) at SciLifeLab. Computations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Projects SNIC 2019/3-305, SNIC 2020/15-58, SNIC 2021/5-50, Uppstore2018069. Finally, we thank Eunsoo Kim and Sally D. Warring for sharing peptide models from Palpitomonas bilix and Roombia truncata.

Author contributions

F.B. and J.G.W. conceived the study. For the Baltic samples, V.V.Z. and J.P. performed sampling and cell sorting; V.V.Z. and J.F.H.S. performed genome amplification and sequencing preparation. For the Pacific samples, C.P., S.W. and A.Z.W. conceived, developed and implemented sort protocols; performed sampling, cell sorting and sequencing; C.P. and A.Z.W. performed initial sequence analyses and phylogenetics. M.E.S. under the supervision of T.J.G.E. and F.B. performed assembly of SAGs and Co-SAGs, phylogenomic analyses, and searched for plastid evidence and gene transfers with the help of V.M. and P.I.K. R.P.S. and J.G.W. assembled mitochondrial genomes. M.E.S., F.B., and J.G.W. drafted the manuscript. T.J.G.E., P.K.J., A.Z.W., J.P., V.V.Z. and J.F.H.S. contributed edits to the manuscript. All authors read and approved the final version.

Funding

Open access funding provided by Uppsala University.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information

The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-26918-0.

Correspondence

Requests for materials should be addressed to Fabien Burki.

Peer reviewer information

Nature Communications thanks David Moreira and Hwan Su Yoon for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information

is available at http://www.nature.com/reprints

Publisher’s note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license 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) 2021