# Evolution of alternative biosynthetic pathways for vitamin C following plastid acquisition in photosynthetic eukaryotes

Glen Wheeler[1,] *, Takahiro Ishikawa[2], Varissa Pornsaksit[3], Nicholas Smirnoff[3] *

[1] Marine Biological Association, Plymouth, UK, PL1 2PB

[2] Department of Life Science and Biotechnology, Shimane University, Shimane 690-8504, Japan.

[3] Biosciences, College of Life and Environmental Sciences, University of Exeter, EX4 4QD

*correspondence to:

glw@mba.ac.uk

n.smirnoff@exeter.ac.uk

Short title: Evolution of vitamin C biosynthesis

**Abstract:**

Ascorbic acid (vitamin C) is an enzyme co-factor in eukaryotes that also plays a critical role in protecting photosynthetic eukaryotes against damaging reactive oxygen species derived from the chloroplast. Many animal lineages, including primates, have become ascorbate auxotrophs due to the loss of the terminal enzyme in their biosynthetic pathway, L-gulonolactone oxidase (GULO). The alternative pathways found in land plants and *Euglena* use a different terminal enzyme, L-galactonolactone dehydrogenase (GLDH). The evolutionary processes leading to these differing pathways and their contribution to the cellular roles of ascorbate remain unclear. Here we present molecular and biochemical evidence demonstrating that GULO was functionally replaced with GLDH in photosynthetic eukaryote lineages following plastid acquisition. GULO has therefore been lost repeatedly throughout eukaryote evolution. The formation of the alternative biosynthetic pathways in photosynthetic eukaryotes uncoupled ascorbate synthesis from hydrogen peroxide production and likely contributed to the rise of ascorbate as a major photoprotective antioxidant.

## Introduction:

Ascorbate (vitamin C) plays an essential role in eukaryotes as an enzyme co-factor in hydroxylation reactions, contributing to diverse processes such as the synthesis of collagen and the demethylation of histones and nucleic acids [1,2]. Ascorbate also plays an antioxidant role in eukaryotes to help protect against reactive oxygen species (ROS) derived from metabolic activity. The majority of hydrogen peroxide ($H_2O_2$) generated in some organelles is likely reduced by other antioxidant systems, such as the peroxiredoxins and glutathione peroxidases in the mitochondria, and catalases in the peroxisome [3,4]. However, ascorbate plays a role in protecting photosynthetic cells against ROS derived from the chloroplast [5]. Ascorbate peroxidase (APX), which is found in both the cytosol and the chloroplast of photosynthetic eukaryotes, is central to this photoprotective role. Thylakoid- and stroma-localised APX removes $H_2O_2$ produced by photosystem I through the activity of the ascorbate-glutathione cycle and this process may account for 10% of photosynthetic electron transport flow. Ascorbate also plays a critical role in preventing lipid peroxidation in the thylakoid membranes and acts as a co-factor for violaxanthin de-epoxidase in the xanthophyll cycle. In addition, ascorbate in the thylakoid lumen may prevent photoinhibition in high light by directly donating electrons to the photosynthetic electron transport chain [5]. These roles have been demonstrated in a range of ascorbate deficient plants that display sensitivity to high light and to oxidants[5].

Photosynthetic eukaryotes arose following the endosymbiotic acquisition of a cyanobacterial ancestor by a non-photosynthetic eukaryote in the Archaeplastida (Plantae) lineage [6]. Several other eukaryote lineages, including the diatoms, haptophytes and euglenids, subsequently gained plastids through a secondary endosymbiosis with either a red or green alga. These plastid endosymbioses were accompanied by lateral gene transfer on a massive scale from the symbiont to the host nuclear genome (known as endosymbiotic gene transfer or EGT), giving rise to the complex physiologies of photosynthetic eukaryotes (ref Timmis 2004). The plastids in the major photosynthetic eukaryote lineages are all ultimately derived from the primary endosymbiosis. Whilst acquisition of a photosynthetic endosymbiont may have been beneficial to the host cell in many ways, the plastid is also a major source of potentially damaging ROS [7]. There is evidence for extensive leakage of $H_2O_2$ out of plastids *via* aquaporins, particularly at high light intensities [8,9]. Plastid acquisition is therefore associated with a greatly increased requirement for cellular antioxidant systems to prevent photodamage. Cyanobacteria do not possess APX or any of the known enzymes for ascorbate biosynthesis

and minimise photo-oxidative stress using alternative mechanisms, such as peroxiredoxins, catalases and glutathione peroxidases [10,11]. This suggests that ascorbate was most likely recruited to its photoprotective role after the acquisition of the plastid in ancestral Archaeplastida [11], although the evolutionary origins of ascorbate biosynthesis are unclear.

There is no clear evidence for ascorbate biosynthesis in prokaryotes (see Materials and Methods), suggesting that the ability to synthesise ascorbate evolved in eukaryotes. The three eukaryote lineages in which ascorbate biosynthesis has been examined extensively (animals, plants and *Euglena*) all exhibit different biosynthetic pathways (Fig. 1). These pathways may have arisen due to convergent evolution, or may represent modifications of an ancestral pathway. An understanding of these evolutionary relationships will provide insight into the cellular roles of ascorbate in eukaryotes, particularly in relation to plastid acquisition in the photosynthetic lineages.

Animals synthesise ascorbate *via* D-glucuronic acid and L-gulonolactone (L-GulL), with L-gulonolactone oxidase (GULO) catalysing the oxidation of L-GulL to ascorbate. Many animal lineages have lost the ability to synthesise ascorbate, including haplorhine primates, guinea pigs, teleost fish, some bats and passerine birds. In all of the animal ascorbate auxotrophs that have been examined, the gene encoding GULO is lost [12-15]. GULO uses molecular $O_2$ as its electron acceptor, resulting in $H_2O_2$ production, which may have contributed to selective pressure to lose this enzyme in some animals [2,16]. The inability of many invertebrates to synthesise ascorbate led to early speculation that ascorbate synthesis may have evolved later in metazoan evolution [17]. However, whilst the loss of *GULO* in vertebrate lineages has been extensively examined [18], little is known about its distribution in invertebrates or the non-metazoan members of the holozoa. This information is required to identify the selective pressures underlying the evolution of ascorbate auxotrophy in animals.

Two alternative routes to ascorbate biosynthesis have been identified in photosynthetic eukaryotes, which both employ L-galactonolactone dehydrogenase (GLDH) as the terminal enzyme instead of GULO. A pathway *via* D-galacturonic acid and L-galactonolactone (L-GalL) was identified in *Euglena* [19]. This pathway is analogous to the animal pathway and also appears to be functional in some stramenopile algae [20,21]. In contrast, ascorbate biosynthesis in land plants was found to occur *via* a different route using D-mannose and L-galactose [22]. Green algae also use the 'plant pathway' [23,24], but evidence is lacking for the nature of

ascorbate biosynthesis in many other evolutionarily important lineages, most notably the rhodophytes (red algae).

This paper focuses on the distribution of the three major pathways of ascorbate biosynthesis described above. Alternative routes of ascorbate biosynthesis have been described in trypanosomes and also in the fungi, which synthesise a range of ascorbate analogues (see Materials and Methods)[25,26]. There is some evidence for the operation of alternative routes to ascorbate in land plants [27-29], although molecular genetic evidence from *Arabidopsis* indicates that the D-mannose/L-galactose pathway is the primary route of ascorbate biosynthesis (see Materials and Methods) [30,31].

The three major pathways of ascorbate biosynthesis therefore all utilise different routes to synthesise an aldonolactone precursor (L-gulonolactone, L-GulL or L-galactonolactone, L-GalL), which is converted to ascorbate by either GULO (animal pathway) or GLDH (plant and euglenid pathways) [19,22,25]. GULO and GLDH exhibit significant sequence similarity and are both members of the vanillyl alcohol oxidase (VAO) family of flavoproteins [32]. These similar enzymes exhibit important biochemical differences. GULO can oxidise L-GulL and L-GalL, whereas GLDH is highly specific for L-GalL [33]. GULO localises to the lumen of the endoplasmic reticulum (ER), whereas GLDH is associated with complex I in the mitochondrial electron transport chain [34]. Importantly, GLDH does not generate $H_2O_2$, as it uses cytochrome *c* rather than $O_2$ as an electron acceptor.

Despite the importance of ascorbate in eukaryote physiology, it is not known how the different pathways of ascorbate biosynthesis arose in animals, plants and algae or relate to its differing cellular roles. This manuscript examines the origins of ascorbate biosynthesis in eukaryotes and seeks to address the following important gaps in our current knowledge: 1) what is the wider distribution of *GULO* loss and ascorbate auxotrophy in the metazoa? 2) do all photosynthetic eukaryotes use an alternative terminal enzyme to animals? 3) why do two different pathways using GLDH exist in photosynthetic eukaryotes? 4) which pathway is used in the rhodophytes? Using a combination of molecular and biochemical analyses, we present evidence that *GULO* is an ancestral gene in eukaryotes that has been functionally replaced by *GLDH* in the photosynthetic lineages, resulting in the development of their alternative biosynthetic pathways.

## Results:

### *Distribution of GULO and GLDH in eukaryote genomes*

To examine the origins of ascorbate biosynthesis in eukaryotes we analysed the distribution of *GULO* and *GLDH* in eukaryote genomes. We found that *GULO* and *GLDH* have a mutually exclusive distribution (Fig 2). *GULO* is absent from many metazoan genomes, including all insects, supporting earlier biochemical evidence that insects are predominately ascorbate auxotrophs (Supplementary File 1) [17,35]. However, *GULO* is present in basally derived metazoans, including sponges and cnidarians, and is also present in a filasterean (*Capsaspora owczarzaki*) and in fungi (Supplementary File 1). This suggests that ascorbate synthesis *via* GULO is an ancestral trait in the Opisthokonta that has been lost in many lineages.

*GULO* is also present the Apusomonadida (*Thecamonas trahens*), a sister group to the Opisthokonts, and in members of the Amoebozoa and Excavata. Surprisingly, we also found *GULO* in basally derived Archaeplastida including the glaucophyte, *Cyanophora paradoxa*, and the rhodophytes *Galdieria sulphuraria* and *Galdieria phlegrea*. The glaucophytes occupy a key position in the evolution of photosynthetic eukaryotes as they diverged from the other Archaeplastida before the split of the red and green algal lineages and have highly unusual chloroplasts (termed cyanelles) that retain several features of the cyanobacterial endosymbiont [36]. *GULO* is absent from all other Archaeplastida genomes, although an enzyme family exhibiting weak similarity to *GULO* has been reported in *Arabidopsis* [37]. However, this enzyme forms a distinct phylogenetic clade from all other *GULO* and *GLDH* sequences and its role in *de novo* ascorbate biosynthesis remains unclear.

*GLDH* was found in all Archaeplastida genomes, except for *Cyanophora* and *Galdieria*, and in all photosynthetic lineages that have acquired a plastid *via* secondary endosymbiosis (including stramenopiles, cryptophytes, haptophytes, chlorarachniophytes and euglenids). *GLDH* was present in several non-photosynthetic organisms including the oomycetes, the foraminifera and in the choanoflagellates, *Monosiga brevicollis* and *Salpingoeca rosetta*. The evolutionary history of algal plastids acquired by secondary endosymbiosis remains uncertain and there is some evidence that non-photosynthetic stramenopile (e.g. oomycetes) and rhizarian (e.g. foraminifera) lineages may have once acquired a plastid that was subsequently lost [6,38,39].

Further identification of *GLDH* or *GULO* in the transcriptomes of 165 eukaryotes within the Marine Microbial Eukaryote Trancriptome dataset [40] confirmed that *GLDH* was found primarily in photosynthetic organisms (Supplementary File 2), but also in the non-photosynthetic stramenopiles such as oomycetes, biocosoecids and labryinthulids and in an acanthoecid choanoflagellate. The presence of *GULO* was restricted to non-photosynthetic organisms, including the heterotrophic flagellate *Palpitomonas bilix*, which is a non-photosynthetic relative of the Cryptophyte algae[41]. The exception was the presence of *GULO* in the chromerids, *Chromera velia* and *Vitrella brassicaformis,* which are photosynthetic relatives of the Apicomplexa [42].

Further searches of Expressed Sequence Tag (EST) and Transcriptome Shotgun Assembly (TSA) databases identified a *GULO* sequence in the green alga, *Chlorokybus atmophyticus* (JO192417.1)[43,44]. *Chlorokybus* represents a basal lineage in the charophyte algae, which are a sister group to the land plants (Figure 2: figure supplement 1). *GLDH* was identified in the other charophytes *Klebsormidium flaccidum, Nitella mirabilis* and *Nitella hyalina* (JO285109.1, JV744884.1, JO253095.1). These searches also revealed the presence of *GULO* in the craspedid choanoflagellate, *Monosiga ovata* (DC478225.1). The Craspedida subgroup of the choanoflagellates is divided into two major clades; clade I contains *M. brevicollis* and *S. rosetta* (which both possess *GLDH*) and clade II contains *M. ovata* [45].

We conclude that all photosynthetic eukaryotes use *GLDH* rather than *GULO* as the terminal enzyme in ascorbate biosynthesis. The exceptions are the basally derived Archaeplastida (*Cyanophora*, *Galdieria* and *Chlorokybus*) and the chromerids (Figure 2: figure supplement 2). No photosynthetic organisms were found to lack both *GULO* and *GLDH*, although both genes were absent in many non-photosynthetic organisms, including all insects, *Daphnia*, *Paramecium* and *Dictyostelium*. The genomes of many parasitic groups also appear to lack both enzymes including the diplomonads, parabasalids and apicomplexa (e.g. *Giardia intestinalis, Trichomonas vaginalis* and *Plasmodium falciparum).* Since all documented pathways of ascorbate biosynthesis require either GULO or GLDH, organisms lacking both of these enzymes are likely to be ascorbate auxotrophs [17].

***Phylogenetic analyses of GULO and GLDH***

A maximum likelihood tree of *GULO* and *GLDH* sequences was produced using the other members of the VAO family as an outgroup to root the tree. *GLDH* is highly conserved and the phylogenetic analyses strongly support a monophyletic origin for all *GLDH* sequences

(100% bootstrap support, posterior probability =1) (Fig. 3). The monophyly of eukaryote *GULO* sequences is well supported (85% bootstrap support, posterior probability =1). This clade includes trypanosome L-GalL oxidase and ascomycete D-arabinonolactone oxidase [26,46], indicating that although these enzymes exhibit altered substrate specificity they should be considered within the *GULO* clade for our evolutionary analyses. There is also moderate support for the monophyly of *GULO* and *GLDH* as a clade within the VAO family, supporting hypotheses that these enzymes may have originated from a gene duplication event. GLDH is highly unusual amongst the VAO flavoprotein family in that it does not use $O_2$ as an electron acceptor, but mutation of a single highly conserved alanine residue in GLDH is required to convert it from a dehydrogenase to an oxidase [47]. The phylogenies within the *GULO* clade and the *GLDH* clade are poorly resolved, and so the trees do not provide evidence on the likelihood of lateral gene transfer, such as endosymbiotic gene transfer (EGT) or horizontal gene transfer (HGT), of either gene. Further phylogenetic analyses, individually examining each gene using unrooted trees with much greater taxonomic sampling, were unable to provide greater resolution (Figure 3: figure supplement 1).

***Distribution of GLDH-dependent pathways in photosynthetic organisms***

Many of the enzymes preceding GULO or GLDH in the animal and euglenid pathways play other roles within the cell, e.g. in uronic acid metabolism or providing pentose intermediates [48]. The presence or absence of these genes is therefore not solely related to ascorbate biosynthesis. However, the plant pathway of ascorbate biosynthesis contains a number of dedicated enzyme steps, allowing a much clearer examination of its distribution. This also enables a distinction to be made between the plant- and euglenid-type pathways, as both utilise GLDH as the terminal enzyme. Plants and green algae use GDP-L-galactose phosphorylase (VTC2) and L-galactose dehydrogenase to generate L-GalL [23,31,49-51], whereas euglenids use D-galacturonate reductase (Fig. 1) [52]. We found that L-galactose dehydrogenase is present in all rhodophytes and Viridiplantae, except the prasinophytes *Ostreococcus* and *Micromonas* (see Materials and Methods). Sequences exhibiting similarity to L-galactose dehydrogenase were also found in the diatoms and in some metazoa, but as some of these species are ascorbate auxotrophs, it appears that this enzyme may play alternative metabolic roles. *VTC2* is found exclusively in the Viridiplantae (including *Chlorokybus*), indicating that the definitive 'plant' pathway is restricted to this lineage (Fig. 2; Supplementary Files 3-4). Biochemical evidence from euglenids and stramenopiles [19-21] suggests that organisms that

lack *VTC2* but possess *GLDH* are likely to operate a 'euglenid' pathway, with D-galacturonic acid acting as the precursor of L-GalL. However, it is not clear whether this is also the case in the rhodophytes. All rhodophytes possess a sequence that is highly similar to characterised L-galactose dehydrogenases from plants and bacteria [49,53] and L-galactose residues are a major constituent of red algal polysaccharides [54]. We therefore examined whether rhodophytes could synthesise ascorbate *via* a modified 'plant' pathway using an alternative route to L-galactose.

***Biochemical analysis of pathways in rhodophytes***

We used the macroalga, *Porphyra umbilicalis,* to examine whether rhodophytes can utilise L-galactose in ascorbate synthesis. We detected $NAD^+$-dependent L-galactose dehydrogenase activity in *Porphyra* thallus extracts (Fig. 4A). Feeding 10 mM L-galactose to *Porphyra* thallus slices increased the concentration of ascorbate (detected as dehydroascorbate by GC-MS) (Fig. 4B). As red algae synthesise GDP-L-galactose from GDP-D-mannose [55], rhodophytes likely use a modified 'plant pathway' to synthesise ascorbate, employing an unidentified enzyme activity to generate L-galactose from GDP-L-galactose instead of VTC2.

We then examined ascorbate biosynthesis in *Galdieria*, which differs from all other rhodophytes (Supplementary File 4) in that it possesses *GULO* rather than *GLDH*. L-galactose, L-GalL and L-GulL were all effective precursors of ascorbate in *G. sulphuraria* (Fig 4C), suggesting that L-galactose is converted to L-GalL, which may then be converted to ascorbate by GULO. A positional isotopic labelling approach indicated that label from D-[1-$^{13}$C]-glucose was incorporated primarily into carbon 1 (C1) of ascorbate (Fig. 4D, Figure 4: figure supplement 1). This labelling pattern is expected for the plant pathway, while the reduction of a uronic acid intermediate in the animal or euglenid pathways would result in the transfer of label from C1 of glucose into C6 of ascorbate/dehydroascorbate [25]. *G. sulphuraria* therefore uses a similar pathway to other rhodophytes, employing GULO instead of GLDH.

In combination, these data identify a clear difference between the Archaeplastida and the photosynthetic lineages that have acquired a plastid *via* secondary endosymbiosis. Whilst both groups use GLDH as the terminal enzyme for ascorbate synthesis, they differ in the route to L-GalL, The Archaeplastida generate L-GalL via L-galactose (without inversion of the carbon chain of glucose), whereas photosynthetic eukaryotes with secondary plastids synthesise L-GalL via D-galacturonate, resulting in inversion of the carbon chain.

253 ***Distribution of ascorbate-dependent antioxidant systems***

254 We have found that nearly all photosynthetic eukaryotes use GLDH to synthesise ascorbate,
255 suggesting that this distribution may be linked to the photoprotective role of ascorbate. We
256 therefore determined the distribution of ascorbate-dependent antioxidant mechanisms in
257 eukaryote genomes. Three main isoforms of APX are found in eukaryotes (APX, APX-R and
258 APX-CCX, a hybrid enzyme containing both ascorbate and cytochrome *c* peroxidase
259 domains) [10,56,57]. APX and APX-R are found in nearly all photosynthetic eukaryotes and in
260 the choanoflagellates, *M. brevicollis* and *S. rosetta*. *Euglena* and *Emiliania* do not possess
261 APX or APX-R, although both possess APX-CCX, which is also found in some fungi and in
262 *Capsaspora* (Fig. 5). The only photosynthetic eukaryote in this analysis that lacks any
263 isoform of APX was *Cyanophora paradoxa*. *Cyanophora* also lacks all of the remaining
264 enzymes of the plant ascorbate-glutathione cycle: monodehydroascorbate reductase
265 (MDHAR), dehydroascorbate reductase (DHAR) and glutathione reductase (GR). The
266 ascorbate-dependent xanthophyll cycle is not present in glaucophytes and so *Cyanophora*
267 does not require ascorbate for non-photochemical quenching (Fig. 5). Moreover, the cellular
268 concentration of ascorbate in *Cyanophora* is either very low or absent, as we could not detect
269 ascorbate in *Cyanophora* extracts using GC-MS (data not shown). It is possible that ascorbate
270 analogues are present that we could not identify. However, in combination with the lack of
271 the plant ascorbate-glutathione cycle and the xanthophyll cycle, we conclude that
272 *Cyanophora* is unlikely to rely on ascorbate to detoxify peroxides derived from
273 photosynthesis. *Cyanophora* does however contain several glutathione peroxidases,
274 peroxiredoxins and catalase, as well as a unique peroxidase (symerythrin) similar to
275 rubrerythrin of prokaryotes[58]. These data suggest that glaucophytes rely on alternative
276 mechanisms to detoxify peroxides derived from photosynthesis. As cyanobacteria also do not
277 appear to use ascorbate for photoprotection[59,60], the photoprotective role of ascorbate may
278 therefore have emerged in the Archaeplastida after the divergence of the glaucophytes.

279

## Discussion

### *Distribution of pathways in eukaryotes*

Ascorbate (vitamin C) is a very familiar metabolite to humans, so it is perhaps surprising that so many aspects of its biosynthesis and metabolism remain uncharacterised. The biosynthetic pathway of ascorbate in plants, which supplies the vast majority of ascorbate in the human diet, remained elusive for many years [22] and the major role of ascorbate in DNA demethylation emerged only very recently [1]. In order to better understand the cellular roles of ascorbate, we have examined the distribution of the three major pathways of ascorbate biosynthesis in eukaryotes. We identify that the Opisthonkonts (animals and fungi), the Amoebozoa and the non-photosynthetic representatives of the Excavata and CCTH (Hacrobia) use *GULO* for ascorbate biosynthesis. In contrast, the photosynthetic organisms in the Archaeplastida, CCTH (Hacrobia), SAR and the photosynthetic members of the Excavata (euglenids) use *GLDH*. In these photosynthetic organisms, the combination of molecular and biochemical evidence suggests that the non-inversion pathway *via* L-galactose (plant pathway) is restricted to Archaeplastida, whereas the inversion pathway *via* D-galacturonate (euglenid pathway) is used by photosynthetic eukaryotes that acquired plastids *via* secondary endosymbiosis. The important exceptions to these trends are: firstly, that *GLDH* is found in several non-photosynthetic organisms, notably in some choanoflagellates (Opisthokonts) and stramenopiles and secondly, that *GULO* is found in several basally derived members of the Archaeplastida.

### *The evolutionary origins of GULO and GLDH*

The processes underlying the distribution of the different terminal enzymes are therefore central to our understanding of the evolution of ascorbate biosynthesis. The mutually exclusive distribution of two highly conserved and functionally similar genes in eukaryotes may be explained by either of two evolutionary scenarios: an ancient gene duplication in the last common eukaryote ancestor (ancient paralogy) followed by differential loss of either gene, or lateral gene transfer of a novel gene followed by functional replacement of the ancestral gene [61]. It is likely that one of these evolutionary scenarios underlies the distribution of *GULO* and *GLDH* amongst eukaryotes (Fig 6).

The model of ancient paralogy requires that both genes were present in the last common eukaryote ancestor, where they both presumably contributed to ascorbate biosynthesis, and

were then differentially lost by every eukaryote lineage. This requires that these two functionally similar enzymes co-existed in multiple lineages throughout eukaryote evolution without significant functional divergence. The distribution of *GULO* and *GLDH* in the Archaeplastida and in the choanoflagellates suggests that these enzymes may have coexisted for a time in these lineages (Figure 2: figure supplement 1). However, there are no clear examples of extant eukaryotes that possess both enzymes, which would be expected if they have co-existed extensively throughout eukaryote evolution. The ancient paralogy model also requires extensive loss of both *GULO* and *GLDH*. There is clear evidence for multiple independent losses of *GULO* in animals and indications that GULO activity may be deleterious under certain conditions, providing a potential selective pressure for gene loss [16,62]. Similar evidence for the loss of *GLDH* in eukaryotes is lacking. In addition, *Arabidopsis* mutants that lack *GLDH* cannot correctly assemble mitochondrial complex I [63], suggesting that loss of *GLDH* is likely to have many wider impacts on metabolism.

The broad distribution of *GULO* supports an ancient evolutionary origin for this gene. It is present in all of the eukaryote supergroups, including basally-derived lineages within the CCTH (Hacrobia) and Archaeplastida and also in the Apusomonads. Although *GLDH* is also present in most of the eukaryote supergroups (except the Amoebozoa), its distribution is primarily restricted to lineages that have acquired a plastid or to isolated lineages (e.g. choanoflagellates). Whilst we cannot discount an ancient origin for *GLDH* in the last common eukaryote ancestor, its distribution may also be reasonably explained by the lateral gene transfer model.

In the lateral gene transfer scenario, either *GULO* or *GLDH* could represent a novel gene that arose in a specific lineage. However, the distribution of *GULO* cannot be reasonably explained by lateral gene transfer, as this requires horizontal gene transfer (HGT) on a massive scale specifically into non-photosynthetic eukaryotes. In contrast, the distribution of *GLDH* can be largely explained by endosymbiotic gene transfer (EGT) during plastid acquisition. *GLDH* may have arisen specifically in ancestral Archaeplastida after the divergence of the glaucophytes and functionally replaced the ancestral gene (*GULO*). *GLDH* could then have been transferred to the other photosynthetic lineages *via* EGT, resulting in the replacement of *GULO* in lineages that acquired their plastids *via* secondary endosymbiosis (Fig. 7). The presence of *GLDH* in some non-photosynthetic eukaryotes may be explained by the evolutionary acquisition of a plastid that was subsequently lost. For example, there is some evidence to support plastid loss in non-photosynthetic stramenopiles,

344 although the number and timing of plastid acquisition events *via* secondary endosymbiosis
345 remains a subject of significant debate[6]. The choanoflagellates have not acquired a plastid at
346 any stage, but there is evidence for large scale horizontal gene transfer (HGT) from algae into
347 this lineage, including the HGT of APX [64]. Choanoflagellates may therefore have acquired
348 *GLDH* *via* HGT along with these other algal genes. EGT of *GLDH* is a more parsimonious
349 scenario than ancient paralogy, as it requires fewer independent loss events. However, the
350 poor resolution of the phylogenies within the *GLDH* clade means that direct evidence for
351 either EGT or HGT between lineages is lacking.

352 *Evolution of alternative pathways*

353 Both the ancient paralogy and EGT evolutionary scenarios are plausible in the wider context
354 of ascorbate biosynthesis. However, the EGT scenario provides a clear rationale to explain
355 why photosynthetic eukaryotes with primary plastids exhibit a different pathway from those
356 with secondary plastids. Our biochemical evidence suggests that ancestral Archaeplastida
357 developed a non-inversion pathway *via* L-galactose that employed the broad specificity of
358 GULO to oxidise L-GalL. The development of *GLDH* in ancestral Archaeplastida would have
359 led to the eventual replacement of *GULO* in all red and green algal lineages, except *Galdieria*
360 and *Chlorokybus*, resulting in the non-inversion plant-type pathway found in extant
361 Archaeplastida. In the photosynthetic eukaryotes with secondary plastids, it is likely that the
362 host initially synthesised ascorbate via an animal-type pathway (involving inversion of chain
363 and GULO) and that the red or green algal symbiont used a plant-type pathway (involving
364 non-inversion of the carbon chain and GLDH). However, neither pathway appears to operate
365 in photosynthetic eukaryotes with secondary plastids, which instead use a euglenid-type
366 pathway. We propose that EGT of *GLDH* from the symbiont could have resulted in
367 functional replacement of *GULO* in the animal-type pathway of the host, leading to a hybrid
368 biosynthetic pathway that employed D-galacturonate rather than D-glucuronate as an
369 intermediate in order to provide L-GalL as a substrate for GLDH. The hybrid pathway
370 therefore involves inversion of the carbon chain of D-glucose and GLDH. The generation of a
371 hybrid pathway suggests that photosynthetic eukaryotes with secondary plastids only
372 acquired *GLDH* by EGT rather than the entire plant pathway.

373 In conclusion, the distribution of *GULO* and *GLDH* in eukaryotes may be explained by either
374 of two evolutionary models; ancient paralogy followed by differential gene loss or EGT of
375 *GLDH* followed by *GULO* loss. We favour the EGT scenario as the most parsimonious and

the most consistent with the biochemical evidence, but we cannot rule out either scenario based on the current evidence. Therefore our evolutionary analyses do not allow us to definitively identify the origin of *GLDH*. However, they do enable clear conclusions to be made on the loss of *GULO*. Both evolutionary models support *GULO* as an ancestral gene in the last common eukaryote ancestor, indicating that *GULO* has been lost in almost all photosynthetic eukaryotes. Therefore, we can conclude that photosynthetic eukaryotes encountered strong selective pressure to replace the function of *GULO* in ascorbate biosynthesis.

*Selective pressures underlying evolution of ascorbate biosynthesis*

The critical role of ascorbate in photoprotection has been demonstrated in a diversity of photosynthetic eukaryotes, including land plants, green algae, diatoms and euglenids [24,65-67]. Our analyses indicate that many photosynthetic eukaryotes possessed GULO prior to plastid acquisition, whereas almost all extant photosynthetic lineages use GLDH to synthesise ascorbate. The selective pressure to replace GULO in ascorbate biosynthesis following plastid acquisition could therefore be linked to the photoprotective role of ascorbate. One intriguing possibility is that the production of $H_2O_2$ by GULO may have limited the ability of the host cell to protect itself against ROS derived from the chloroplast.

Ancestral eukaryotes developed multiple antioxidant mechanisms to protect themselves from ROS derived from organelles such as the peroxisome and the mitochondria. However, the acquisition of a photosynthetic cyanobacterial endosymbiont in the Archaeplastida would have resulted in a greatly increased requirement for cellular antioxidants to protect the host cell from $H_2O_2$ secreted by the plastid. Ascorbate, synthesised in the host cell but not in the cyanobacterial endosymbiont, appears to have been recruited to this role after the divergence of the glaucophytes. The recruitment of ascorbate as a major cellular antioxidant in photosynthetic eukaryotes may have led to an increased requirement for ascorbate biosynthesis. However, ascorbate biosynthesis via GULO results in the production of $H_2O_2$ in the ER lumen. In mammalian cells, this results in a damaging depletion and oxidation of the glutathione pool when ascorbate synthesis is increased by feeding L-GulL [16,68]. Ancestral photosynthetic eukaryotes may have been unable to balance their increasing requirements for ascorbate biosynthesis with maintenance of the redox status within the ER, providing selective pressure to uncouple ascorbate biosynthesis from $H_2O_2$ production.

This hypothesis is consistent with the presence of *GULO* rather than *GLDH* in the glaucophytes. As glaucophytes do not appear to use ascorbate for photoprotection, ascorbate biosynthesis would not have been subjected to the same selective pressures as other photosynthetic eukaryotes. This rationale may also apply to the retention of *GULO* in *Galdieria*, which is likely to have both possessed both *GULO* and *GLDH*. *Galdieria* is photosynthetic and expresses a functional APX [69], but it is very sensitive to even moderate light intensities and grows primarily in an endolithic environment utilising heterotrophic carbon sources [70]. Thus, photo-oxidative stress in this environment may be minimal, reducing the selective pressure in *Galdieria* to replace *GULO*.

*Evolution of ascorbate auxotrophy in animals*

The evolution of vitamin auxotrophy underpins many important nutritional and ecological interactions between organisms [71]. The selective pressures resulting in *GULO* loss in animals represent a combination of the costs of ascorbate synthesis (including detoxification of $H_2O_2$ derived from GULO), the physiological requirements for ascorbate and the ecological factors that determine the supply of dietary ascorbate throughout their life cycle. The development of the photoprotective role of ascorbate in photosynthetic eukaryotes would have significantly altered its availability to many heterotrophic organisms. The leaves of land plants have particularly high cellular concentrations of ascorbate relative to other photosynthetic eukaryotes [11], which may result from their inability to remove intracellular $H_2O_2$ *via* diffusion to an aquatic medium. Our dataset reveals that in almost all documented cases of ascorbate auxotrophy in animals, the major source of dietary ascorbate derives from GLDH rather than GULO (Table 1). This is the case even for insectivorous animals, as insects appear to lack *GULO* and must also obtain ascorbate in their diet, primarily from land plants. Thus, the replacement of *GULO* with *GLDH* in photosynthetic organisms may have ultimately been an important contributory factor in the loss of *GULO* in many animal auxotrophs.

The pseudogenisation of *GULO* in primates, bats and guinea pigs is one of the best known examples of evolutionary gene loss [15]. Through a wider analysis of ascorbate biosynthesis, we have identified that *GULO* has also been lost in photosynthetic eukaryotes. Photosynthetic eukaryotes functionally replaced *GULO* with an alternative terminal enzyme, *GLDH*, which uncoupled ascorbate biosynthesis from $H_2O_2$ production and potentially aided the important photoprotective role of ascorbate. These developments in photosynthetic eukaryotes may

438 have ultimately contributed to the loss of *GULO* in many herbivorous animals, by influencing
439 their supply of dietary ascorbate.

440

## Materials and Methods

### *Bioinformatics*

A broad range of eukaryote genomes were selected for detailed analyses (Supplementary File 5). Sequence similarity searches were used to identify candidate genes involved in ascorbate biosynthesis. Genomic searches were initially performed using BLASTP with mouse or *Arabidopsis* proteins. All instances of protein absence were confirmed using TBLASTN against the genome with additional searches using sequences from closely related organisms. Sequence similarity searches of transcriptomic datasets were used to identify trends in the presence of ascorbate biosynthesis genes, but were not used to infer absence. Proteins recovered from sequence similarity searches were identified using a combination of BLAST score, manual inspection of conserved residues in multiple sequence alignments and their position in phylogenetic trees generated by both neighbour-joining and maximum likelihood method within the MEGA5 software package [72].

For detailed phylogenetic analysis of *GULO* and *GLDH*, multiple sequence alignments were generated using MUSCLE. Poorly aligned regions were removed by manual inspection and the alignments were further refined using GBLOCKS 0.91b to remove ambiguously aligned sites [73], resulting in an alignment of 263 amino acids. ProtTest [74] was used to determine the best substitution model (WAG with gamma and invariant sites) [75]. Maximum likelihood phylogenetic trees were generated using PhyML3.0 software with 100 bootstraps. Bayesian posterior probabilities were calculated using BEAST v1.8 [76], running for 10,000,000 generations, with a burn-in of 1,000,000 generations. As the phylogenetic relationships within the *GLDH* clade were not well resolved, further unrooted phylogenetic analyses of *GLDH* were performed using an individual multiple sequence alignment to allow more positions to be used.

### *Biochemical analyses of ascorbate metabolism in rhodophytes*

*Galdieria sulphararia* 074G was grown in Galdieria Medium (GM) (CCCryo, Potsdam-Golm, Germany) at 30ºC, light intensity 50 μmol $m^{-2}$ $s^{-1}$. *Porphyra umbilicalis* was collected from Maer Rocks, Exmouth, UK (50° 36′ 31.6″ N 3° 23′ 27.0″ W). L-Galactose dehydrogenase activity was measured in ammonium sulphate (50% saturation) precipitates of *Porphyra* thallus protein extracts [49]. To determine the impact of exogenous precursors on ascorbate, *Galdieria* cultures or slices of *Porphyra* thallus were incubated with sugars or

aldonic acid lactones (25 mM *Galdieria*, 10 mM *Porphyra*) for 24 h in GM or artificial sea water (Instant Ocean, Aquarium Systems, Sarrebourg, France). *Galdieria* cultures (15 ml) were harvested by centrifugation and extracted with 0.5 ml 80% methanol containing 0.1% formic acid using sonication in the presence of glass beads. *Porphyra* thallus was powdered in liquid nitrogen followed by homogenisation in 80% methanol (0.1 g thallus in 0.5 ml extractant). Homogenates were centrifuged (10 min at 16,000 g, 4 °C). Supernatants (100 µl) were dried into glass vials, methoximated trimethylsilyl derivatives were prepared [77] and analysed by accurate mass GC-EI-qToF MS (Agilent 7200, Agilent Technologies, Santa Clara, CA, USA). Derivatives were injected (0.2-0.4 µl, 1/2 to 1/50 split ratio) onto a Zebron SemiVolatiles GC column (30 m analytical + 10 m guard length, 0.25 mm internal diameter, 0.25 µm film thickness, Phenomenex, Macclesfield, UK) using He carrier gas (1.2 ml $min^{-1}$). Injector temperature was 250 °C and column temperature program was 70 °C for 4 min, followed by an increase to 310 °C at 15 °C/min. The column was held at the final temperature for 6 min. Compounds were fragmented at 70 eV and MS spectra were collected (50-600 amu at 5 spectra $s^{-1}$). Ascorbate is oxidised to dehydroascorbate (DHA) during derivatisation and DHA was identified by co-chromatography and comparison of accurate mass spectra with an ascorbate standard. The position of stable isotope incorporation from [1-$^{13}$C]-D-glucose and [6-$^{13}$C]-D-glucose into ascorbate was identified using mass spectra obtained from injection of [1-$^{13}$C]-ascorbate derivatives. C6 of DHA was found in fragments with m/z values of 157.046 and 245.1029, while m/z 316.1038 contained C1 and C6. $^{13}$C enrichment was assessed by the relative abundance of *m*+1 for each fragment relative to the $^{12}$C ascorbate standard.$^{13}$C-labelled compounds were obtained from Omicron Biochemicals (South Bend, IN, USA) and all other chemicals were from Sigma-Aldrich (Dorset, UK).

*Molecular analyses*

Reverse transcriptase PCR was used to verify the expression of *GULO* in *Cyanophora paradoxa* (CCAP 981/1) and confirm its coding sequence. RNA was prepared using the TRIzol method (Invitrogen) from *Cyanophora paradoxa* cultures grown in standard medium (MWC), at 20ºC, 16:8 light:dark, light intensity 50 µmol $m^{-2}$ $s^{-1}$. Reverse-transcriptase PCR was performed using a gene specific primer (GGAACTCCTCGAACTTGGGG) for reverse transcription, followed by amplification of a 1017 base pair region using the following PCR primers: GTCGCCCCTTCTGAGCATAG (forward) and CATGAGCGCGTCGAAGTCT (reverse). NCBI accession number KJ957823.

Euglena *transcriptome sequencing*

*Euglena gracilis* (strain Z) was grown in Koren-Hutner medium (KH) under continuous illumination (24 μmol $m^{-2}$ $s^{-1}$) at 26°C. RNA was harvested by the TRIzol method and used to prepare cDNA. Paired end reads were generated by Illumina sequencing technology resulting in a total of 193,472,913 reads. *De novo* assembly was carried out using Trinity [78], followed by further clustering with TGICL [79].

*Alternative routes for ascorbate biosynthesis*

The three major routes of ascorbate biosynthesis described in Figure 1 are well supported by biochemical and molecular evidence. However, there is some evidence to suggest that some classes of eukaryotes may use alternative routes to ascorbate or use multiple routes. These pathways are reviewed comprehensively elsewhere [25,51,80] but the implications for our findings are highlighted below.

*i) Land plants*

Ascorbate in land plants is synthesised predominately *via* a non-inversion pathway through GDP-D-mannose and L-galactose in which carbon atom 1 (C1) of the precursor hexose remains as C1 in ascorbate [25,81]. Genetic evidence from a range of ascorbate-deficient *Arabidopsis* mutants indicates that the D-mannose/L-galactose pathway is the primary route of ascorbate biosynthesis [5,30,31,82]. For example, *Arabidopsis vtc2 vtc5* double mutants lacking GDP-L-Gal phosphorylase activity are not *via*ble unless supplemented with ascorbate, suggesting no other route can supply sufficient ascorbate to rescue this defect [31]. However, a number of other routes to ascorbate in plants have been proposed. Plants can potentially use D-galacturonate as an ascorbate precursor, and this alternative route may contribute to ascorbate synthesis in certain tissue types, such as fruits [27,83]. GDP-L-gulose, formed during the activity of GDP-mannose epimerase *in vitro* [29], could provide L-GulL, which could contribute to ascorbate synthesis *via* the GULO-like enzymes found in land plants [37]. Overexpression of *Arabidopsis* GULO-like genes in tobacco cell lines increased cellular ascorbate concentration in the presence of exogenous L-GulL [37]. Oxidation of *myo*-inositol is another potential source of L-GulL [28,84]. However, more definitive biochemical and genetic evidence for these alternative pathways is required in order to assess whether these pathways contribute significantly to ascorbate biosynthesis in plants. The evidence for the alternative pathways is largely based on increased ascorbate production after ectopic expression of genes

or addition of exogenous substrates. This approach leaves open the possibility that these pathways are not active in wild type plants. The critical experiments, in which the proposed enzymes are mutated or knocked out, have not been reported.

Our evolutionary analyses indicate that *GULO* and *GLDH* co-existed in ancestral Archaeplastida, suggesting that multiple pathways of ascorbate biosynthesis were operational in these organisms. Whilst the 'plant pathway' *via* GLDH is clearly the primary biosynthetic route in extant land plants, the presence of minor alternative pathways such as those listed above may reflect this ancestry of shared pathways.

*ii) Prasinophytes*

The genomes of the prasinophyte algae *Ostreococcus* and *Micromonas* are unusual amongst the Viridiplantae in that they lack L-galactose dehydrogenase and GDP-mannose pyrophosphorylase [24]. However, it is likely that these prasinophytes can generate GDP-D-mannose, either through an unidentified alternative enzyme activity (as proposed in brown algae [85]) or through the transferase activity of VTC2. Furthermore, we identified L-fucose (6-deoxy-L-galactose) dehydrogenase in all *Ostreococcus* and *Micromonas* genomes. L-fucose dehydrogenase also exhibits activity with L-galactose [86] and it therefore may functionally replace L-galactose dehydrogenase in these prasinophytes. Analysis of the Marine Microbial Eukaryote Transcriptome dataset revealed that L-fucose dehydrogenase is present in many other prasinophytes including members of the Prasinococcales, Pycnococcaceae and the Pyramimonadaceae. The only prasinophyte identified with L-galactose dehydrogenase was *Nephroselmis pyriformis* (Nephroselmidophyceae). The Chlorodendrophyceae lineage containing *Tetraselmis* spp. also possess L-galactose dehydrogenase.

*iii) Trypanosomes*

GULO from trypanosomes exhibits activity with D-AraL or L-GalL but not L-GulL, and is therefore referred to as L-GalL oxidase [26,87,88]. The source of L-GalL in trypanosomes has not been determined. We did not find evidence for L-galactose dehydrogenase or the other enzymes in the D-mannose/L-galactose pathway in trypanosome genomes.

*iv) Fungi*

The fungi synthesise a range of ascorbate analogues, including 6-deoxy-L-ascorbate, ascorbate glycosides and the five carbon analogue, D-erythroascorbate [25]. Yeasts synthesise

D-erythroascorbate from D-arabinose *via* D-AraL and these final steps are therefore analogous to the D-mannose/L-galactose pathway. Deletion of *ALO1* encoding D-AraL oxidase in *Saccharomyces cerevisiae* results in increased sensitivity to oxidative stress but the mutants are still viable [46]. Until both the biosynthesis and the physiological roles of these analogues are better understood, it is difficult to understand how the biosynthesis of the fungal ascorbate analogues may have evolved.

*v) Prokaryotes*

There is little evidence to suggest that prokaryotes synthesise ascorbate *de novo*. *GULO* and *GLDH* are essentially absent from prokaryotes. Extensive sequence similarity searches identified a single *GULO* sequence in the cyanobacterium *Rivularia* sp PCC 7116 (WP_015122198.1), but all other cyanobacteria lack *GULO*. An enzyme exhibiting L-GulL dehydrogenase activity has been cloned from *Mycobacterium tuberculosis* and a similar enzyme is present in a range of other prokaryotes [89]. While the presence of this enzyme suggests some prokaryotes have the capacity for ascorbate synthesis, specific evidence for *in vivo* ascorbate biosynthesis is lacking. There is a single report indicating that some cyanobacteria contain very low concentrations of ascorbate [90], although the analytical method used (2,4-dinitrophenylhydrazine) may not be sufficiently specific for ascorbate at these concentrations [91]. Prokaryotes lack ascorbate peroxidase and it is likely that early reports of ascorbate peroxidase activity in cyanobacteria were due to the activity of other forms of peroxidase [59].

*vi) Chromerids*

The chromerids are photosynthetic relatives of the apicomplexa [42] and are the only organisms in our analyses that have acquired a plastid *via* secondary endosymbiosis and retained *GULO*. *Chromera velia* utilises the xanthophyll cycle for non-photochemical quenching [92], suggesting that ascorbate performs a photoprotective role in chromerids by acting as a co-factor for violaxanthin de-epoxidase (VDE). We found homologues of VDE and VDE-like proteins in the *C. velia* transcriptome, as well as monodehydroascorbate reductase (MDHAR). The presence of *GULO* rather than *GLDH* in *C. velia* may relate to the unusual mitochondria of apicomplexans. Apicomplexans lack complex I of the mitochondrial electron transport chain [93], which may have influenced their ability to acquire and/or utilise GLDH, as GLDH is incorporated into complex I in land plants [34]. However, the genomes of the parasitic

apicomplexa all appear to lack both *GULO* and *GLDH*. It should be noted that dinoflagellates also lack mitochondrial complex I but possess *GLDH* (Table S2).

## Acknowledgements

The authors acknowledge funding from the Exeter University Science Strategy Fund, Biotechnology and Biological Sciences Research Council (BB/G021678) and the Natural Environment Research Council (NE/J021008/1). Hannah Florance (Exeter University Mass Spectrometry Facility) provided analytical assistance. The *Euglena* transcriptome was sequenced by the Exeter University Sequencing Service with assistance from Konrad Paszkiewicz and Karen Moore. Access to the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) was provided by National Center for Genome Resources. This paper is dedicated to the memory of Frank A. Loewus (1919-2014) who pioneered research into ascorbate biosynthesis.

## Competing interests

The authors declare no competing interests.

## References


1 Blaschke, K. *et al.* Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* **500**, 222-226, doi:10.1038/nature12362 (2013).

2 Mandl, J., Szarka, A. & Banhegyi, G. Vitamin C: update on physiology and pharmacology. *British journal of pharmacology* **157**, 1097-1110, doi:10.1111/j.1476-5381.2009.00282.x (2009).

3 Mhamdi, A., Noctor, G. & Baker, A. Plant catalases: peroxisomal redox guardians. *Arch Biochem Biophys* **525**, 181-194, doi:10.1016/j.abb.2012.04.015 (2012).

4 Sies, H. Role of metabolic H2O2 generation: redox signaling and oxidative stress. *J Biol Chem* **289**, 8735-8741, doi:10.1074/jbc.R113.544635 (2014).

5 Smirnoff, N. Vitamin C: The Metabolism and Functions of Ascorbic Acid in Plants. *Adv Bot Res* **59**, 107-177, doi:Doi 10.1016/B978-0-12-385853-5.00003-9 (2011).

6 Keeling, P. J. The endosymbiotic origin, diversification and fate of plastids. *Philos Trans R Soc Lond B Biol Sci* **365**, 729-748, doi:10.1098/rstb.2009.0103 (2010).

7 Dorrell, R. G. & Howe, C. J. What makes a chloroplast? Reconstructing the establishment of photosynthetic symbioses. *J Cell Sci* **125**, 1865-1875, doi:10.1242/jcs.102285 (2012).

8 Mubarakshina, M. M. *et al.* Production and diffusion of chloroplastic H2O2 and its implication to signalling. *J Exp Bot* **61**, 3577-3587, doi:10.1093/jxb/erq171 (2010).

9 Naydov, I. A., Mubarakshina, M. M. & Ivanov, B. N. Formation kinetics and H2O2 distribution in chloroplasts and protoplasts of photosynthetic leaf cells of higher plants under illumination. *Biochemistry. Biokhimiia* **77**, 143-151, doi:10.1134/S0006297912020046 (2012).

10 Zamocky, M., Furtmuller, P. G. & Obinger, C. Evolution of structure and function of Class I peroxidases. *Arch Biochem Biophys* **500**, 45-57, doi:10.1016/j.abb.2010.03.024 (2010).

11 Gest, N., Gautier, H. & Stevens, R. Ascorbate as seen through plant evolution: the rise of a successful molecule? *J Exp Bot* **64**, 33-53, doi:10.1093/jxb/ers297 (2013).

12 Cui, J., Pan, Y. H., Zhang, Y., Jones, G. & Zhang, S. Progressive pseudogenization: vitamin C synthesis and its loss in bats. *Mol Biol Evol* **28**, 1025-1031, doi:10.1093/molbev/msq286 (2011).

13 Nishikimi, M., Fukuyama, R., Minoshima, S., Shimizu, N. & Yagi, K. Cloning and chromosomal mapping of the human nonfunctional gene for L-gulono-gamma-lactone oxidase, the enzyme for L-ascorbic acid biosynthesis missing in man. *J Biol Chem* **269**, 13685-13688 (1994).

14 Nishikimi, M., Kawai, T. & Yagi, K. Guinea pigs possess a highly mutated gene for L-gulono-gamma-lactone oxidase, the key enzyme for L-ascorbic acid biosynthesis missing in this species. *J Biol Chem* **267**, 21967-21972 (1992).

15 Drouin, G., Godin, J. R. & Page, B. The genetics of vitamin C loss in vertebrates. *Current genomics* **12**, 371-378, doi:10.2174/138920211796429736 (2011).

16 Banhegyi, G., Csala, M., Braun, L., Garzo, T. & Mandl, J. Ascorbate synthesis-dependent glutathione consumption in mouse liver. *FEBS Lett* **381**, 39-41 (1996).

17 Chatterjee, I. B. Evolution and the biosynthesis of ascorbic acid. *Science* **182**, 1271-1272 (1973).

18 Yang, H. Conserved or lost: molecular evolution of the key gene GULO in vertebrate vitamin C biosynthesis. *Biochemical genetics* **51**, 413-425, doi:10.1007/s10528-013-9574-0 (2013).

19 Shigeoka, S., Nakano, Y. & Kitaoka, S. The biosynthetic pathway of L-ascorbic acid in Euglena gracilis Z. *Journal of nutritional science and vitaminology* **25**, 299-307 (1979).

20 Grun, M. & Loewus, F. A. L-Ascorbic-acid biosynthesis in the euryhaline diatom Cyclotella cryptica. *Planta* **160**, 6-11, doi:10.1007/BF00392459 (1984).

21 Helsper, J. P., Kagan, L., Hilby, C. L., Maynard, T. M. & Loewus, F. A. l-Ascorbic Acid Biosynthesis in Ochromonas danica. *Plant Physiol* **69**, 465-468 (1982).
22 Wheeler, G. L., Jones, M. A. & Smirnoff, N. The biosynthetic pathway of vitamin C in higher plants. *Nature* **393**, 365-369, doi:10.1038/30728 (1998).
23 Running, J. A., Burlingame, R. P. & Berry, A. The pathway of L-ascorbic acid biosynthesis in the colourless microalga Prototheca moriformis. *J Exp Bot* **54**, 1841-1849, doi:10.1093/jxb/erg207 (2003).
24 Urzica, E. I. *et al.* Impact of oxidative stress on ascorbate biosynthesis in Chlamydomonas via regulation of the VTC2 gene encoding a GDP-L-galactose phosphorylase. *J Biol Chem* **287**, 14234-14245, doi:10.1074/jbc.M112.341982 (2012).
25 Loewus, F. A. Biosynthesis and metabolism of ascorbic acid in plants and of analogs of ascorbic acid in fungi. *Phytochemistry* **52**, 193-210, doi:Doi 10.1016/S0031-9422(99)00145-4 (1999).
26 Logan, F. J., Taylor, M. C., Wilkinson, S. R., Kaur, H. & Kelly, J. M. The terminal step in vitamin C biosynthesis in Trypanosoma cruzi is mediated by a FMN-dependent galactonolactone oxidase. *Biochem J* **407**, 419-426, doi:10.1042/BJ20070766 (2007).
27 Badejo, A. A. *et al.* Translocation and the alternative D-galacturonate pathway contribute to increasing the ascorbate level in ripening tomato fruits together with the D-mannose/L-galactose pathway. *J Exp Bot* **63**, 229-239, doi:10.1093/jxb/err275 (2012).
28 Lorence, A., Chevone, B. I., Mendes, P. & Nessler, C. L. myo-inositol oxygenase offers a possible entry point into plant ascorbate biosynthesis. *Plant Physiol* **134**, 1200-1205, doi:10.1104/pp.103.033936 (2004).
29 Wolucka, B. A. & Van Montagu, M. GDP-mannose 3',5'-epimerase forms GDP-L-gulose, a putative intermediate for the de novo biosynthesis of vitamin C in plants. *J Biol Chem* **278**, 47483-47490, doi:10.1074/jbc.M309135200 (2003).
30 Conklin, P. L. *et al.* Genetic evidence for the role of GDP-mannose in plant ascorbic acid (vitamin C) biosynthesis. *Proc Natl Acad Sci U S A* **96**, 4198-4203 (1999).
31 Dowdle, J., Ishikawa, T., Gatzek, S., Rolinski, S. & Smirnoff, N. Two genes in Arabidopsis thaliana encoding GDP-L-galactose phosphorylase are required for ascorbate biosynthesis and seedling viability. *Plant J* **52**, 673-689, doi:TPJ3266 [pii]

10.1111/j.1365-313X.2007.03266.x (2007).
32 Leferink, N. G., Heuts, D. P., Fraaije, M. W. & van Berkel, W. J. The growing VAO flavoprotein family. *Arch Biochem Biophys* **474**, 292-301, doi:10.1016/j.abb.2008.01.027 (2008).
33 Smirnoff, N. L-ascorbic acid biosynthesis. *Vitamins and hormones* **61**, 241-266 (2001).
34 Schertl, P. *et al.* L-galactono-1,4-lactone dehydrogenase (GLDH) forms part of three subcomplexes of mitochondrial complex I in Arabidopsis thaliana. *J Biol Chem* **287**, 14412-14419, doi:10.1074/jbc.M111.305144 (2012).
35 Dadd, R. H. Insect nutrition: current developments and metabolic implications. *Annual review of entomology* **18**, 381-420, doi:10.1146/annurev.en.18.010173.002121 (1973).
36 Price, D. C. *et al.* Cyanophora paradoxa genome elucidates origin of photosynthesis in algae and plants. *Science* **335**, 843-847, doi:10.1126/science.1213561 (2012).
37 Maruta, T. *et al.* The contribution of Arabidopsis homologs of L-gulono-1,4-lactone oxidase to the biosynthesis of ascorbic acid. *Bioscience, biotechnology, and biochemistry* **74**, 1494-1497 (2010).
38 Tyler, B. M. *et al.* Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* **313**, 1261-1266, doi:10.1126/science.1128796 (2006).
39 Glockner, G. *et al.* The Genome of the Foraminiferan Reticulomyxa filosa. *Curr Biol* **24**, 11-18, doi:10.1016/j.cub.2013.11.027 (2014).
40 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, doi:10.1371/journal.pbio.1001889 (2014).
41 Yabuki, A. *et al.* Palpitomonas bilix represents a basal cryptist lineage: insight into the character evolution in Cryptista. *Scientific reports* **4**, 4641, doi:10.1038/srep04641 (2014).
42 Janouskovec, J., Horak, A., Obornik, M., Lukes, J. & Keeling, P. J. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proc Natl Acad Sci U S A* **107**, 10949-10954, doi:10.1073/pnas.1003335107 (2010).
43 Leliaert, F. *et al.* Phylogeny and Molecular Evolution of the Green Algae. *Crit Rev Plant Sci* **31**, 1-46, doi:Doi 10.1080/07352689.2011.615705 (2012).
44 Timme, R. E., Bachvaroff, T. R. & Delwiche, C. F. Broad Phylogenomic Sampling and the Sister Lineage of Land Plants. *Plos One* **7**, doi:ARTN e29696

DOI 10.1371/journal.pone.0029696 (2012).
45 Jeuck, A., Arndt, H. & Nitsche, F. Extended phylogeny of the Craspedida (Choanomonada). *Eur J Protistol* **50**, 430-443, doi:10.1016/j.ejop.2014.06.001 (2014).
46 Huh, W. K. *et al.* Characterisation of D-arabinono-1,4-lactone oxidase from Candida albicans ATCC 10231. *European journal of biochemistry / FEBS* **225**, 1073-1079 (1994).
47 Leferink, N. G. *et al.* Identification of a gatekeeper residue that prevents dehydrogenases from acting as oxidases. *J Biol Chem* **284**, 4392-4397, doi:10.1074/jbc.M808202200 (2009).
48 Linster, C. L. & Van Schaftingen, E. Vitamin C. Biosynthesis, recycling and degradation in mammals. *The FEBS journal* **274**, 1-22, doi:10.1111/j.1742-4658.2006.05607.x (2007).
49 Gatzek, S., Wheeler, G. L. & Smirnoff, N. Antisense suppression of l-galactose dehydrogenase in Arabidopsis thaliana provides evidence for its role in ascorbate synthesis and reveals light modulated l-galactose synthesis. *Plant J* **30**, 541-553 (2002).
50 Laing, W. A., Wright, M. A., Cooney, J. & Bulley, S. M. The missing step of the L-galactose pathway of ascorbate biosynthesis in plants, an L-galactose guanyltransferase, increases leaf ascorbate content. *Proc Natl Acad Sci U S A* **104**, 9534-9539, doi:10.1073/pnas.0701625104 (2007).
51 Linster, C. L. *et al.* Arabidopsis VTC2 encodes a GDP-L-galactose phosphorylase, the last unknown enzyme in the Smirnoff-Wheeler pathway to ascorbic acid in plants. *J Biol Chem* **282**, 18879-18885, doi:10.1074/jbc.M702094200 (2007).
52 Ishikawa, T. *et al.* Functional characterization of D-galacturonic acid reductase, a key enzyme of the ascorbate biosynthesis pathway, from Euglena gracilis. *Bioscience, biotechnology, and biochemistry* **70**, 2720-2726 (2006).
53 Hobbs, M. E., Williams, H. J., Hillerich, B., Almo, S. C. & Raushel, F. M. l-Galactose metabolism in Bacteroides vulgatus from the human gut microbiota. *Biochemistry* **53**, 4661-4670, doi:10.1021/bi500656m (2014).
54 Percival, E. Polysaccharides of Green, Red and Brown Seaweeds - Their Basic Structure, Biosynthesis and Function. *Brit Phycol J* **14**, 103-117 (1979).
55 Su, J. C. & Hassid, W. Z. Carbohydrates and nucleotides in the red alga Porphyra perforata. I. Isolation and identification of carbohydrates. *Biochemistry* **1**, 468-474 (1962).
56 Fawal, N. *et al.* PeroxiBase: a database for large-scale evolutionary analysis of peroxidases. *Nucleic Acids Res* **41**, D441-444, doi:10.1093/nar/gks1083 (2013).
57 Lazzarotto, F. *et al.* Ascorbate peroxidase-related (APx-R) is a new heme-containing protein functionally associated with ascorbate peroxidase but evolutionarily divergent. *New Phytologist* **191**, 234-250, doi:10.1111/j.1469-8137.2011.03659.x (2011).
58 Cooley, R. B., Arp, D. J. & Karplus, P. A. Symerythrin structures at atomic resolution and the origins of rubrerythrins and the ferritin-like superfamily. *J Mol Biol* **413**, 177-194, doi:10.1016/j.jmb.2011.08.019 (2011).
59 Bernroitner, M., Zamocky, M., Furtmuller, P. G., Peschek, G. A. & Obinger, C. Occurrence, phylogeny, structure, and function of catalases and peroxidases in cyanobacteria. *J Exp Bot* **60**, 423-440, doi:10.1093/jxb/ern309 (2009).

60 Latifi, A., Ruiz, M. & Zhang, C. C. Oxidative stress in cyanobacteria. *FEMS microbiology reviews* **33**, 258-278, doi:10.1111/j.1574-6976.2008.00134.x (2009).
61 Keeling, P. J. & Inagaki, Y. A class of eukaryotic GTPase with a punctate distribution suggesting multiple functional replacements of translation elongation factor 1alpha. *Proc Natl Acad Sci U S A* **101**, 15380-15385, doi:10.1073/pnas.0404505101 (2004).
62 Hiller, M. *et al.* A "forward genomics" approach links genotype to phenotype using independent phenotypic losses among related species. *Cell reports* **2**, 817-823, doi:10.1016/j.celrep.2012.08.032 (2012).
63 Pineau, B., Layoune, O., Danon, A. & De Paepe, R. L-galactono-1,4-lactone dehydrogenase is required for the accumulation of plant respiratory complex I. *J Biol Chem* **283**, 32500-32505, doi:10.1074/jbc.M805320200 (2008).
64 Nedelcu, A. M., Miles, I. H., Fagir, A. M. & Karol, K. Adaptive eukaryote-to-eukaryote lateral gene transfer: stress-related genes of algal origin in the closest unicellular relatives of animals. *Journal of evolutionary biology* **21**, 1852-1860, doi:10.1111/j.1420-9101.2008.01605.x (2008).
65 Ishikawa, T. *et al.* Euglena gracilis ascorbate peroxidase forms an intramolecular dimeric structure: its unique molecular characterization. *Biochem J* **426**, 125-134, doi:10.1042/BJ20091406 (2010).
66 Lavaud, J. & Kroth, P. G. In diatoms, the transthylakoid proton gradient regulates the photoprotective non-photochemical fluorescence quenching beyond its control on the xanthophyll cycle. *Plant Cell Physiol* **47**, 1010-1016, doi:10.1093/pcp/pcj058 (2006).
67 Mellado, M., Contreras, R. A., Gonzalez, A., Dennett, G. & Moenne, A. Copper-induced synthesis of ascorbate, glutathione and phytochelatins in the marine alga Ulva compressa (Chlorophyta). *Plant physiology and biochemistry : PPB / Societe francaise de physiologie vegetale* **51**, 102-108, doi:10.1016/j.plaphy.2011.10.007 (2012).
68 Puskas, F. *et al.* Gulonolactone oxidase activity-dependent intravesicular glutathione oxidation in rat liver microsomes. *FEBS Lett* **430**, 293-296 (1998).
69 Sano, S. *et al.* Characterization of ascorbate peroxidases from unicellular red alga Galdieria partita. *Plant Cell Physiol* **42**, 433-440 (2001).
70 Oesterhelt, C., Schmalzlin, E., Schmitt, J. M. & Lokstein, H. Regulation of photosynthesis in the unicellular acidophilic red alga Galdieria sulphuraria. *Plant J* **51**, 500-511, doi:10.1111/j.1365-313X.2007.03159.x (2007).
71 Helliwell, K. E., Wheeler, G. L. & Smith, A. G. Widespread decay of vitamin-related pathways: coincidence or consequence? *Trends in genetics : TIG* **29**, 469-478, doi:10.1016/j.tig.2013.03.003 (2013).
72 Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731-2739, doi:10.1093/molbev/msr121 (2011).
73 Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic biology* **56**, 564-577, doi:10.1080/10635150701472164 (2007).
74 Abascal, F., Zardoya, R. & Posada, D. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104-2105, doi:10.1093/bioinformatics/bti263 (2005).
75 Whelan, S. & Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* **18**, 691-699 (2001).
76 Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**, 1969-1973, doi:10.1093/molbev/mss075 (2012).
77 Lisec, J., Schauer, N., Kopka, J., Willmitzer, L. & Fernie, A. R. Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nature protocols* **1**, 387-396, doi:10.1038/nprot.2006.59 (2006).

78 Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols* **8**, 1494-1512, doi:10.1038/nprot.2013.084 (2013).
79 Pertea, G. *et al.* TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* **19**, 651-652 (2003).
80 Smirnoff, N. Ascorbate biosynthesis and function in photoprotection. *Philos Trans R Soc Lond B Biol Sci* **355**, 1455-1464, doi:10.1098/rstb.2000.0706 (2000).
81 Loewus, F. A. & Jang, R. The conversion of C14-labeled sugars to L-ascorbic acid in ripening strawberries. II. Labeling patterns in the free sugars. *J Biol Chem* **232**, 505-519 (1958).
82 Conklin, P. L. *et al.* Arabidopsis thaliana VTC4 encodes L-galactose-1-P phosphatase, a plant ascorbic acid biosynthetic enzyme. *J Biol Chem* **281**, 15662-15670, doi:10.1074/jbc.M601409200 (2006).
83 Isherwood, F. A. & Mapson, L. W. Biological synthesis of ascorbic acid: the conversion of derivatives of D-galacturonic acid into L-ascorbic acid by plant extracts. *Biochem J* **64**, 13-22 (1956).
84 Lisko, K. A. *et al.* Ontogenetic changes in vitamin C in selected rice varieties. *Plant Physiol Bioch* **66**, 41-46, doi:DOI 10.1016/j.plaphy.2013.01.016 (2013).
85 Michel, G., Tonon, T., Scornet, D., Cock, J. M. & Kloareg, B. Central and storage carbon metabolism of the brown alga Ectocarpus siliculosus: insights into the origin and evolution of storage carbohydrates in Eukaryotes. *New Phytol* **188**, 67-81, doi:10.1111/j.1469-8137.2010.03345.x (2010).
86 Schachter, H., Sarney, J., McGuire, E. J. & Roseman, S. Isolation of diphosphopyridine nucleotide-dependent L-fucose dehydrogenase from pork liver. *J Biol Chem* **244**, 4785-4792 (1969).
87 Kudryashova, E. V., Leferink, N. G., Slot, I. G. & van Berkel, W. J. Galactonolactone oxidoreductase from Trypanosoma cruzi employs a FAD cofactor for the synthesis of vitamin C. *Biochim Biophys Acta* **1814**, 545-552, doi:10.1016/j.bbapap.2011.03.001 (2011).
88 Wilkinson, S. R., Prathalingam, S. R., Taylor, M. C., Horn, D. & Kelly, J. M. Vitamin C biosynthesis in trypanosomes: a role for the glycosome. *Proc Natl Acad Sci U S A* **102**, 11645-11650, doi:10.1073/pnas.0504251102 (2005).
89 Wolucka, B. A. & Communi, D. Mycobacterium tuberculosis possesses a functional enzyme for the synthesis of vitamin C, L-gulono-1,4-lactone dehydrogenase. *The FEBS journal* **273**, 4435-4445, doi:10.1111/j.1742-4658.2006.05443.x (2006).
90 Telor, E., Huflejt, M. E. & Packer, L. Hydroperoxide Metabolism in Cyanobacteria. *Archives of Biochemistry and Biophysics* **246**, 396-402, doi:Doi 10.1016/0003-9861(86)90485-6 (1986).
91 Roe, J. H. Comparative analyses for ascorbic acid by the 2,4-dinitrophenylhydrazine method with the coupling reaction at different temperatures: a procedure for determining specificity. *J Biol Chem* **236**, 1611-1613 (1961).
92 Kotabova, E., Kana, R., Jaresova, J. & Prasil, O. Non-photochemical fluorescence quenching in Chromera velia is enabled by fast violaxanthin de-epoxidation. *FEBS Lett* **585**, 1941-1945, doi:10.1016/j.febslet.2011.05.015 (2011).
93 Danne, J. C., Gornik, S. G., Macrae, J. I., McConville, M. J. & Waller, R. F. Alveolate mitochondrial metabolic evolution: dinoflagellates force reassessment of the role of parasitism as a driver of change in apicomplexans. *Mol Biol Evol* **30**, 123-139, doi:10.1093/molbev/mss205 (2013).
94 Yoon, H. S., Hackett, J. D. & Bhattacharya, D. A genomic and phylogenetic perspective on endosymbiosis and algal origin. *Journal of Applied Phycology* **18**, 475-481, doi:10.1007/s10811-006-9054-6 (2006).
95 Johnson, M. D., Oldach, D., Delwiche, C. F. & Stoecker, D. K. Retention of transcriptionally active cryptophyte nuclei by the ciliate Myrionecta rubra. *Nature* **445**, 426-428, doi:10.1038/nature05496 (2007).

96 Adler, L. N., Gomez, T. A., Clarke, S. G. & Linster, C. L. A novel GDP-D-glucose phosphorylase involved in quality control of the nucleoside diphosphate sugar pool in Caenorhabditis elegans and mammals. *J Biol Chem* **286**, 21511-21523, doi:10.1074/jbc.M111.238774 (2011).
97 Chan, C. X. *et al.* Porphyra (Bangiophyceae) Transcriptomes Provide Insights Into Red Algal Development And Metabolism. *Journal of Phycology* **48**, 1328-1342, doi:DOI 10.1111/j.1529-8817.2012.01229.x (2012).
98 Chan, C. X. *et al.* Red and Green Algal Monophyly and Extensive Gene Sharing Found in a Rich Repertoire of Red Algal Genes. *Current Biology* **21**, 328-333, doi:DOI 10.1016/j.cub.2011.01.037 (2011).
99 Milton, K. & Jenness, R. Ascorbic-Acid Content of Neotropical Plant-Parts Available to Wild Monkeys and Bats. *Experientia* **43**, 339-342, doi:Doi 10.1007/Bf01945577 (1987).
100 Birney, E. C., Jenness, R. & Ayaz, K. M. Inability of bats to synthesise L-ascorbic acid. *Nature* **260**, 626-628 (1976).
101 Cui, J., Yuan, X., Wang, L., Jones, G. & Zhang, S. Recent loss of vitamin C biosynthesis ability in bats. *PLoS One* **6**, e27114, doi:10.1371/journal.pone.0027114 (2011).
102 Dabrowski, K. Gulonolactone oxidase is missing in teleost fish. The direct spectrophotometric assay. *Biol Chem Hoppe Seyler* **371**, 207-214 (1990).
103 Desjardins, L. M., Castell, J. D. & Kean, J. C. Synthesis of Dehydroascorbic Acid by Subadult Lobsters (Homarus-Americanus). *Can J Fish Aquat Sci* **42**, 370-373, doi:Doi 10.1139/F85-048 (1985).
104 Hapette, A. M. & Poulet, S. A. Application of High-Performance Liquid-Chromatography to the Determination of Ascorbic-Acid in Marine Plankton. *J Liq Chromatogr* **13**, 357-370, doi:Doi 10.1080/01483919008049549 (1990).
105 Pierre, L. L. Synthesis of ascorbic acid by the normal fatbody of the cockroach, Leucophaea maderae (F.), and by its symbionts. *Nature* **193**, 904-905 (1962).

## Figure Legends

### Figure 1: Major ascorbate biosynthetic pathways in eukaryotes

The scheme depicts the three major ascorbate biosynthetic pathways found in eukaryotes [19,22,48]. The plant pathway (also known as the Smirnoff-Wheeler or D-mannose/L-galactose pathway) involves no inversion of the carbon chain (i.e. C1 of D-glucose becomes C1 of L-ascorbate), whereas the euglenid and animal pathways involve inversion of the carbon chain in the conversion from uronic acid to aldonolactone (i.e. C1 of D-glucose becomes C6 of L-ascorbate). Our analyses focus on enzymes with a dedicated role in ascorbate biosynthesis (shown in red): GULO – L-GulL oxidase; VTC2 - GDP-L-galactose phosphorylase; VTC4 – L-galactose-1-phosphate phosphatase; L-galDH – L-galactose dehydrogenase; GLDH – L-GalL dehydrogenase. The other enzymes are: PGM - phosphoglucomutase; UGP - UDP-D-glucose pyrophosphorylase; UGDH - UDP-D-glucose dehydrogenase; UGUR - UDP-glucuronidase; GlcUAR - D-glucuronate reductase; SMP30 – regucalcin/lactonase; GAE - UDP-D-glucuronate-4-epimerase; GalUAR - D-galacturonate reductase. Enzyme names are not listed for steps where multiple enzymes may be involved or where specific enzymes have not been identified.

### Figure 2: Coulson plot indicating the taxonomic distribution of the different ascorbate pathways

40 eukaryote genomes were analysed for the presence of genes in the ascorbate biosynthetic pathways. The two potential terminal enzymes in the pathway are boxed. *GLDH* is common to both the 'plant' and 'euglenid' type pathways. A schematic tree depicts the currently accepted phylogenetic relationships between organisms. The predicted route of ascorbate biosynthesis in each organism is shown. Note that 'euglenid' and 'rhodophyte' type pathways cannot currently be distinguished from sequence analysis alone and the predictions are based on biochemical evidence. Asterisk denotes a genome assembly was not available for *Euglena gracilis* and its transcriptome was analysed (methods). Grey circles in VTC4 represent the presence of a highly similar enzyme, *myo*-inositol-1-phosphate phosphatase that exhibits L-galactose-1-phosphatase activity. GULO in trypanosomes and yeasts acts to oxidise the alternative substrates L-galactonolactone or D-arabinonolactone respectively. VTC3 is not a biosynthetic enzyme, but represents a dual function Ser/Thr protein kinase/protein

phosphatase 2C that may play a regulatory role in the plant pathway. Black cross represents a pseudogene encoding a non-functional enzyme. Organisms with sequenced genomes that were found to lack both of the terminal enzymes in the known pathways (GULO and GLDH) are likely to be ascorbate auxotrophs and were not included in the plot. These include *Giardia intestinalis*, *Trichomonas vaginalis*, *Entamoeba invadens*, *Plasmodium falciparum* and *Perkinsus marinus*.

**Figure 2: figure supplement 1: Distribution of GULO and GLDH in the Archaeplastida**

A schematic tree demonstrating the currently accepted phylogenetic positions of the major lineages in the Archaeplastida [43,94]. The presence of either *GULO* or *GLDH* in representatives of each lineage is shown. The boxes denote the Viridiplantae (green), Rhodophyta (red) and Glaucophyta.

**Figure 2: figure supplement 2: Distribution of the different ascorbate pathways**

The schematic tree summarises the distribution of the two terminal enzymes in ascorbate biosynthesis, along with VTC2, the first committed enzyme in plant pathway. Blue lines indicate photosynthetic lineages derived by the primary endosymbiosis (of a cyanobacterium). Red or green lines indicate lineages that have become photosynthetic following a secondary endosymbiosis event with either a red or a green alga respectively. It should be noted that the timing and origin of many secondary endosymbioses remain unclear, particularly within the SAR supergroup where several non-photosynthetic lineages within the stramenopiles, alveolates and even rhizaria may potentially have lost an ancestral plastid. *GULO* is found in basally derived lineages of the Archaeplastida, Excavata, Opisthokonta, Amoebozoa and the CCTH group. In contrast, *GLDH* is found predominately in photosynthetic eukaryotes, although it is also found in non-photosynthetic stramenopiles and rhizaria and also in some choanoflagellates. Lineages where there is biochemical evidence determining inversion or non-inversion of the carbon chain in the conversion from D-glucose to ascorbate are shown.

**Figure 3: Phylogenetic analysis of L-gulonolactone oxidase and L-galactonolactone dehydrogenase**

A maximum likelihood phylogenetic tree demonstrating the relationships between aldonolactone oxidoreductases involved in ascorbate biosynthesis. A multiple sequence alignment of 263 amino acid residues was used with alditol oxidases from the vanillyl alcohol oxidase (VAO) family acting as the outgroup. Photosynthetic organisms are shown in green. There is strong support for a monophyletic origin for *GLDH* in eukaryotes. Bootstrap values >80% are shown above nodes (100 bootstraps) and Bayesian posterior probabilities >0.95 are shown below (10000000 generations), except for selected key nodes (circled) where all values are displayed.

**Figure 3: figure supplement 1: Phylogenetic analysis of L-galactonolactone dehydrogenase**

An unrooted maximum likelihood phylogenetic tree of *GLDH*. To improve resolution of *GLDH* phylogeny, an individual phylogenetic analysis was performed using a larger alignment (302 amino acids) with greater taxonomic sampling (151 sequences), although relationships between major taxonomic groups remain poorly resolved. Bootstrap values >70% are shown (100 bootstraps).

**Figure 4: Biochemical evidence for a modified D-mannose/L-galactose pathway in rhodophytes**

A) Crude extracts of *Porphyra umbilicalis* thallus demonstrate L-galactose dehydrogenase activity using 5 mM L-galactose (L-Gal) as a substrate. No activity was demonstrated with 5 mM L-fucose (6-deoxy-L-galactose) as a substrate. The result is representative of three different enzyme preparations. B) Feeding 10 mM L-Gal to *Porphyra* thallus for 24 h resulted in an accumulation of ascorbate (detected as dehydroascorbate - DHA). D-mannose (10 mM) did not cause an increase in ascorbate in *Porphyra*, but exogenous D-mannose does not elevate ascorbate in land plants even though it is an intermediate in ascorbate biosynthesis. The bar chart shows mean peak areas of selected fragments (± s.d.). n=3. C) Feeding ascorbate precursors (25 mM) to *Galdieria sulphararia* from both the plant and animal pathways results in increased cellular ascorbate (detected as dehydroascorbate using GC-MS) (± s.d.). The extent of the increase in cellular ascorbate is influenced by the rate of conversion of the intermediate and the rate of its uptake into the cell. n=3. D) Feeding D-[1-$^{13}$C]-glucose (25 mM) to *Galdieria sulphararia* results in enrichment of $^{13}$C in the 316/317 m/z fragment

of dehydroascorbate (which includes C1), but not in the 245/246 m/z or 157/158 m/z fragment (which exclude C1) (± s.d.). In contrast, feeding D-[6-$^{13}$C]-glucose (25 mM) labels all fragments, suggesting that they all include C6. In combination, this labelling pattern indicates plant-like non-inversion of the carbon chain in the conversion of hexoses to ascorbate. n=3.

**Figure 4: figure supplement 1: Positional isotopic labelling of ascorbate biosynthesis**

Analysis of ascorbate by GC-MS. Ascorbate is oxidised to dehydroascorbate during the derivatisation process and representative accurate mass spectra are shown. Analysis of a L-[1-$^{13}$C]-ascorbate standard indicates that the m/z 316 fragment of dehydroascorbate contains C1 whilst other fragments (m/z 157 and 245) do not. Dehydroascorbate from *G. sulphuraria* extracts exhibits an identical retention time and mass spectra to that of the ascorbate standard.

**Figure 5: Coulson Plot showing the distribution of photoprotective ascorbate-dependent enzymes**

Eukaryote genomes were analysed for the presence of enzymes from the plant ascorbate-glutathione cycle, the xanthophyll cycle and other ascorbate-dependent enzymes. We found that eukaryotes possess two distinct isoforms of GSH reductase. PeroxiBase was used to distinguish between the different forms of ascorbate peroxidase [56]. Boxes highlight the terminal enzymes in the biosynthetic pathway and the ascorbate peroxidase family (APX, APX-R and APX-CCX). MDHAR - monodehydroascorbate reductase; DHAR - dehydroascorbate reductase; GR-I - glutathione reductase isoform I; GR-II - glutathione reductase isoform II; APX - ascorbate peroxidase; APX-R - ascorbate peroxidase-related; APX-CCX - hybrid ascorbate peroxidase/cytochrome c peroxidase; VDE - violaxanthin de-epoxidase; VDE-like - violaxanthin de-epoxidase like; AO - ascorbate oxidase.

**Figure 6: Evolutionary scenarios for *GULO* and *GLDH***

The scheme illustrates two most likely evolutionary scenarios responsible for the distribution of *GULO* and *GLDH* in eukaryotes. In the ancient paralogy scenario, an ancient gene duplication in the last common eukaryote ancestor results in the presence of two functionally similar genes, *GULO* and *GLDH*, followed by differential loss of either gene in each lineage.

In the endosymbiotic gene transfer (EGT) scenario, *GULO* represents the ancestral gene and *GLDH* represents a novel gene that arose in a specific lineage. EGT of *GLDH* (red dashed arrow) to other photosynthetic lineages (green ovals) enables functional replacement of the ancestral gene. Note that *GULO* represents an ancestral gene in both of these evolutionary scenarios.

**Figure 7: A proposed evolutionary model of ascorbate biosynthesis**

The scheme illustrates the proposed events in the EGT evolutionary model of eukaryote ascorbate biosynthesis. In this scenario, ancestral eukaryotes synthesised ascorbate *via* GULO. *GLDH* arose in the Archaeplastida following primary endosymbiosis of a cyanobacterium, after the divergence of the glaucophyte lineage. *GLDH* functionally replaced *GULO* in the red and green algal lineages, coinciding with the rise of the photoprotective role of ascorbate. Plastid acquisition *via* secondary endosymbiosis of either a green or red alga resulted in endosymbiotic gene transfer of *GLDH* and replacement of *GULO*. As these organisms became the dominant primary producers in many ecosystems, a series of trophic interactions (dotted lines) resulted in the loss of *GULO* in non-photosynthetic organisms, either by providing a ready supply of dietary ascorbate (resulting in ascorbate auxotrophy in heterotrophic organisms) or through putative horizontal gene transfer of *GLDH* (e.g. choanoflagellates). For clarity, not all potential trophic interactions are shown.

**Table 1: Dietary sources of ascorbate in animal ascorbate auxotrophs**

Major sources of dietary ascorbate were identified in known animal auxotrophs. This information allows us to assess which terminal enzyme contributed to the production of dietary ascorbate. In nearly all cases the major source of dietary ascorbate is most likely to have been derived from GLDH. Phylogenetic analyses suggest GULO has been lost on multiple independent occasions throughout the Chiroptera (bats). Although ancestral bats may have been primarily insectivores, various sources of dietary ascorbate may have contributed to GULO loss. The passerine birds that are unable to synthesise ascorbate are primarily herbivores or insectivores. However, some members of the *Lanius* genus (shrikes) feed also on small vertebrates, in addition to insects. Most teleost fish are believed to be ascorbate auxotrophs due to loss of GULO. As zooplankton (primarily crustacea) are also ascorbate auxtrophs, phytoplankton are likely to be the ultimate source of dietary ascorbate. Reports suggest the ability of crustacea to synthesise ascorbate is either absent or very weak, although the taxonomic sampling and currently available genomic resources are limited. Most, but not all, phytophagous insects have a dietary requirement for ascorbate, and we did not find GULO in any insect genomes. Note also that some species of insect (e.g. cockroaches) may obtain ascorbate from eukaryote endosymbionts, which may allow them to survive on ascorbate-poor diets.

## Supplementary File 1: Distribution of *GULO* and *GLDH* in opisthokont and apusomonad genomes

Genomes of the opisthokonts (including animals and fungi) were examined for the presence of *GULO* and *GLDH*. Strikethrough indicates non-functional pseudogenes. The absence of *GULO* is well documented in the known ascorbate auxotrophs such as haplorhine primates, guinea pigs, bats, teleost fish and passerine birds. Some fungi (e.g. ascomycetes) use D-arabinonolactone oxidase to produce five carbon ascorbate analogue, erythroascorbate. D-arabinonolactone oxidase has different substrate specificity to GULO, but exhibits a high degree of sequence similarity and has been classed as GULO in the table.

## Supplementary File 2: Identification of *GULO* and *GLDH* in marine microbial eukaryote transcriptomes

Data from the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP, http://marinemicroeukaryotes.org/) were analysed for the terminal enzymes in ascorbate biosynthesis. This dataset contains 679 transcriptomes from 320 different species. Sequence similarity searches used a stringent length cut off to avoid ambiguous results from incompletely sequenced gene products (minimum length 300 amino acids). Using these criteria, we identified *GULO* or *GLDH* in 165 species. Although the absence of a gene in a transcriptome cannot be used to infer absence, we found no examples of organisms that possess both *GULO* and *GLDH*, even when a more relaxed length criterion was used (minimum length 100 amino acids). Note that the underlined *GLDH* sequences from the ciliates *Myrionecta* and *Strombidinopsis* are 100% identical to sequences recovered from their prey (respectively *Geminigera cryophila* and *Isochrysis galbana*). These sequences may therefore be due to contamination. Alternatively, as both these ciliates exhibit kleptoplasty, the presence of algal *GLDH* sequences in the ciliate transcriptome may also represent examples of plastid-related nuclear genes that are retained and transcribed to aid plastid function [95]. In the latter scenario, these ciliates could therefore use GLDH to temporarily synthesise ascorbate during plastid acquisition.

## Supplementary File 3: Identification of *VTC2* in marine microbial eukaryote transcriptomes

Data from the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) was analysed for *VTC2*, encoding GDP-L-galactose phosphorylase, the first committed step

in land plant ascorbate biosynthesis. Sequence similarity searches used a stringent cut off to avoid ambiguous results from incompletely sequenced gene products (minimum length 300 amino acids). *VTC2* was identified in 37 species, all of which belong to the Chlorophyta. Genes exhibiting weak similarity to GDP-L-galactose phosphorylase, which may represent homologues of GDP-D-glucose phosphorylase [96], were not included in these results.

**Supplementary File 4: Distribution of ascorbate biosynthetic genes in Archaeplastida transcriptomes**

Rhodophyte transcriptomes from the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) or Genbank[97,98] were examined for the presence of ascorbate biosynthesis genes. The rhodophytes transcriptomes all exhibit the pathway found in the genomes of *Cyanidioschyzon merolae, Chondrus crispus* and *Porphyridium purpureum*, possessing *GLDH* rather than *GULO*. *Chlorokybus atmophyticus* is a green alga belonging to the Streptophyte lineage containing land plants and charophyte algae. The *Chlorokybus* transcriptome appears unique amongst the Viridiplantate in that it contains *GULO* rather than *GLDH*. All of the other enzymes of the plant pathway are present.

**Supplementary File 5: Genome resources used in this study**

A list of the eukaryote genomes used to study the distribution of genes relating to ascorbate biosynthesis and metabolism.

| Animal auxotroph | Primary dietary source of ascorbate | Ultimate dietary source of ascorbate | Enzyme for ascorbate synthesis | References |
|---|---|---|---|---|
| **Primates** | Land plants | | GLDH | [99] |
| **Guinea pig** | Land plants | | GLDH | |
| **Bats** | Land plants | | GLDH | [99-101] |
| | Insects | Land plants | GLDH | |
| | Fish | Phytoplankton | GLDH | |
| | Blood | | GULO | |
| **Passerine birds** | Land plants | | GLDH | [15] |
| | Insects | Land plants | GLDH | |
| | Small vertebrates | | GULO | |
| **Teleost fish** | Zooplankton (crustacea) | Phytoplankton | GLDH | [102] |
| | Phytoplankton | | GLDH | |
| **Crustacea** | Phytoplankton | | GLDH | [103,104] |
| **Phytophagous insects** | Land plants | | GLDH | [35,105] |

**Table 1: Dietary sources of ascorbate in animal auxotrophs**