Molecular evidence for the early evolution of photosynthetic water oxidation

Tanai Cardona[1*], Patricia Sánchez-Baracaldo[2], A. William Rutherford[1], Anthony Larkum[3]

[1]Department of Life Sciences, Imperial College London, London, UK
[2]School of Geographical Sciences, University of Bristol, Bristol, UK
[3]University of Technology Sydney, Ultimo NSW, Australia

*Correspondence to: t.cardona@imperial.ac.uk

Abstract

The evolution of Photosystem II changed the history of life by oxygenating the Earth's atmosphere. However, there is currently no consensus on when and how oxygenic photosynthesis originated. Here we present an alternative perspective on the evolution of oxygenic photosynthesis by studying the evolution of D1 and D2, the core subunits of Photosystem II, as a function of time. A Bayesian relaxed molecular clock approach was applied to the phylogeny of Type II reaction center proteins using geochemical constraints and calibrations derived from the fossil record of Cyanobacteria and photosynthetic eukaryotes. The resulting age estimates were interpreted in the context of the structure and function of photochemical reaction centers. Firstly, we point out it is likely that the ancestral protein to D1 and D2 gave rise to a photosystem that was capable of water oxidation. Secondly, our results indicate that the gene duplication event that led to the divergence of D1 and D2 is likely to have occurred more than a billion years before the emergence of the last common ancestor of extant Cyanobacteria. Thirdly, we show that it is unlikely that Cyanobacteria obtained photosynthesis via horizontal gene transfer. Furthermore, the data suggest that the origin of photosynthesis in the early Archaean was necessarily followed by a rapid diversification of reaction center proteins, which included the divergence of D1 and D2. It is therefore concluded that primordial forms of water oxidation could have originated relatively soon after the emergence of photochemical reaction centers in the early Archaean.

Introduction

Molecular evolution of Type II reaction centers

The transition from anoxygenic to oxygenic photosynthesis must have initiated when an ancestral photochemical reaction center evolved the capacity to oxidize water to oxygen (Michel & Deisenhofer, 1988; Rutherford, 1989). Today, water oxidation is catalyzed by the Mn_4CaO_5 cluster of Photosystem II (PSII) of Cyanobacteria and photosynthetic eukaryotes. Therefore, a complete scenario for the origin of oxygenic photosynthesis must explain how and when Type II reaction centers diversified and how and when one of these reaction centers evolved the capacity to oxidize water. The evolution of Type II reaction center proteins has been described and discussed in some detail before (Beanland, 1990; Blankenship, 1992; Cardona, 2015, 2016b; Nitschke & Rutherford, 1991; Rutherford & Nitschke, 1996; Sadekar et al., 2006) and it is schematized in Fig. 1A.

Fig. 1A illustrates that the core subunits of PSII, D1 and D2, and the core subunits of the anoxygenic Type II reaction centers, L and M, share a common origin (Beanland, 1990). That is to say that D1, D2, L and M, all descended from a single protein. This ancestral protein is denominated **II** in Fig. 1A. However, D1 and D2 share between each other significantly more sequence identity than either of them share with L or M, and this is reflected in greater structural identity between D1 and D2. The same is true the other way around, L and M share between each other significantly more sequence identity than either of them does with D1 or D2, and this is accordingly translated into greater structural similarity between L and M (Fig. 1). Thus, the ancestral core protein **II** diversified into two forms, one ancestral to D1 and D2, and the other ancestral to L and M. The ancestral protein to D1 and D2 is denominated **D0** in Fig. 1. The ancestral protein to L and M is denominated **K**. Hence, D1 and D2 originated from a gene duplication event and makes a monophyletic clade of Type II reaction center proteins, distinct from that which gave rise to L and M (Cardona, 2015, 2016b).

It is worth noting that the monophyletic relationship of D1 and D2 implies that PSII core proteins cannot be derived from L and M proteins, and thus Cyanobacteria could not have obtained Type II reaction center proteins from Proteobacteria or Chloroflexi. Fig. 1A highlights that the evolutionary event that led to the divergence of **D0** and **K**, marked **2**, predate the diversification event that led to the divergence of L and M, marked **4** and **5**. In other words, while D1 and D2 share

common ancestry with L and M, the former did not originate from the latter, instead both families evolved in parallel. This was first noted by Beanland (1990) and it is corroborated by more detailed phylogenetic studies that integrate structural and functional comparisons of the reaction center proteins, see for example Sadekar et al. (2006). A first corollary of the early divergence of **D0** and **K** is that the most ancestral Type II reaction centers were homodimeric: thus, a homodimeric reaction center made with subunit **K** was ancestral to the anoxygenic Type II reaction centers, and a homodimeric reaction center made of **D0** was ancestral to PSII. A second corollary is that the ancestral Type II reaction center, made of a homodimer of ancestral protein **II**, cannot be described as being more similar to L or M, than to D1 and D2. For this reason and rather counterintuitively, there is no phylogenetic evidence against the possibility that the most ancestral Type II reaction centers (**II**) were capable of water oxidation. A third corollary is that the divergence of D1 and D2 (**2**) is likely to have predated the evolution of Chloroflexi-type and Proteobacteria-type L and M proteins (**5**). However, this third corollary cannot be corroborated until a study of the rates of evolution of the distinct reaction center proteins is carried out.

Molecular evolution of Photosystem II core proteins

When we refer to *the last common ancestor of Cyanobacteria* we imply here the most recent ancestor between the genus *Gloeobacter* and all other extant strains of Cyanobacteria capable of oxygenic photosynthetic; excluding the recently described non-photosynthetic Cyanobacteria classes, Sericytochromatia and Melainabacteria (Soo et al., 2017; Soo et al., 2014).

The last common ancestor of Cyanobacteria was already capable of efficient oxygenic photosynthesis because it had a PSII able to coordinate a Mn_4CaO_5 cluster virtually indistinguishable from that of all other extant strains (Cardona et al., 2015). This is based on the observation that even the earliest diverging clade, the strains of the genus *Gloeobacter* (Blank & Sanchez-Baracaldo, 2010; Honda et al., 1999; Sanchez-Baracaldo et al., 2005; Shih et al., 2013), encode in their genomes a standard set of genes for the assembly of PSII (Nakamura et al., 2003; Saw et al., 2013) and can oxidize water by a mechanism identical to those of later evolving strains (Koyama et al., 2008).

The Mn_4CaO_5 cluster of PSII is coordinated mainly by the D1 protein (Ferreira et al., 2004; Nixon & Diner, 1994), which provides six ligands to the cluster (Fig. 2). The structure of the cluster has been resolved in detail by Umena et al. (2011) and Suga et al. (2015). Among reaction center proteins D1 has the most complex evolutionary history, which was described in detail by Cardona et al. (2015) and is schematized in Fig. 1B. At least five different well-defined types of D1 have been identified. The early evolving forms, referred to as atypical D1 forms, are characterized by numerous amino acid positions predicted to be ancestral, as well as unique traits evolved secondarily as these adapted to novel functions, but most remarkably, by the absence of some of the ligands to the Mn_4CaO_5 cluster . There are three distinct groups of atypical forms: the earliest evolving D1 found in the genome of *Gloeobacter kilaueensis* (G0 in Fig. 2) has retained sequence similarity with D2 at key positions now lost in other forms of D1 (Cardona et al., 2015); the super-rogue D1 (Murray, 2012), now known to be required for the synthesis of chlorophyll *f* (G1) (M. Y. Ho et al., 2016), and the rogue D1 (Murray, 2012) or sentinel D1 (G2) thought to be required for the shut-down of water oxidation activity without disassembly of the PSII complex (Wegener et al., 2015).

The late evolving forms of D1, also referred to as the standard D1 forms, are characterized by having a complete ligand sphere for the Mn_4CaO_5 cluster and can be subdivided into two groups: the microaerobic form of D1 (G3) expressed under low oxygen concentrations (Sicora et al., 2009) and the dominant form of D1 (G4), which is present in the genome of all Cyanobacteria and photosynthetic eukaryotes with PSII and it is the main form of D1 associated with water oxidation catalysis. Previous phylogenetic studies seemed to suggest that the microaerobic form (G3) branched out before the dominant form of D1 (G4) diversified (Cardona et al., 2015). Remarkably, a water-splitting PSII with a microaerobic form of D1 is more efficient and robust than a PSII with a dominant form of D1 under low-oxygen concentrations (Crawford et al., 2016). This observation may suggest that the dominant form of D1 evolved as a response to increasing oxygen concentrations within the cell and the environment (Cardona et al., 2015).

Both the phylogeny of D1 proteins and the ancestral traits retained in the atypical D1 sequences suggest that these atypical forms might have branched out at different stages during the evolution of water oxidation but prior to the evolution of the standard forms of D1. In this way, it can

be deduced from the available sequence data that the Mn_4CaO_5 cluster reached its current configuration (Umena et al., 2011), at least in the ancestral state of D1 before the divergence of the microaerobic (G3) and dominant forms (G4).

The phylogeny of D1 is also characterized by an extensive number of duplication events occurring at every taxonomic level (Cardona et al., 2015), some of which are ancient and gave rise to the atypical forms, but many of which are much more recent. Therefore, every cyanobacterium with multiple *psbA* genes can carry several identical copies encoding identical D1 subunits, which are the result of recent gene duplications, and copies with different degrees of divergence, which are the result of more ancestral duplication events. The phylogeny of D1 also shows that every cyanobacterium with PSII has at least one *psbA* gene encoding a dominant form of D1 (G4). For example, *Gloeobacter kilaueensis* has a genome with six *psbA* genes, five encode dominant forms (G4), three of these five have identical amino acid sequences, while the other two show some level of sequence divergence. The sixth gene encodes a unique atypical sequence, designated here G0. In contrast, *Gloeobacter violaceous* lacks the atypical sequence but has the other five D1 sequences. Another example is *Chroococcidiopsis thermalis*, which has in its genome six different *psbA* genes, four encoding different dominant forms of D1 (G4), in addition to a rogue D1 (G2) and a super-rogue D1 (G1) with chlorophyll *f* synthase activity (M. Y. Ho et al., 2016).

Altogether, the distribution and phylogeny of D1 proteins indicate that the last common ancestor of Cyanobacteria already carried in its genome a dominant form of D1 (G4), which implies that the gene duplication events that led to the origin of the microaerobic form of D1 (G3) and the atypical sequences (G0, G1, G2) predated this last common ancestor and the diversification of crown group Cyanobacteria (Cardona et al., 2015). The phylogeny of D1 shows that the last common ancestor of Cyanobacteria existed at a very advanced stage in the evolution of oxygenic photosynthesis and PSII water oxidation. If so, how long did it take for an ancestral Type II reaction center to evolve into the sophisticated water-oxidizing photosystem inherited by Cyanobacteria and photosynthetic eukaryotes?

The ancestral photosystem prior to the divergence of D1 and D2

As a result of the monophyletic relation of D1 and D2 and the conserved structural and functional characteristics between these proteins, it is possible to reconstruct with confidence some of the traits of the ancestral photosystem. These conserved characteristics, present in both D1 and D2, but absent in L and M, strongly suggest that the ancestral homodimeric photosystem, made of a **D0** dimer, had most of the traits known to be associated with the highly oxidizing potential of standard PSII and required for water oxidation (Cardona, 2016b; Cardona et al., 2012; Rutherford & Faller, 2003; Rutherford & Nitschke, 1996).

Four structural domains differentiate D1 and D2 from L and M (Fig. 2A). These are: 1) the N-terminus, which is essential for the efficient disassembly and repair of these subunits upon photodamage, which occurs by a mechanism that is identical for D1 and D2 (Komenda et al., 2007; Krynicka et al., 2015). 2) A segment between the 1st and 2nd transmembrane helices, which is required for protein-protein interactions with the small subunits and the extrinsic polypeptides (Cardona, 2015, 2016b). The small subunits are required for assembly and stability of the complex and none of them is present in anoxygenic reaction centers (Dobakova et al., 2007; Komenda et al., 2012; Sugiura et al., 2015). The extrinsic polypetides are important for the stability of the cluster, in particular PsbO (De Las Rivas et al., 2004; Franzen et al., 1985; Roose et al., 2016). 3) An extended loop between the 4th and the 5th transmembrane helices required for the coordination of bicarbonate, which is important for regulation and photoprotection (Brinkert et al., 2016; Ferreira et al., 2004). 4) An extended C-terminus necessary for the coordination of the Mn_4CaO_5 cluster, Cl^- binding, water channels, and proton pathways (Debus, 2001; Linke & Ho, 2013; Umena et al., 2011). All four of these structural differences are either directly required for water oxidation or are necessary for stability of the complex, assembly, repair, and photoprotection. The strong inference that can be made from these structural characteristics is that before the divergence of D1 and D2, the ancestral photosystem was already undergoing evolutionary changes to support water oxidizing chemistry and to deal with radical oxygen species.

Further evidence for water oxidation prior to the divergence of D1 and D2 can be observed at the sequence and cofactor level. A redox tyrosine-histidine pair is conserved in both D1 and D2, this demonstrates that the $P/P^{+\bullet}$ chlorophyll redox pair in this ancestral photosystem was able to generate

enough oxidizing power to form a neutral tyrosyl radical on either side of the homodimeric reaction center (Rutherford & Faller, 2003; Rutherford & Nitschke, 1996). In addition, several potential metal ligands are traceable to **D0**, at least two glutamate at positions equivalent to 170 and 189 of standard D1, a histidine at position equivalent to 337, and an additional ligand to the Mn_4CaO_5 cluster is provided from the ancestral protein of the CP43 and the CP47 subunits (Cardona, 2016b; Cardona et al., 2015), Fig. 2B. Conserved carotenoids near the peripheral reaction center chlorophylls, Chl_Z and Chl_D, implies the presence of photoprotective mechanism to quench triplet chlorophyll states and the existence of electron transfer side pathways (Cardona et al., 2012; Krieger-Liszkay et al., 2008), which are in place to prevent the formation of singlet oxygen. Conserved tyrosine residues on both D1 (Y246) and D2 (Y244), which provide hydrogen bonds to the bicarbonate ligand of the non-heme iron (Umena et al., 2011), suggest that this was present in the ancestral homodimeric photosystem, made of **D0,** Fig. 2C. The binding of bicarbonate has now been shown to modulate the potential of the quinones as a protective mechanism against singlet oxygen formation (Brinkert et al., 2016). It can be concluded that the ancestral photosystem made of **D0** was able to oxidize water accompanied by the unavoidable production of radical oxygen species, which then required the innovation of protective measures. It is plausible that the ancestral photosystem oxidized manganese and coordinated a primordial catalytic cluster. From this it follows that the heterodimerization process of D1 and D2, culminating with the appearance of the standard form of D1, was consequently a path to optimization of: 1) water oxidation efficiency (Rutherford et al., 2012), 2) photoprotective mechanisms against radical oxygen species (Krieger-Liszkay et al., 2008), and 3) photoactivation of the Mn_4CaO_5 cluster, assembly, and repair of the protein complex in the presence of oxygen (Nickelsen & Rengstl, 2013; Nixon et al., 2010).

Geochemical evidence for photosynthesis

It is thought that some form of photosynthesis was well stablished by 3.5 Ga. This is demonstrated by both sedimentological and isotopic evidence for photoautotrophic microbial communities recorded in Paleoarchean rocks (Butterfield, 2015; Nisbet & Fowler, 2014; Tice & Lowe, 2004). In addition, sedimentary rocks and banded iron formations from Isua, Greenland, hint at the presence of

photosynthetic bacteria in the marine photic zone as early as 3.7-3.8 Ga (Czaja et al., 2013; Grassineau et al., 2006; Knoll, 2015; Rosing, 1999; Rosing & Frei, 2004; Schidlowski, 1988). There is consensus that during most of the Archean, the world was largely anaerobic with oxygen concentrations staying below 10^{-5} of present atmospheric levels (Lyons et al., 2014). Only about 2.4 Ga did atmospheric oxygen concentrations increase by several orders of magnitude (Bekker et al., 2004) to about 10^{-2} to 10^{-1} of current levels; this point in time is known as the Great Oxidation Event (GOE) and sets a firm constraint on the minimum age for the evolution of oxygenic photosynthesis (Fischer et al., 2016; Lyons et al., 2014). However, the origin of oxygenic photosynthesis still remains controversial. There is much evidence for local oxygen 'oases' in the late Archean record, possibly going back to 3.0 Ga (Riding et al., 2014), but it has been reported that oxygen was present in sedimentary deposits as old as 3.7 Ga based on redox-sensitive trace elements (Frei et al., 2016; Lyons et al., 2014; Planavsky et al., 2014; Rosing & Frei, 2004; Satkoski et al., 2015). While there is consensus from geochemical evidence that anoxygenic photosynthesis originated in the early Archaean, there is much greater uncertainty for the origin of oxygenic photosynthesis.

Two conclusions relevant to this study can be made on the evolution of photochemical reaction centers from the geochemical and microfossil record. 1) The emergence of any form of photosynthesis before 3.5 Ga means that the origin of photochemical reaction centers must predate this date. 2) A standard form of Photosystem II must have originated before, or coincident with, the Great Oxidation Event. If this is so, then what happened during the evolution of photochemical reaction centers between the emergence of photoautotrophy in the early Archaean and the appearance of the standard configuration of Photosystem II in crown group photosynthetic Cyanobacteria?

Reformulating the evolution of oxygenic photosynthesis

Prior to the evolution of the standard form of D1 several stages in the evolution of oxygenic photosynthesis can be envisaged: the earliest of these stages is the divergence of Type I and Type II reaction center proteins (**1**, Fig. 1A); this is then followed by the divergence of the *anoxygenic* family (L/M) and *oxygenic* family (D1/D2) of Type II reaction center proteins (**2**), then by the duplication event that led to the divergence of D1 and D2 (**3**), and the subsequent (**7**) gene duplication events and

specializations that created the known diversity of D1 forms, which ultimately resulted in the emergence of the dominant form of D1 (G4) that characterizes PSII today. If **D0** was able to support water oxidation chemistry, then the earliest stages in the evolution of oxygenic photosynthesis occurred between stages **2** and **3** as depicted in Fig. 1A.

The evolution of oxygenic photosynthesis can be conceptualized from the molecular phylogeny of Type II reaction centers as the span of time between **D0** and the ancestral standard form of D1 (marked **8** in Fig. 1B). This span of time, or the difference in time between the earliest possible occurrence of water oxidation recorded in the molecular sequence record and the appearance of the standard form of D1, representing modern-day PSII, we denote by **ΔT**. If **ΔT** is small, a few million years or less for example, it could be concluded that the evolution of oxygenic photosynthesis was a sudden and fast process. This scenario would imply that oxygenic photosynthesis evolved rapidly just before the GOE as suggested recently by Ward et al. (2016) or Shih et al. (2017). On the other hand, if **ΔT** is large, several hundred million years or more for example, it would be consistent with oxygenic photosynthesis emerging long before the GOE as suggested by some geochemical (Planavsky et al., 2014; Satkoski et al., 2015) and phylogenetic data (Blank & Sanchez-Baracaldo, 2010; Schirrmeister et al., 2013; Schirrmeister et al., 2015).

Here we applied Bayesian relaxed molecular clock analyses to assess the magnitude of **ΔT** and the dynamics of the rates of evolution of Type II reaction center proteins trough time. To do this we started with the basic assumptions that the earliest stages in the evolution of photochemical reaction centers occurred in the early Archaean and that the last common ancestor of crown group Cyanobacteria inherited a standard form of Photosystem II. We then constricted the clock with geochemical constraints of photosynthesis and previously implemented fossil constraints across Cyanobacteria and photosynthetic eukaryotes from the Proterozoic and Phanerozoic. The data presented here indicate that the evolution of oxygenic photosynthesis is more likely to have been a long and gradual process starting early during the evolutionary history of life.

Results

Change in sequence identity as a function of time

A first approximation to the evolution of Type II reaction centers as a function of time can be derived from the level of sequence identity between D1 and D2 of different species with known divergence times as shown in Fig. 3. For example, the D1 protein of the dicotyledon *Arabidopsis thaliana* shares 99.7% amino acid sequence identity with that of the dicotyledon *Populus trichocarpa*, which are estimated to have diverged between 127.2 and 82.8 Ma (Clarke et al., 2011), Fig. 3 and supplementary Table S1. On the other hand, *A. thaliana*'s D1 shares 87.7% sequence identity with that of the red alga *Cyanidioschyzon merolae*. Red algae are known to have diverged at least 1.2 Ga ago (Butterfield, 2000) and recently described exceptionally well preserved fossils have pushed this date back to 1.6 Ga (Bengtson et al., 2017). At the other end of this evolutionary line, the three dominant forms of D1 from *Gloeobacter violaceous* share on average 79.2% sequence identity with that of *C. merolae* or 78.5% with that of *A. thaliana*. If the percentage of sequence identity between pairs of species is plotted as a function of their divergence time, a linear decrease of identity is observed among reaction center proteins at a rate of less than 1% per 100 million years (supplementary Table S2). In other words, the rate of evolution of D1 and D2 since the GOE and since the emergence of photosynthetic eukaryotes has remained relatively slow until the present time, if considered over a large geological time scale, with less than 20% change in sequence identity in the past 2.0 Ga.

Now, if the last common ancestor of Cyanobacteria existed hundreds of millions of years before the GOE, this would presuppose an even slower rate of evolution of the core subunits of PSII. On the other hand, if the rate of evolution of D1 and D2 are taken at face value, following the roughly uniform rate observed in photosynthetic eukaryotes, this would locate the divergence of *Gloeobacter* after the GOE (Fig. 3, red spot): in consequence, the older the last common ancestor of Cyanobacteria, the slower the rate of evolution of the dominant form of D1 and D2. Therefore, large uncertainties in the fossil record of photosynthetic eukaryotes would result in only small changes to this trend. For example, if the divergence of red algae occurred as late as 1.0 Ga or as early as 2.0 Ga, this will only cause a small shift in the overall rate. Or for example, if the last common ancestor of angiosperms is actually 100 million years older than currently understood, this would result in almost

a negligible change in the rate of evolution of the dominant form of D1 and D2 over the large time scale of the planet.

All reaction center proteins originated from a single ancestral protein that diversified as the multiple groups of photosynthetic bacteria arose. As a result of this common ancestry, any standard D1 shares on average about 27% sequence identity with any D2 across their entire sequence. Any standard D1 or D2 share on average 17% sequence identity with any L or M. The level of sequence identity falls well below 5% if any Type II reaction center protein is compared with any Type I reaction center protein (Cardona, 2015). In consequence, the rate of evolution of D1 and D2 since the GOE, as estimated from the decrease of sequence identity (<1% per 0.1 Ga), is too slow to account for the evolution of photochemical reaction centers within a reasonable amount of time (Fig. 3, dashed line). In other words, the rate of evolution of reaction center proteins since the origin of life could not have been constant, and any scenario for the origin of photochemical reaction centers at any point in the Archaean requires initially faster rates of evolution than any rate observed since the Proterozoic (Fig. 3, light blue line).

Taking into consideration that D1 and D2 share only about 27% sequence identity, this would suggests, as illustrated in Fig. 3, that the event that led to the divergence of D1 and D2 is more likely to have occurred closer to the origin of the primordial reaction center proteins during the emergence of the first forms of photoautotrophy in the early Archaean, than closer to the Great Oxidation Event.

Bayesian relaxed molecular clock analysis

While the simple approach used above indicates that the divergence of D1 and D2 must be placed well before the GOE, it is possible to make further inferences on the evolution of oxygenic photosynthesis by applying a relaxed molecular clock approach to the phylogeny of Type II reaction center proteins and from the resulting rates of evolution. Fig. 4 shows a Bayesian relaxed log-normal autocorrelated molecular clock built using the CAT + GTR + Γ model allowing for flexible boundaries on the calibration points (Lartillot et al., 2009). We specified the age of the root (root prior) at 3.5 Ga with a standard deviation of 0.05 Ga. That is to say, that the most ancestral form of a Type II reaction center protein is assumed to have already evolved by 3.5 Ga. Under these conditions,

the last common ancestral protein to the standard form of D1 prior to the divergence of the G3 and G4 types (Fig. 4, green dot) is timed at 2.19 ± 0.22 Ga. On the other hand, **D0** (Fig. 4, orange dot) is timed at 3.22 ± 0.19 Ga. It follows then that the difference in time between **D0** and the first standard form of D1, **ΔT**, is 1.02 Ga.

Table 1 shows estimates of divergence times of key ancestral Type II reaction center proteins and the respective **ΔT** value using different root priors. It can be seen, for example, that under the assumption that Type I and Type II reaction center proteins had already diverged 3.8 Ga ago, **ΔT** is found to be 1.17 Ga. Similarly, if the origin of photosynthesis is assumed to be a late event centered at 3.2 Ga, though unlikely, **ΔT** is still 0.90 Ga. These patterns show that regardless of the exact timing of the origin of the first photochemical reaction centers in the early Archaean, the evolution of oxygenic photosynthesis is unlikely to have been a sudden event occurring in a few million years prior to the GOE.

Relaxing the boundaries on the calibration points by increasing the standard deviation on the root prior pushes the timing of the earliest events in the evolution of Type II reaction center to even older ages rather than younger ages (supplementary Table S3). For example, a root prior of 3.5 Ga with a standard deviation of 0.1 Ga pushes the estimated time for the root to 3.65 Ga; and with a standard deviation of 0.2 Ga to 4.00 Ga. In the former case, **ΔT** shifts to 1.16 Ga and in the latter to 1.31 Ga.

Fig. 4 additionally indicates that the divergence of the L and M subunits occurred *after* the divergence of D1 and D2. The estimated time for the divergence of L and M is centered at 2.87 ± 0.16 Ga, while the time for the divergence of D1 and D2, as we saw above, is 3.22 ± 0.19 Ga. The late divergence of L and M suggests that the divergence of D1 and D2 predates significantly the diversification event that led to the evolution of Chloroflexi-type and Proteobacteria-type L and M subunits, in agreement with the third corollary of the evolution of reaction center proteins as described in the introduction.

The Bayesian analysis using flexible boundaries on the calibration points consistently produced ages for the divergence of the D2 subunit of *G. violaceous* and the dominant form of D1 (G4) after the GOE, in agreement with the work by Shih et al. (2017). Yet, previous molecular clocks

have suggested that the last common ancestor of Cyanobacteria might predate the GOE (Schirrmeister et al., 2015), so we also performed a similar analysis that allowed us to explore this scenario. This was only achieved by the application of less flexible boundaries under an empirical amino acid substitution model (LG + Γ) instead of the non-parametric approach described above. We found this to be the only way to locate the D2 of *G. violaceous* and the dominant form of D1 (G4) before the GOE. The effect of less flexible boundaries on the estimated divergence times is shown in Table 1. Assuming a root prior of 3.5 ± 0.05 Ga, the estimated divergence time for the standard form of D1 becomes 2.64 ± 0.15 Ga and pushes **D0** back to 3.40 ± 0.09 Ga, making **ΔT** 0.77 Ga. On the other hand, if we allowed flexibility on the root prior by increasing the standard deviation to 0.4 Ga (supplementary Table S3), the estimated divergence time for the standard form of D1 becomes 2.83 ± 0.21, but the estimated age of the root is pushed back to 3.92 ± 0.21 Ga with **D0** at 3.79 ± 0.21, making **ΔT** 0.96 Ga. Overall, placing the last common ancestor of Cyanobacteria before the GOE pushes the gene duplication event that led to the divergence of D1 and D2 even closer to the origin of Type II reaction centers and to the origin of photosynthesis.

Rates of evolution

Fig. 3 suggests that the rates of evolution had to be faster at the initial stages in the early Archaean than in the Proterozoic, even when **ΔT** is as large as 1.0 Ga. To have a better understanding of the changes of the rate of evolution of Type II reaction center proteins we plotted the rates as a function of divergence time. The rates are generated from the molecular clock analysis as described in Lepage et al. (2007) and references therein.

Fig. 5A plots the rate of evolution (**v**) of each node in the tree, expressed as amino acid substitutions per site per unit of time, against the estimated divergence time for each respective node. It can be seen that the rate at the earliest stage is much faster than the rates observed since the Proterozoic. What is remarkable is that faster rates at the origin of photosynthesis are required even if the origin of Type II reaction centers is assumed to be located at 4.2 Ga, as seen in Fig. 5B. The decrease in the rate of evolution is consistent with the observations derived from Fig. 3 and appears to approximate a first-order exponential decay curve (supplementary Table S4). Fig. 5A additionally

shows that L and M have been evolving at a slightly faster rate than D1 and D2. From this slow-down of the rates of evolution it can be calculated that since each respective duplication event (**2** and **4** in Fig. 1) it took about 168 million years for D1 and D2 to fall to 50% sequence identity and about 115 million years for L and M.

The maximum rate of evolution in the D1 and D2 family of reaction center proteins is placed at the node that represents **D0**. We will refer to this rate as $\boldsymbol{\nu_{max}}$. Fig. 5A shows that the rate of evolution flattens out at comparatively slow rates during the Proterozoic. These rates correspond to the rates of Group 4 D1 and D2. We will refer to the average rate of evolution during this zone of slow change as $\boldsymbol{\nu_{min}}$. The average rate $\boldsymbol{\nu_{min}}$ is thus defined as the average rate from each node in Group 4 D1 and D2. In Fig. 5A, $\boldsymbol{\nu_{max}}$ is 5.03 ± 1.42 amino acid substitutions per site per Ga, while $\boldsymbol{\nu_{min}}$ is 0.12 ± 0.04 substitutions per site per Ga. Under the assumption that reaction centers had evolved by 3.5 Ga, these rates indicate that to account for the divergence of D1 and D2 in one billion years (**ΔT**), the initial rate of evolution had to be about 40 times larger than that observed since the last common ancestor of Cyanobacteria. For a comparison of published rates of evolution of diverse proteins see Table 2.

A fast rate of evolution at the earliest stage implies that $\boldsymbol{\nu_{max}}$ must increase if **ΔT** is smaller, as shown in Fig. 5B and 5C. The decrease in **ΔT** can only be obtained computationally by making the root prior younger, which correspondingly forces all nodes to younger ages. The increase of $\boldsymbol{\nu_{max}}$ with decreasing **ΔT** obeys a power law-function (supplementary Table S5). For example, if **ΔT** is twenty times smaller; that is, if the divergence of D1 and D2 occurred in only 50 million years, then $\boldsymbol{\nu_{max}}$ had to be many times larger, in this case 483.83 times larger than $\boldsymbol{\nu_{min}}$ as calculated by extrapolating from Fig. 5C. In simple terms, if D1 and D2 diverged in 50 Ma the initial rate of evolution had to be 2443 amino acid changes per site per Ga, which is only comparable to the rate of evolution of some viruses, see Table 2. Taking into consideration that reaction center proteins and many other proteins of importance in bioenergetics have changed little for billions of years, with some being as old as the last universal common ancestor (Weiss et al., 2016), a sudden virus-like rate of evolution to account for the divergence of D1 and D2, seems quite improbable. Therefore, it is unlikely that the divergence of the core proteins of PSII was a sudden or very rapid event.

Sensitivity analysis

To test the variability of the method we explored a range of contrasting models. We compared the effect of the model of relative exchange rates on the estimated divergence times: Fig. 6A plots estimated divergence times calculated with the CAT + GTR model (Yang & Rannala, 2006) against divergence times calculated using the CAT model with a uniform (Poisson) model of equilibrium frequencies. Fig. 6A shows that the GTR model does not have a strong effect on the calculated divergence times as the slope of the graphs does not deviate from unity when paired with the uniform model (supplementary Fig. S1 and Table S6). Thus under a root prior of 3.5 ± 0.05 Ga a CAT + Poisson model also generated a **ΔT** of 1.02 Ga, see Table 4.

To understand the effects of the oldest calibration point (point 11, Fig. 4) on the estimated divergence time, we tested a second set of boundaries restricting this point to a minimum age of 2.7 Ga (Calibration 2) under the assumption that the divergence of the genus *Gloeobacter* may significantly predate the GOE. Fig. 6B shows a comparison of the two calibration choices on the overall estimated times for a flexible and a non-flexible model. If the divergence times using both calibrations are plotted against each other, a linear relationship is obtained (see supplementary Fig. S2 and Table S7). Calibration 2 did not seem to have a very strong effect on the estimated divergence times nor **ΔT**. For example, under a root prior of 3.5 ± 0.05 Ga and employing Calibration 2, **ΔT** was 0.97 Ga, see Table 4. We also tested the effect of removing the oldest calibration point from the analysis (point 11, Fig. 4), but this only shifted **ΔT** to values greater than 1.0 Ga.

In contrast, the choice of model for the evolution of substitution rates had a strong impact on the estimated divergence times (supplementary Fig. S3 and Table S8). Fig. 6C shows divergence time estimates of a tree calculated using a relaxed log-normal autocorrelated molecular clock plotted against a tree calculated using an uncorrelated gamma model on the rates of evolution. The autocorrelated model assumes that close branches are more likely to evolve at a similar rate, while the uncorrelated model assumes that the rate of evolution of each branch can vary independently (S. Y. W. Ho & Duchene, 2014; Lepage et al., 2007). Under the uncorrelated model the estimated divergence times of many nodes were aberrantly shifted to younger ages: for example, most

Cyanobacterial and eukaryotic D1 clustered in the range of 700 to 0 Ma, which is inconsistent with the fossil record, and results in **ΔT** values of about 2.0 Ga. The molecular reason behind this difference could be related to the fact that photochemistry imposes strong constrains on the evolution of reaction center proteins. As all of them must coordinate and maintain all redox cofactors, chlorophylls, quinones, carotenoids, and a non-heme iron, at a precise orientation and distance from each other to allow for control of electron transfer rates and redox potentials. These rates and potentials are crucial not only for function but also for protection against the formation of radical oxygen species (Cardona et al., 2012; Rutherford et al., 2012). It seems reasonable then, that the rates of evolution of all Type II reaction center proteins should be, to some extent, similar, and even more so among closely related groups, thus corresponding to an autocorrelated model.

Discussion

A complete scenario for the origin and evolution of oxygenic photosynthesis needs to account for the evolution of Type II reaction centers and in particular, for the divergence of D1 and D2, the core subunits of PSII. Our results indicate that the evolution of oxygenic photosynthesis cannot be described as a rapid event happening within a few million years before the GOE. Our results suggests that the evolution of oxygenic photosynthesis is better described as a long process, spanning nearly a billion years, and starting soon after the origin of the first photochemical reaction centers in ancestral populations of bacteria.

Change in sequence identity as a function of time

As an approximation to the evolution of the core subunits of PSII we plotted the level of sequence identity of D1 and D2 as a function of known divergence times (Fig. 3). From this plot two main characteristics of the evolution of photosynthesis can be defined. Firstly, the three earliest stages in the evolution of photochemical reaction centers: 1) the divergence of Type I and Type II reaction centers, 2) the divergence of *anoxygenic* and *oxygenic* families of Type II reaction center proteins, and 3) the divergence of D1 and D2, were more likely to have occurred soon after the origin of the first reaction centers rather than near the GOE. Taking into consideration that there is strong evidence for

photoautotrophy at 3.5 Ga, then these three stages could well predate this time. In addition, if photosynthesis is confirmed at 3.8 Ga, then the occurrence of these three events must, consequently, be pushed further back in time accordingly. Secondly, the last common ancestor of Cyanobacteria is more likely to have lived around the time of the GOE rather than shortly after the origin of photosynthesis. This common ancestor must have had a dominant form of D1 (G4), which places it relatively far after the origin of photosynthesis. The early divergence of D1 and D2 implies that the earliest stages in the evolution of oxygenic photosynthesis may predate the last common ancestor of Cyanobacteria by about 1.0 Ga.

Bayesian relaxed molecular clock analysis

The application of a molecular clock to the phylogeny of Type II reaction centers is strongly constrained by evidence of some form of photosynthesis at 3.5 Ga, by the GOE at 2.45 Ga, and by the relatively slow rate of evolution of D1 and D2 for at least 2.0 Ga. These constraints could be in question if: 1) the geochemical evidence for photosynthesis in the early Archaean is the result of abiotic processed or geological artifacts, 2) the GOE has an entirely abiotic cause, and 3) photosynthetic eukaryotes radiated at a much earlier time, in the Archaean instead of the Proterozoic. However, given the current state of knowledge these three opposing alternatives not only appear improbable, but also contradictory.

The application of the constraints listed above on the molecular phylogeny of Type II reaction centers revealed that the divergence of D1 and D2 must be a long process traceable to the earliest events in the history of photosynthesis, as shown in Fig. 3. We highlight this long process by introducing the concept of **ΔT**, which defines two specific points in time in the evolution of D1 and D2. The magnitude of **ΔT** is dictated by the phylogenetic distance between D1 and D2 and their rate of evolution derived from the geochemical and fossil constrains. Therefore, it is not surprising that under most reasonable models employed in this analysis **ΔT** is always in the range of 1.0 Ga.

We have considered two possible evolutionary scenarios that are both consistent with a large **ΔT**. In the first scenario the standard forms of D1 start to diverge at about 2.4 Ga, as seen in Fig. 4, and diversify into G3 and G4 after the GOE. If we consider that the last common ancestor of

Cyanobacteria had a G4 D1, this would set it after the GOE. This scenario, derived from the application of a relaxed molecular clock using a non-parametric CAT model with flexible boundaries, is in agreement with the recent observations by Shih et al. (2017) and other molecular clock studies that placed the divergence of *Gloeobacter* after the GOE (David & Alm, 2011; Feng et al., 1997; Marin et al., 2017). In this scenario, under a root prior of 3.5 Ga, the divergence of D1 and D2 is set at about 3.2 Ga.

In the second scenario, we consider that the last common ancestor of Cyanobacteria occurs before the GOE as suggested by other molecular clock analyses (Falcon et al., 2010; Sanchez-Baracaldo, 2015; Schirrmeister et al., 2015). In our work, this scenario is obtained with the application of a relaxed molecular clock using an empirical amino acid substitution model (LG + Γ) with less flexible boundaries. In this scenario, under a root prior of 3.5 Ga, the appearance of the ancestral standard form of D1 is set at about 2.6 Ga. An age for a standard form of D1 of 2.6 Ga has the consequence of pushing the divergence of D1 and D2 closer to the root, and thus it is set at about 3.4 Ga (Table 1). What can be concluded from this is that the older the last common ancestor of Cyanobacteria is, the more likely it is that the divergence of D1 and D2 occurred near the origin of photochemical reaction centers and thus, near the origin of photosynthesis.

In addition, the relaxation of the root prior constraints by increasing the standard deviation, invariably results in older estimated ages for the root (supplementary Table S2). This result implies that it is more likely that photochemical reaction centers originated significantly before 3.5 Ga than after this time, which is consistent with the detection of ^{13}C-depleted organic carbon in 3.8 Ga rocks (Grassineau et al., 2006; Rosing, 1999) and by the existence of stromatolites of similar age (Nutman et al., 2016), both from the Isua supracrustal belt in Greenland.

Horizontal gene transfer

There are two main competing hypotheses regarding the evolution of oxygenic photosynthesis in Cyanobacteria. One hypothesis states that Cyanobacteria obtained photosynthesis via extensive horizontal gene transfer (HGT) of photosynthetic components from anoxygneic photosynthetic bacteria: see Raymond (2009) for a review and Soo et al. (2017) for a recent variant of this

hypothesis. This is designated here ‘the HGT hypothesis’. The second hypothesis states that the evolution of two distinct reaction centers is a trait of the last common ancestor of all photosynthetic organisms and that the divergence of the two types of reaction centers initiated after an ancestral gene duplication event, see for example (Harel et al., 2015; Mulkidjanian et al., 2006; Sousa et al., 2013). This is designated here as ‘the duplication hypothesis’.

The HGT hypothesis implies that the two distinct reaction centers diverged in distinct phyla of bacteria and were transferred to an ancestral *non-photosynthetic* cyanobacterium before the evolution of water oxidation. The issue with this hypothesis is that it is inconsistent with the evidence for the molecular evolution of photochemical reaction centers, which shows that the core subunits of PSI and PSII branched out before the diversification of the anoxygenic reaction center proteins (Cardona, 2015; Sadekar et al., 2006). On the other hand, the issue with the gene duplication hypothesis is that it leads to the question: “in which lineage did the gene duplication take place?” (Sousa et al., 2013). Given that Cyanobacteria are the only bacteria with Type I and Type II reaction centers, this question often leads to the conclusion that photosynthesis originated in Cyanobacteria or some form of ancestral Cyanobacteria, usually referred to as procyanobacteria or protocyanobacteria (Harel et al., 2015; Mulkidjanian et al., 2006; Sousa et al., 2013). However, this conclusion is in conflict with phylogenomic data, which consistently do not locate Cyanobacteria as a particularly early branching phylum in comparison with other phototrophic groups (Ciccarelli et al., 2006; David & Alm, 2011; Jun et al., 2010; Marin et al., 2017; Rinke et al., 2013; Segata et al., 2013; Wu et al., 2009). Our observations might help resolve some of these issues.

From Fig. 4 it is concluded that the divergence event, marked **II**, predated the diversification events that led to the divergence of L and M, and also the divergence of this pair into Proteobacteria-type and Chloroflexi-type core proteins. Similarly, event **II** also predated the diversification events that led to the divergence of D1 and D2, and all the variety of D1 proteins found in Cyanobacteria. It follows then that event **II** had to have occurred before the diversification of the phyla Proteobacteria, Chloroflexi, and Cyanobacteria. Put another way, it is likely that the phyla known today to have Type II reaction centers, including crown group photosynthetic Cyanobacteria, had not yet diversified during the earliest events in the evolution of photosynthesis, when the ancestral reaction centers first

diversified into Type I and Type II, and the subsequent gene duplications that resulted in the ancestral forms of L, M, D1, and D2. In agreement with this, an independent study performed by David and Alm (2011) identified a period of genetic innovation that coincided with a rapid diversification of the major bacterial lineages and the birth of a quarter of modern gene families. This period, between 3.33 Ga and 2.85 Ga, which the authors referred to as the Archaean Expansion, was suggested to be linked to the "bioavailability of oxygen", and was supported by the ancient origin of what the authors referred to as: "the basic genetic components required for oxygenic photosynthesis and respiration". Our results put the divergence of D1 and D2 at 3.22 Ga and that of L and M at 2.87 Ga, which as we have seen, requires that the split between the *anoxygenic* and *oxygenic* branches of Type II reaction center proteins, event **II**, occurred by 3.5 Ga. It is worth noting that 3.5 Ga is likely to be a conservative estimate for the origin of photosynthesis. If photosynthesis is considerably older, say 3.8 Ga (Czaja et al., 2013; Grassineau et al., 2006; Knoll, 2015; Rosing, 1999; Schidlowski, 1988), then this would imply an even older branching event for the ancestral core protein of PSII.

Taking into account that some form of photoautotrophy was well established by 3.5 Ga and that the major diversification of bacteria that resulted in the appearance of the described phyla is likely to have occurred later, in this so-called Archaean Expansion (David & Alm, 2011), and as a consequence of the oxygenation of Earth and the evolution of eukaryotes (Battistuzzi & Hedges, 2009; Brucker & Bordenstein, 2012; Harel et al., 2015; Marin et al., 2017), it seems increasingly plausible that photosynthesis evolved near the root of the bacterial tree of life. The origin of photosynthesis was then rapidly followed by the divergence of the *anoxygenic* and *oxygenic* branches of reaction center proteins. The implication of this is that the earliest stages in the evolution of photosynthesis, including the appearance of the ancestral forms of D1 and D2, concomitant with water oxidation, is likely to have started prior to the divergence of most described groups of bacteria now known to bear photochemical reaction centers. Another implication of this early divergence is that the results provided by Mulkidjanian et al. (2006), Sousa et al. (2013), and Harel et al. (2015) can be interpreted as evidence for two distinct reaction centers as a trait of ancestral populations of photosynthetic bacteria. A trait that today has been retained in Cyanobacteria and not acquired via

horizontal gene transfer. From this perspective it can be concluded that this trait has been retained in Cyanobacteria since shortly after the origin of photosynthesis.

There are additional evolutionary considerations derived from the structure of Photosystem II and its Mn_4CaO_5 cluster that require an origin of both types of reaction centers from a gene duplication event long before the last common ancestor of crown group Cyanobacteria, this have been discussed extensively elsewhere (Cardona, 2016b, 2017).

Recently, Shih et al. (2017) reported a phylogenetic analysis of Cyanobacteria and their closest non-photosynthetic relatives, the Melainabacteria (Soo et al., 2017; Soo et al., 2014). Shih et al. (2017) postulated that Cyanobacteria acquired photosynthetic capacity via horizontal gene transfer after they diverged from the Melainabacteria. They hypothesized that the transfer of photosynthetic genes occurred from anoxygenic photosynthetic bacteria into a *non-photosynthetic* Cyanobacterial ancestor with water oxidation evolving quickly before the GOE. Our results—which traced the evolution of the core subunits that allow the oxidation of water—show that not only is it unlikely that Cyanobacteria obtained photosynthesis via horizontal gene transfer, but it is also unlikely that the divergence of D1 and D2 from ancestral anoxygenic Type II reaction center proteins occurred rapidly just before the GOE. This leads to the conclusion that the origin of the newly described classes of non-photosynthetic Cyanobacteria is more likely to have been initiated after ancestral losses of oxygenic photosynthesis.

The early evolution of photosynthesis as suggested by the geochemical record and the data we report here is consistent with photoautotrophy being an ancestral trait of early evolving bacteria and implies that the loss of it might be a derived trait in many phyla of bacteria. It is sometimes argued that the loss of photosynthesis is unlikely due to the scattered distribution of reaction center proteins across the bacterial tree of life (Boucher et al., 2003; Fischer et al., 2016; Raymond, 2009; Xiong et al., 1998). Nevertheless, it has been shown repeatedly that gene loss is a very common process across all domains of life (Albalat & Canestro, 2016; D'Souza et al., 2014; Koskiniemi et al., 2012; Ku et al., 2015). For example, the majority of the proteins suggested to be in the universal ancestor by Weiss et al. (2016) are not universally distributed, including those in carbon (e.g. Wood-Linjbal pathway), nitrogen (e.g. nitrogenase, glutamine synthase), and cofactor metabolism (e.g. siroheme, cobalamin),

among many others, most of which are found to be scattered across the bacterial and archaeal domains. This fact implies ancestral and widespread gene loss across the entire tree of life. While horizontal gene transfer can only explain reacquisition of these processes after the universal ancestor, it cannot explain absence. Consistent with these observations, David and Alm (2011) showed that the period of rapid genetic innovation that peaked about 3.2 Ga is followed by a peak of massive gene loss a hundred million years later. It should not be at all surprising that clades of bacteria, which are specialized to live symbiotically and in aphotic environments (Di Rienzi et al., 2013), or as predators (Soo et al., 2015), such as the Melainabacteria, originated from ancestors that lost photosynthetic capacity. In fact, losses of photosynthesis across the tree of life is a continues process: several examples of the recent loss of oxygenic photosynthesis are the secondary loss of PSII by the symbiont *Atelocyanobacterium talassa* (Cornejo-Castillo et al., 2016; Thompson et al., 2012), the complete loss of photosynthesis of the cyanobacerium endosymbiont of *Epithemia turgida* (Nakayama et al., 2014), the case of non-photosynthetic diatoms (Kamikawa et al., 2017), parasitic Apicomplexa (Moore et al., 2008) and holoparasitic angiosperms (Ravin et al., 2016).

Ancestral losses of photosynthetic capacity do not imply that horizontal gene transfer of photochemical reaction centers has never occurred during the evolution of photosynthesis. Only a single unambiguous case of non-photosynthetic bacteria gaining phototrophy via horizontal gene transfer has been demonstrated so far; that is the case of *Gemmatimonas phototrophica* and closely related strains (Huang et al., 2016), which obtained Type II reaction centers and bacteriochlorophyll synthesis from a gammaproteobacterium (Zeng et al., 2014). All other suggested cases of ancestral horizontal gene transfer of photosynthesis in bacteria seem to be either very ancient, so that the direction of transfer is ambiguous—if it occurred at all—or between two photosynthetic strains. Examples of the latter case are the replacement of Cyanobacterial rubisco for proteobacterial rubisco in red algae and dinoflagellates (Delwiche & Palmer, 1996), the replacement of Cyanobacterial protochlorophyllide reductase for the proteobacterial enzyme in marine *Synechococcus* and *Prochlorococcus* (Bryant et al., 2012), and horizontal gene transfer of photosynthesis components between strains of Proteobacteria (Igarashi et al., 2001) or the transfer of reaction center genes or the far-red acclimation response gene cluster between strains of Cyanobacteria (Magnuson & Cardona,

2016; Sullivan et al., 2006). On the other hand, the most feasible case of an ancient event of horizontal gene transfer of Type II reaction centers supported by phylogenetic data might be between an ancestor of the Chloroflexi that predated the radiation event that led to the divergence of the Anaerolineae and Chloroflexia classes (Klatt et al., 2011), and an ancestor of the Proteobacteria predating the radiation event that led to the Alpha-, Beta-, and Gammaproteobacteria classes (Swingley et al., 2009; Tank et al., 2009), but postdating the divergence of Acidobacteria (Cardona, 2016a).

Rates of evolution

If a strict molecular clock model were to be applied to the phylogeny of Type II reaction centers, the root of the tree would be placed long before the formation of the planet (dashed red line in Fig. 3). Given the fact that the rate of evolution of D1 and D2 is restrained at comparatively slow rates from the Proterozoic until present times, the period of fast evolutionary change has to be located at the earliest stages to account for the evolution of photosynthesis within a reasonable timeframe. We found that even considering an age for the root of Type II reaction centers by 3.8 Ga the maximum rate during the evolution of D1 and D2, $\mathbf{v_{max}}$, is many times greater than the measured rates in the Proterozoic.

For the sake of argument, let us consider that an ancestral *non-photosynthetic* cyanobacterium obtained an anoxygenic Type II reaction center via horizontal gene transfer shortly after the divergence of Melainabacteria, roughly 2.63 Ga ago, as suggested by Shih et al. (2017). By 2.45 Ga this transferred anoxygenic Type II reaction center had evolved into a standard water-oxidizing PSII; that is 180 Ma after the Melainabacteria/Oxyphotobacteria split. It has been known for some time that the source of the transferred anoxygenic Type II reaction center core proteins could not have been a member of the Proteobacteria or the Chloroflexi, as D1 and D2 are not direct descendants of the L and M subunits found in these organisms. It becomes necessary to invoke the existence of a new lineage of undiscovered anoxygenic photosynthetic organisms that contained an anoxygenic reaction center made of D1- and D2-like proteins, at a point in time before the divergence of the atypical and standard forms of D1. A new question arises from these considerations, did the divergence of D1 and D2 occur

before or after the horizontal gene transfer event? If it occurred after, that means that the divergence of D1 and D2 must have occurred in 180 Ma, which as we have seen is not supported by the rates of evolution, as this would require the existence of reaction center that, all of a sudden, undergoes an extreme acceleration of the rate of evolution upon transfer, followed by a precipitous decline in the rate. In this case, the rate of evolution upon transfer would peak at a rate more than 800 times larger than the measured rates of evolution in the Proterozoic. On the other hand, in the even more unlikely case that the divergence of D1 and D2 occurred before the horizontal gene transfer, this would presuppose that D1 and D2 would have been diverging for a long period of time in this unknown lineage of bacteria. What is more, it would also presuppose that the capacity to oxidize water evolved before the horizontal gene transfer. In conclusion, a careful inspection of the evolution of reaction center proteins and their rates of evolution indicates that the HGT hypothesis cannot account for the observed diversification of reaction center proteins.

Another finding of the molecular clock analysis is that the divergence of L and M seems to occur after the divergence of D1 and D2. This is due to the slightly faster rates of evolution observed for L and M at all periods of time. If we add to this the fact that L and M subunits are about 50 to 90 amino acids shorter than D1 and D2, it should take less time for L and M to converge into the ancestral node **K**. This raises the following question: if the heterodimerization of D1 and D2 was driven by optimization of water oxidation and the incorporation of photoprotective mechanisms, what selective pressures triggered the heterodimerization of L and M?

It has been proposed that the heterodimerization of Type II reaction centers was driven by the optimization of electron transfer efficiency and the avoidance of photodamage (Cardona et al., 2012; Rutherford & Faller, 2003; Rutherford & Nitschke, 1996; Rutherford et al., 2012). In a homodimeric Type II ancestor containing two identical exchangeable quinones, charge separation would result in enhanced back-reactions caused by electron transfer to a quinone site when empty or when occupied by a reduced form of the quinone (Rutherford & Faller, 2003; Rutherford & Nitschke, 1996); and additionally, by shorter back-reaction pathways (Cardona et al., 2012; Rutherford et al., 2012). Back-reactions from the early radical pairs give rise to chlorophyll triplet states in the heart of the reaction center (Dutton et al., 1972; Rutherford et al., 1981) and these triplet states are dangerous when they

encounter oxygen, giving rise to the highly damaging singlet oxygen (Krieger-Liszkay et al., 2008; Rutherford et al., 2012; Vass & Cser, 2009). In the oxygenic **D0** homodimeric ancestor of PSII, avoidance of photodamage could be a significant selective pressure for heterodimerization (Rutherford et al., 2012). It seems clear that the enhanced back-reactions intrinsic to the homodimeric Type II reaction centers would have become a major problem whenever oxygen was present. The inefficient **D0** reaction center, as the primordial oxygenic photosystem, would have encountered this problem first and thus come under strong selection pressure toward heterodimerization at an early time. The results of the present work fit with this picture. Furthermore, our present finding that heterodimerization started significantly later in the anoxygenic homodimeric Type II reaction center (**K**) compared to its oxygenic counterpart (**D0**)**,** can be explained in this context as well, since **K** would have encountered these selection pressures much later, at a time when oxygen concentrations began to rise in localized environments in the late Archaean period, nearing the GOE (Bosak et al., 2009; Lyons et al., 2014; Riding et al., 2014).

It is rather remarkable that anoxygenic Type II reaction centers contain an asymmetrically located carotenoid in contact with a core bacteriochlorophyll (Deisenhofer & Michel, 1989). The role of this carotenoid is to quench bacteriochlorophyll triplet states in order to prevent the formation of singlet oxygen (Cogdell et al., 2000). In addition, light-harvesting complexes in most anoxygenic photosynthetic bacteria contain carotenoids and perform a similar photoprotective function (Kim et al., 2007; Melo et al., 2000; Tsukatani et al., 2012). As an extension of this, it does not seem unreasonable to think that even ancestral populations of anoxygenic phototrophic bacteria were under strong selective pressure by the threat of chlorophyll triplet-induced formation of radical oxygen species early during the evolutionary history of photosynthesis.

The phenomenon described here of an initial fast rate of evolution followed by an exponential decrease demands an explanation. Two possible mechanisms may account for this observation. The first one is the temperature dependent deamination of cytosine, as suggested by Lewis and coworkers (2016). They calculated that as the Earth cooled during its first 4 Ga, the rate of spontaneous mutation would have fallen exponentially by a factor of more than 4000. That is to say that the rate of spontaneous mutation during the earliest stages in the history of life would have been about three

orders of magnitude greater than those observed since the Proterozoic. Lewis and coworkers (2016) calculated that 50% of all spontaneous mutations occurred in the first 0.2 Ga, which matches well with the exponential decay trend seen in Fig. 5A, especially if an origin of photosynthesis is considered to be about 3.8 Ga. The second possibility is higher UV radiation on the planet's surface during the early Earth in the absence of an ozone layer, which could have resulted in rates of DNA damage up to three orders of magnitude greater than in present day Earth, as calculated by Cockell (2000); this higher rate of damage may have led as well to faster rates of change. Alternatively, both mechanisms could have contributed simultaneously.

Conclusion

This work shows for the first time that the molecular evolution of photochemical reaction centers is consistent with an early Archaean origin of photosynthesis as inferred from geochemical evidence. We show that the origin of photosynthesis is likely to have been followed by a relatively rapid diversification leading to the divergence of *oxygenic* and *anoxygenic* families of reaction center proteins in ancestral populations of bacteria. In addition, our results support the premise that some form of water oxidation is highly likely to have evolved before the divergence of D1 and D2; because the ancestral protein to D1 and D2 branched out from the root of Type II reaction center proteins, it is plausible that primordial forms of biological water oxidation contributed to the deposition of the Isua banded iron formations (Frei et al., 2016; Rosing & Frei, 2004) and it is consistent with the persistent detection of oxygen hundreds of millions of years prior to the GOE (Lyons et al., 2014). From the gene duplication event that permitted the divergence of D1 and D2, a billion years could have passed before the emergence of the standard form of D1 inherited by the last common ancestor of crown group Cyanobacteria. Our results also show that the last common ancestor of Cyanobacteria existed at a relatively advanced stage in the evolution of oxygenic photosynthesis and yet it is likely that this ancestor descended from a lineage with an ancestry that can be traced, without obvious discontinuity, to the origin of photosynthesis. From this perspective, and taking into consideration the extensive diversity of D1 proteins, it is plausible that transitional forms of oxygenic photosynthesis existed with

various degrees of complexity and efficiency at different stages following the origin of water oxidation catalysis in the ancestral homodimeric photosystem (Cardona et al., 2015).

Materials and Methods

Type II reaction center proteins covering the known diversity of phototrophic bacteria were selected for this analysis. These include the L and M subunits from representatives of the phyla Proteobacteria, Gemmatimonadetes, and Chloroflexi and D1 and D2 proteins specific to the phylum Cyanobacteria and photosynthetic eukaryotes. A representative selection of D1 forms were obtained from Cardona et al. (2015). These include the atypical D1 sequence found in the genome of *Gloeobacter kilaueensis* (Group 0), the super-rogue D1 or chlorophyll *f* synthase (Group 1), the rogue D1 or sentinel D1 (Group 2), the microaerobic D1 (Group 3), and the dominant form of D1 with sequences from both Cyanobacteria and photosynthetic eukaryotes (Group 4). In total 74 sequences were selected for the analysis, 30 D1, 20 D2, 10 L, and 14 M. Sequence alignments were performed using Clustal Omega employing ten combined guide trees and Hidden Markov Model iterations (Sievers et al., 2011). To confirm that the alignment conformed with known structures, the 3D structures of the D1, D2, L and M, from the crystal structures 3WU2 (Umena et al., 2011) of PSII and 2PRC (Lancaster & Michel, 1997) of the anoxyenic Type II reaction center were overlapped using the CEalign (Jia et al., 2004) plug-in for PyMOL (Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC) and structural homologous positions were cross-checked with the alignment. Maximum Likelihood phylogenetic analysis was performed using PhyML 3.1 (Guindon et al., 2010) using the Blosum62 model of amino acid substitution. The amino acid equilibrium frequencies, proportion of invariable sites, and across site rate variation were allowed to be determined by the software from the dataset. Tree search operations were set as the best from the Nearest Neighbor Interchange and Subtree Pruning and Regrafting method, with the tree topology optimized using five random starts. The tree, which reproduced the phylogeny of Type II reaction center proteins, was then used as input for the Bayesian relaxed molecular clock analyses.

The phylogeny of Type II reaction centers was cross-calibrated on D1 and D2 as listed in Table 4 and calibration points were allocated as presented in Fig. 4, red dots. Dates for the

Arabidopsis/*Populus* divergence, the appearance of the angiosperms, gymnosperms, and land plants were obtained from Clarke et al. (2011). Dates for the divergence of Coscinodiscophyceae (Sims et al., 2006), Floridiophyceae (Xiao et al., 2004), and Rhodophyta (Butterfield, 2000; Knoll et al., 2013) were also included in the calibration. Furthermore, well-known Cyanobacterial fossils were assigned to Nostocales (Golubic et al., 1995; Schirrmeister et al., 2016), Pleurocapsales (Golubic & Lee, 1999; Sergeev et al., 2002), and the first occurrence of multicellularity as represented by the early branching *Pseudanabeana* (Golubic & Lee, 1999; Schirrmeister et al., 2016; Sergeev et al., 2002). The oldest calibration point, point 11, was selected to be the branching of the Group 4 D1 of *Gloeobacter violaceous* and was set to be around the age for the GOE with a minimum of 2.45 Ga and no maximum age (Calibration 1). In some cases, a calibration point 11 of 2.7 Ga was used for comparison (Calibration 2). The calibration points on D1 were allocated on Group 4 as this type of D1 is the only one retained in all Cyanobacteria with PSII and it is the only type of D1 inherited by photosynthetic eukaryotes. The analysis by Cardona et al. (2015) indicated that the gene duplication events that led to the atypical forms and the microaerobic forms of D1 are likely to predate the last common ancestor of Cyanobacteria. Therefore, it is the dominant form of D1 (G4) that more closely reproduced a species trees as the emergence of this type of D1 is roughly coincidental with the last common ancestor of Cyanobacteria, as described in Cardona et al. (2015). The tree in Fig. 4 was calculated using a root prior of 3.5 ± 0.05 Ga as a likely age for the evolution of photosynthesis. The evidence for photosynthesis in 3.5 Ga Archaean rocks has been reviewed extensively recently, see for example Nisbet and Fowler (2014), Butterfield (2015), or Knoll (2015). Nevertheless, we tested the effect of the root prior on the estimated divergence times as shown in Table 1, by varying the root prior from 3.2 to 4.1 Ga and as shown in Fig. 5B from 0.8 to 4.2 Ga.

A Bayesian Markov chain Monte Carlo approach was used to calculate the estimated divergence times using Phylobayes 3.3f , which allows for the application of a relaxed log-normal autocorrelated molecular clock under the CAT + GTR + Γ model (Lartillot et al., 2009; Lartillot & Philippe, 2004) necessary for the implementation of flexible boundaries on the calibration points (Yang & Rannala, 2006). This model was compared with 1) a LG + Γ model that sets less flexible boundaries on the calibration points, 2) a CAT + Γ model assuming a uniform distribution of amino

acid equilibrium frequencies (Poisson), or 3) an uncorrelated gamma model where the rates of substitution can vary independently. The flexible bounds were set to allow for 2.5% tail probability falling outside each calibration boundary or 5% in the case of a single minimum boundary. All molecular clocks were computed using four discrete categories for the gamma distribution and four chains were run in parallel until convergence.

In this work we define the period of time between the duplication event that led to the divergence of D1 and D2 and the appearance of the ancestral standard D1 as **ΔT**. This value is calculated as the subtraction of the mean age of the latter node (Fig. 4, green dot) from the former's mean node age (Fig. 4, D0, orange dot). The instant rates of evolution were retrieved from the output files of Phylobayes, these rates are computed by the software as described elsewhere (Kishino et al., 2001; Lepage et al., 2007) and are expressed as amino acid changes per site per unit of time. The rate at node **D0** was termed $\mathbf{v_{max}}$ and a baseline rate of evolution during the Proterozoic was obtained as the average of all node rates in Group 4 D1 and D2 and denoted $\mathbf{v_{min}}$. All estimated divergence times for each node of each tree used in this analysis are provided in the supplementary data files.

Acknowledgements

This work was supported by an Imperial College London Junior Research Fellowship (T.C.), a Royal Society Dorothy Hodgkin Fellowship (P.S.-B.), the Biotechnology and Biological Sciences Research Council (BB/K002627/1 to A.W.R. and BB/L011206/1 to A.W.R.), and the Royal Society Wolfson Research Merit Award (A.W.R). The authors are immensely grateful to Prof. Euan Nisbet, Prof. Donald E. Canfield, Prof. Minik Rosing, and Dr. Andrea Fantuzzi for valuable comments and feedback. Computing resources were provided by the HPC facility of Imperial College London. We declare no conflict of interests.

References

Albalat, R., & Canestro, C. (2016). Evolution by gene loss. *Nature Reviews Genetics, 17*(7), 379-391. doi:10.1038/nrg.2016.39

Battistuzzi, F. U., & Hedges, S. B. (2009). A major clade of prokaryotes with ancient adaptations to life on land. *Molecular Biology and Evolution, 26*(2), 335-343. doi:10.1093/molbev/msn247

Beanland, T. J. (1990). Evolutionary relationships between Q-type photosynthetic reaction centers - hypothesis-testing using parsimony. *Journal of Theorerical Biology, 145*(4), 535-545. doi:10.1016/S0022-5193(05)80487-4

Bekker, A., Holland, H. D., Wang, P. L., Rumble, D., 3rd, Stein, H. J., Hannah, J. L., . . . Beukes, N. J. (2004). Dating the rise of atmospheric oxygen. *Nature, 427*(6970), 117-120. doi:10.1038/nature02260

Bengtson, S., Sallstedt, T., Belivanova, V., & Whitehouse, M. (2017). Three-dimensional preservation of cellular and subcellular structures suggests 1.6 billion-year-old crown-group red algae. *PLoS Biology, 15*(3), e2000735. doi:10.1371/journal.pbio.2000735

Blank, C. E., & Sanchez-Baracaldo, P. (2010). Timing of morphological and ecological innovations in the Cyanobacteria - a key to understanding the rise in atmospheric oxygen. *Geobiology, 8*(1), 1-23. doi:10.1111/j.1472-4669.2009.00220.x

Blankenship, R. E. (1992). Origin and early evolution of photosynthesis. *Photosynthesis Research, 33*, 91-111.

Bosak, T., Liang, B., Sim, M. S., & Petroff, A. P. (2009). Morphological record of oxygenic photosynthesis in conical stromatolites. *Proceedings of the National Academy of Sciences of the United States of America, 106*(27), 10939-10943. doi:10.1073/pnas.0900885106

Boucher, Y., Douady, C. J., Papke, R. T., Walsh, D. A., Boudreau, M. E., Nesbo, C. L., . . . Doolittle, W. F. (2003). Lateral gene transfer and the origins of prokaryotic groups. *Annual Review of Genetics, 37*, 283-328. doi:10.1146/annurev.genet.37.050503.084247

Brinkert, K., De Causmaecker, S., Krieger-Liszkay, A., Fantuzzi, A., & Rutherford, A. W. (2016). Bicarbonate-induced redox tuning in Photosystem II for regulation and protection. *Proceedings of the National Academy of Sciences of the United States of America, 113*(43), 12144-12149. doi:10.1073/pnas.1608862113

Brucker, R. M., & Bordenstein, S. R. (2012). Speciation by symbiosis. *Trends in Ecology & Evolution, 27*(8), 443-451. doi:10.1016/j.tree.2012.03.011

Bryant, D., Liu, Z., LI, T., Zhao, F., Klatt, C. G., Ward, D., . . . Overmann, J. (2012). Comparative and functional genomics of anoxygenic green bacteria from the taxa chlorobi, chloroflexi, and acidobacteria. In R. L. Burnap & W. Vermaas (Eds.), *Functional genomics and evolution of photosynthetic systems* (Vol. 33, pp. 47-102). Dordrecht Springer.

Bukh, J., Pietschmann, T., Lohmann, V., Krieger, N., Faulk, K., Engle, R. E., . . . Bartenschlager, R. (2002). Mutations that permit efficient replication of hepatitis c virus rna in huh-7 cells prevent productive replication in chimpanzees. *Proceedings of the National Academy of Sciences of the United States of America, 99*(22), 14416-14421. doi:10.1073/pnas.212532699

Butterfield, N. J. (2000). Bangiomorpha pubescens n. Gen., n. Sp.: Implications for the evolution of sex, multicellularity, and the mesoproterozoic/neoproterozoic radiation of eukaryotes. *Paleobiology, 26*(3), 386-404. doi:10.1666/0094-8373(2000)026<0386:Bpngns>2.0.Co;2

Butterfield, N. J. (2015). Proterozoic photosynthesis - a critical review. *Palaeontology, 58*(6), 953-972. doi:10.1111/pala.12211

Canestro, C., Albalat, R., Hjelmqvist, L., Godoy, L., Jornvall, H., & Gonzalez-Duarte, R. (2002). Ascidian and amphioxus Adh genes correlate functional and molecular features of the Adh family expansion during vertebrate evolution. *Journal of Molecular Evolution, 54*(1), 81-89. doi:10.1007/s00239-001-0020-2

Cardona, T. (2015). A fresh look at the evolution and diversification of photochemical reaction centers. *Photosynthesis Research, 126*(1), 111-134. doi:10.1007/s11120-014-0065-x

Cardona, T. (2016a). Origin of bacteriochlorophyll *a* and the early diversification of photosynthesis. *PLoS One, 11*(3), e0151250. doi:10.1371/journal.pone.0151250

Cardona, T. (2016b). Reconstructing the origin of oxygenic photosynthesis: Do assembly and photoactivation recapitulate evolution? *Frontiers in Plant Science, 7*, 257. doi:10.3389/fpls.2016.00257

Cardona, T. (2017). Photosystem II is a chimera of reaction centers. *Journal of Molecular Evolution, 84*(2-3), 149-151. doi:10.1007/s00239-017-9784-x

Cardona, T., Murray, J. W., & Rutherford, A. W. (2015). Origin and evolution of water oxidation before the last common ancestor of the Cyanobacteria. *Molecular Biology and Evolution, 32*(5), 1310-1328. doi:10.1093/molbev/msv024

Cardona, T., Sedoud, A., Cox, N., & Rutherford, A. W. (2012). Charge separation in Photosystem II: A comparative and evolutionary overview. *Biochimica et Biophysica Acta, 1817*(1), 26-43. doi:10.1016/j.bbabio.2011.07.012

Carrat, F., & Flahault, A. (2007). Influenza vaccine: The challenge of antigenic drift. *Vaccine, 25*(39-40), 6852-6862. doi:10.1016/j.vaccine.2007.07.027

Ciccarelli, F. D., Doerks, T., von Mering, C., Creevey, C. J., Snel, B., & Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science, 311*(5765), 1283-1287. doi:10.1126/science.1123061

Clarke, J. T., Warnock, R. C. M., & Donoghue, P. C. J. (2011). Establishing a time-scale for plant evolution. *New Phytologist, 192*(1), 266-301. doi:10.1111/j.1469-8137.2011.03794.x

Cockell, C. S. (2000). Ultraviolet radiation and the photobiology of earth's early oceans. *Origins of Life and Evolution of Biospheres, 30*(5), 467-499.

Cogdell, R. J., Howard, T. D., Bittl, R., Schlodder, E., Geisenheimer, I., & Lubitz, W. (2000). How carotenoids protect bacterial photosynthesis. *Philosophical Transactions of the Royal Society B, 355*(1402), 1345-1349.

Cornejo-Castillo, F. M., Cabello, A. M., Salazar, G., Sanchez-Baracaldo, P., Lima-Mendez, G., Hingamp, P., . . . Acinas, S. G. (2016). Cyanobacterial symbionts diverged in the late cretaceous towards lineage-specific nitrogen fixation factories in single-celled phytoplankton. *Nature Communications, 7*. doi:11071

10.1038/ncomms11071

Crawford, T. S., Hanning, K. R., Chua, J. P., Eaton-Rye, J. J., & Summerfield, T. C. (2016). Comparison of D1'- and D1-containing PS II reaction centre complexes under different environmental conditions in synechocystis sp. Pcc 6803. *Plant, Cell & Environment, 39*(8), 1715-1726. doi:10.1111/pce.12738

Czaja, A. D., Johnson, C. M., Beard, B. L., Roden, E. E., Li, W. Q., & Moorbath, S. (2013). Biological Fe oxidation controlled deposition of banded iron formation in the ca. 3770 Ma Isua supracrustal belt (West Greenland). *Earth and Planetary Science Letters, 363*, 192-203. doi:10.1016/j.epsl.2012.12.025

D'Souza, G., Waschina, S., Pande, S., Bohl, K., Kaleta, C., & Kost, C. (2014). Less is more: Selective advantages can explain the prevalent loss of biosynthetic genes in bacteria. *Evolution, 68*(9), 2559-2570. doi:10.1111/evo.12468

David, L. A., & Alm, E. J. (2011). Rapid evolutionary innovation during an archaean genetic expansion. *Nature, 469*(7328), 93-96. doi:10.1038/Nature09649

De Las Rivas, J., Balsera, M., & Barber, J. (2004). Evolution of oxygenic photosynthesis: Genome-wide analysis of the OEC extrinsic proteins. *Trends in Plant Science, 9*(1), 18-25. doi:10.1016/j.tplants.2003.11.007

Debus, R. J. (2001). Amino acid residues that modulate the properties of tyrosine Y_Z and the manganese cluster in the water oxidizing complex of Photosystem II. *Biochimica et Biophysica Acta, 1503*(1-2), 164-186. doi:10.1016/S0005-2728(00)00221-8

Deisenhofer, J., & Michel, H. (1989). Nobel lecture. The photosynthetic reaction centre from the purple bacterium *Rhodopseudomonas viridis*. *EMBO Journal, 8*(8), 2149-2170.

Delwiche, C. F., & Palmer, J. D. (1996). Rampant horizontal transfer and duplication of rubisco genes in Eubacteria and plastids. *Molecular Biology and Evolution, 13*(6), 873-882.

Di Rienzi, S. C., Sharon, I., Wrighton, K. C., Koren, O., Hug, L. A., Thomas, B. C., . . . Ley, R. E. (2013). The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling toCyanobacteria. *Elife, 2*. doi:10.7554/eLife.01102

Dobakova, M., Tichy, M., & Komenda, J. (2007). Role of the PsbI protein in Photosystem II assembly and repair in the cyanobacterium *Synechocystis* sp PCC 6803. *Plant Physiology, 145*(4), 1681-1691. doi:10.1104/pp.107.107805

Dutton, P. L., Leight, J. S., & Seibert, M. (1972). Primary processes in photosynthesis: In situ ESR studies on the light induced oxidized and triplet state of reaction center bacteriochlorophyll. *Biochemical and Biophysical Research Communications, 46*(2), 406-413.

Falcon, L. I., Magallon, S., & Castillo, A. (2010). Dating the Cyanobacterial ancestor of the chloroplast. *ISME Journal, 4*(6), 777-783. doi:10.1038/ismej.2010.2

Feng, D. F., Cho, G., & Doolittle, R. F. (1997). Determining divergence times with a protein clock: Update and reevaluation. *Proceedings of the National Academy of Sciences of the United States of America, 94*(24), 13028-13033.

Ferreira, K. N., Iverson, T. M., Maghlaoui, K., Barber, J., & Iwata, S. (2004). Architecture of the photosynthetic oxygen-evolving center. *Science, 303*(5665), 1831-1838. doi:10.1126/science.1093087

Fischer, W. W., Hemp, J., & Johnson, J. E. (2016). Evolution of oxygenic photosynthesis. *Annual Review of Earth and Planetary, 44*, 647-683. doi:10.1146/annurev-earth-060313-054810

Franzen, L. G., Hansson, O., & Andreasson, L. E. (1985). The roles of the extrinsic subunits in Photosystem II as revealed by electron paramagnetic resonance. *Biochimica et Biophysica Acta, 808*(1), 171-179. doi:10.1016/0005-2728(85)90040-4

Frei, R., Crowe, S. A., Bau, M., Polat, A., Fowle, D. A., & Døssing, L. N. (2016). Oxidative elemental cycling under the low o_2 eoarchean atmosphere. *Scientific Reports, 6*, 21058. doi:10.1038/srep21058

Georgelis, N., Braun, E. L., & Hannah, L. C. (2008). Duplications and functional divergence of adp-glucose pyrophosphorylase genes in plants. *BMC Evolutionary Biology, 8*. doi:232
10.1186/1471-2148-8-232

Golubic, S., & Lee, S. J. (1999). Early Cyanobacterial fossil record: Preservation, palaeoenvironments and identification. *European Journal of Phycology, 34*(4), 339-348. doi:10.1017/S0967026299002279

Golubic, S., Sergeev, V. N., & Knoll, A. H. (1995). Mesoproterozoic archaeoellipsoides: Akinetes of heterocystous Cyanobacteria. *Lethaia, 28*(4), 285-298. doi:10.1111/j.1502-3931.1995.tb01817.x

Grassineau, N. V., Abell, P. I., Appel, P. W. U., Lowry, D., & Nisbet, E. G. (2006). Early life signatures in sulfur and carbon isotopes from Isua, Barberton, Wabigoon (Steep Rock), and Belingwe Greenstone belts (3.8 to 2.7 Ga). *Geological Society of America Memoirs, 198*, 33-52.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology, 59*(3), 307-321. doi:10.1093/sysbio/syq010

Harel, A., Karkar, S., Cheng, S., Falkowski, P. G., & Bhattacharya, D. (2015). Deciphering primordial Cyanobacterial genome functions from protein network analysis. *Current Biology, 25*(5), 628-634. doi:10.1016/j.cub.2014.12.061

Ho, M. Y., Shen, G., Canniffe, D. P., Zhao, C., & Bryant, D. A. (2016). Light-dependent chlorophyll *f* synthase is a highly divergent paralog of PsbA of Photosystem II. *Science, 353*(6302). doi:10.1126/science.aaf9178

Ho, S. Y. W., & Duchene, S. (2014). Molecular-clock methods for estimating evolutionary rates and timescales. *Molecular Ecology, 23*(24), 5947-5965. doi:10.1111/mec.12953

Honda, D., Yokota, A., & Sugiyama, J. (1999). Detection of seven major evolutionary lineages in Cyanobacteria based on the 16S rRNA gene sequence analysis with new sequences of five marine *Synechococcus* strains. *Journal of Molecular Evolution, 48*(6), 723-739. doi:10.1007/Pl00006517

Huang, Y., Zeng, Y., Lu, H., Feng, H., Zeng, Y., & Koblizek, M. (2016). Novel *acsf* gene primers revealed a diverse phototrophic bacterial population, including Gemmatimonadetes, in lake Taihu (China). *Applied Environmental Microbiology, 82*(18), 5587-5594. doi:10.1128/AEM.01063-16

Igarashi, N., Harada, J., Nagashima, S., Matsuura, K., Shimada, K., & Nagashima, K. V. (2001). Horizontal transfer of the photosynthesis gene cluster and operon rearrangement in purple bacteria. *Journal of Molecular Evolution, 52*(4), 333-341. doi:10.1007/s002390010163

Jia, Y. T., Dewey, G., Shindyalov, I. N., & Bourne, P. E. (2004). A new scoring function and associated statistical significance for structure alignment by CE. *Journal of Computational Biology, 11*(5), 787-799. doi:10.1089/1066527042432260

Jun, S. R., Sims, G. E., Wu, G. H. A., & Kim, S. H. (2010). Whole-proteome phylogeny of prokaryotes by feature frequency profiles: An alignment-free method with optimal feature resolution. *Proceedings of the National Academy of Sciences of the United States of America, 107*(1), 133-138. doi:10.1073/pnas.0913033107

Kagan, R. M., McFadden, H. J., McFadden, P. N., OConnor, C., & Clarke, S. (1997). Molecular phylogenetics of a protein repair methyltransferase. *Comparative Biochemistry and Physiology - Part B: Biochemistry & Molecular Biology, 117*(3), 379-385. doi:10.1016/S0305-0491(96)00333-1

Kamikawa, R., Moog, D., Zauner, S., Tanifuji, G., Ishida, K. I., Miyashita, H., . . . Inagaki, Y. (2017). A non-photosynthetic diatom reveals early steps of reductive evolution in plastids. *Molecular Biology and Evolution*. doi:10.1093/molbev/msx172

Kay, E. H., & Hoekstra, H. E. (2008). Rodents. *Current Biology, 18*(10), R406-R410. doi:10.1016/j.cub.2008.03.019

Kim, H., Li, H., Maresca, J. A., Bryant, D. A., & Savikhin, S. (2007). Triplet exciton formation as a novel photoprotection mechanism in chlorosomes of *Chlorobium tepidum*. *Biophysical Journal, 93*(1), 192-201. doi:10.1529/biophysj.106.103556

Kishino, H., Thorne, J. L., & Bruno, W. J. (2001). Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Molecular Biology and Evolution, 18*(3), 352-361.

Klatt, C. G., Wood, J. M., Rusch, D. B., Bateson, M. M., Hamamura, N., Heidelberg, J. F., . . . Ward, D. M. (2011). Community ecology of hot spring Cyanobacterial mats: Predominant populations and their functional potential. *ISME Journal, 5*(8), 1262-1278. doi:10.1038/ismej.2011.73

Knoll, A. H. (2015). Paleobiological perspectives on early microbial evolution. *Cold Spring Harbor Perspectives Biology, 7*(7), a018093. doi:10.1101/cshperspect.a018093

Knoll, A. H., Worndle, S., & Kah, L. C. (2013). Covariance of microfossil assemblages and microbialite textures across an upper mesoproterozoic carbonate platform. *Palaios, 28*(7-8), 453-470. doi:10.2110/palo.2013.p13-005r

Komenda, J., Sobotka, R., & Nixon, P. J. (2012). Assembling and maintaining the Photosystem II complex in chloroplasts and Cyanobacteria. *Current Opinion in Plant Biology, 15*(3), 245-251. doi:10.1016/j.pbi.2012.01.017

Komenda, J., Tichy, M., Prasil, O., Knoppova, J., Kuvikova, S., de Vries, R., & Nixon, P. J. (2007). The exposed N-terminal tail of the D1 subunit is required for rapid D1 degradation during

Photosystem II repair in *Synechocystis* sp PCC 6803. *Plant Cell, 19*(9), 2839-2854. doi:10.1105/tpc.107.053868

Koskiniemi, S., Sun, S., Berg, O. G., & Andersson, D. I. (2012). Selection-driven gene loss in bacteria. *PLoS Genet, 8*(6), e1002787. doi:10.1371/journal.pgen.1002787

Koyama, K., Suzuki, H., Noguchi, T., Akimoto, S., Tsuchiya, T., & Mimuro, M. (2008). Oxygen evolution in the thylakoid-lacking cyanobacterium G*loeobacter violaceus* PCC 7421. *Biochimica et Biophysica Acta, 1777*(4), 369-378. doi:10.1016/j.bbabio.2008.01.009

Krieger-Liszkay, A., Fufezan, C., & Trebst, A. (2008). Singlet oxygen production in Photosystem II and related protection mechanism. *Photosynthesis Research, 98*(1-3), 551-564. doi:10.1007/s11120-008-9349-3

Krynicka, V., Shao, S., Nixon, P. J., & Komenda, J. (2015). Accessibility controls selective degradation of Photosystem II subunits by FtsH protease. *Nature Plants, 1*, 15168. doi:10.1038/nplants.2015.168

Ku, C., Nelson-Sathi, S., Roettger, M., Sousa, F. L., Lockhart, P. J., Bryant, D., . . . Martin, W. F. (2015). Endosymbiotic origin and differential loss of eukaryotic genes. *Nature, 524*(7566), 427-432. doi:10.1038/nature14963

Lancaster, C. R. D., & Michel, H. (1997). The coupling of light-induced electron transfer and proton uptake as derived from crystal structures of reaction centres from *Rhodopseudomonas viridis* modified at the binding site of the secondary quinone, Q_B. *Structure, 5*(10), 1339-1359. doi:10.1016/S0969-2126(97)00285-2

Lartillot, N., Lepage, T., & Blanquart, S. (2009). Phylobayes 3: A bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics, 25*(17), 2286-2288. doi:10.1093/bioinformatics/btp368

Lartillot, N., & Philippe, H. (2004). A bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution, 21*(6), 1095-1109. doi:10.1093/molbev/msh112

Lepage, T., Bryant, D., Philippe, H., & Lartillot, N. (2007). A general comparison of relaxed molecular clock models. *Molecular Biology and Evolution, 24*(12), 2669-2680. doi:10.1093/molbev/msm193

Lewis, C. A., Crayle, J., Zhou, S. T., Swanstrom, R., & Wolfenden, R. (2016). Cytosine deamination and the precipitous decline of spontaneous mutation during earth's history. *Proceedings of the National Academy of Sciences of the United States of America, 113*(29), 8194-8199. doi:10.1073/pnas.1607580113

Linke, K., & Ho, F. M. (2013). Water in Photosystem II: Structural, functional and mechanistic considerations. *Biochimica et Biophysica Acta, 1837*(1), 14-32. doi:10.1016/j.bbabio.2013.08.003

Lyons, T. W., Reinhard, C. T., & Planavsky, N. J. (2014). The rise of oxygen in earth's early ocean and atmosphere. *Nature, 506*(7488), 307-315. doi:10.1038/nature13068

Magnuson, A., & Cardona, T. (2016). Thylakoid membrane function in heterocysts. *Biochimica et Biophysica Acta, 1857*(3), 309-319. doi:10.1016/j.bbabio.2015.10.016

Marin, J., Battistuzzi, F. U., Brown, A. C., & Hedges, S. B. (2017). The timetree of prokaryotes: New insights into their evolution and speciation. *Molecular Biology and Evolution, 34*, 437-446. doi:10.1093/molbev/msw245

Matsunami, M., Yoshioka, T., Minoura, T., Okano, Y., & Muto, Y. (2011). Evolutionary features and intracellular behavior of the PrtB protein. *Biochemical Genetics, 49*(7-8), 458-473. doi:10.1007/s10528-011-9422-z

Melo, T. B., Frigaard, N. U., Matsuura, K., & Naqvi, K. R. (2000). Electronic energy transfer involving carotenoid pigments in chlorosomes of two green bacteria: *Chlorobium tepidum* and *Chloroflexus aurantiacus*. *Spectrochim Acta Mol Biomol Spectrosc, 56*(10), 2001-2010. doi:10.1016/S1386-1425(00)00289-4

Michel, H., & Deisenhofer, J. (1988). Relevance of the photosynthetic reaction center from purple bacteria to the structure of Photosystem II. *Biochemistry, 27*(1), 1-7. doi:10.1021/Bi00401a001

Moore, R. B., Obornik, M., Janouskovec, J., Chrudimsky, T., Vancova, M., Green, D. H., . . . Carter, D. A. (2008). A photosynthetic alveolate closely related to apicomplexan parasites. *Nature, 451*(7181), 959-963. doi:10.1038/nature06635

Mulkidjanian, A. Y., Koonin, E. V., Makarova, K. S., Mekhedov, S. L., Sorokin, A., Wolf, Y. I., . . . Galperin, M. Y. (2006). The Cyanobacterial genome core and the origin of photosynthesis. *Proceedings of the National Academy of Sciences of the United States of America, 103*(35), 13126-13131. doi:10.1073/pnas.0605709103

Murray, J. W. (2012). Sequence variation at the oxygen-evolving centre of Photosystem II: A new class of 'rogue' Cyanobacterial d1 proteins. *Photosynthesis Research, 110*(3), 177-184. doi:10.1007/s11120-011-9714-5

Nakamura, Y., Kaneko, T., Sato, S., Mimuro, M., Miyashita, H., Tsuchiya, T., . . . Tabata, S. (2003). Complete genome structure of *Gloeobacter violaceus* PCC 7421, a cyanobacterium that lacks thylakoids. *DNA Research, 10*(4), 137-145. doi:10.1093/dnares/10.4.137

Nakayama, T., Kamikawa, R., Tanifuji, G., Kashiyama, Y., Ohkouchi, N., Archibald, J. M., & Inagaki, Y. (2014). Complete genome of a nonphotosynthetic cyanobacterium in a diatom reveals recent adaptations to an intracellular lifestyle. *Proceedings of the National Academy of Sciences of the United States of America, 111*(31), 11407-11412. doi:10.1073/pnas.1405222111

Nickelsen, J., & Rengstl, B. (2013). Photosystem II assembly: From Cyanobacteria to plants. *Annual Review of Plant Biology, 64*, 609-635. doi:10.1146/annurev-arplant-050312-120124

Nisbet, E. G., & Fowler, C. F. R. (2014). The early history of life. In K. D. M. & W. H. Schlesinger (Eds.), *Treatise on geochemistry* (2nd ed., Vol. 10, pp. 1-42). Amsterdam: Elsevier Science.

Nitschke, W., & Rutherford, A. W. (1991). Photosynthetic reaction centers - variations on a common structural theme. *Trends in Biochemical Sciences, 16*(7), 241-245. doi:10.1016/0968-0004(91)90095-D

Niwa, S., Yu, L. J., Takeda, K., Hirano, Y., Kawakami, T., Wang-Otomo, Z. Y., & Miki, K. (2014). Structure of the LH1-RC complex from *Thermochromatium tepidum* at 3.0 a. *Nature, 508*(7495), 228-232. doi:10.1038/nature13197

Nixon, P. J., & Diner, B. A. (1994). Analysis of water-oxidation mutants constructed in the cyanobacterium *Synechocystis* sp. PCC 6803. *Biochemical Society Transactions, 22*(2), 338-343.

Nixon, P. J., Michoux, F., Yu, J. F., Boehm, M., & Komenda, J. (2010). Recent advances in understanding the assembly and repair of Photosystem II. *Annals of Botany, 106*(1), 1-16. doi:10.1093/Aob/Mcq059

Nutman, A. P., Bennett, V. C., Friend, C. R. L., Van Kranendonk, M. J., & Chivas, A. R. (2016). Rapid emergence of life shown by discovery of 3,700-million-year-old microbial structures. *Nature*. doi::10.1038/nature19355

Planavsky, N. J., Asael, D., Hofmann, A., Reinhard, C. T., Lalonde, S. V., Knudsen, A., . . . Rouxel, O. J. (2014). Evidence for oxygenic photosynthesis half a billion years before the great oxidation event. *Nature Geoscience, 7*(4), 283-286. doi:10.1038/ngeo2122

Ravin, N. V., Gruzdev, E. V., Beletsky, A. V., Mazur, A. M., Prokhortchouk, E. B., Filyushin, M. A., . . . Skryabin, K. G. (2016). The loss of photosynthetic pathways in the plastid and nuclear genomes of the non-photosynthetic mycoheterotrophic eudicot *monotropa hypopitys*. *BMC Plant Biology, 16*(Suppl 3), 238. doi:10.1186/s12870-016-0929-7

Raymond, J. (2009). The role of horizontal gene transfer in photosynthesis, oxygen production, and oxygen tolerance. *Methods in Molecular Biology, 532*, 323-338. doi:10.1007/978-1-60327-853-9_19

Riding, R., Fralick, P., & Liang, L. Y. (2014). Identification of an archean marine oxygen oasis. *Precambrian Research, 251*, 232-237. doi:10.1016/j.precamres.2014.06.017

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J. F., . . . Woyke, T. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature, 499*(7459), 431-437. doi:10.1038/Nature12352

Roose, J. L., Frankel, L. K., Mummadisetti, M. P., & Bricker, T. M. (2016). The extrinsic proteins of Photosystem II: Update. *Planta, 243*(4), 889-908. doi:10.1007/s00425-015-2462-6

Rosing, M. T. (1999). C-13-depleted carbon microparticles in > 3700-Ma sea-floor sedimentary rocks from west greenland. *Science, 283*(5402), 674-676. doi:10.1126/science.283.5402.674

Rosing, M. T., & Frei, R. (2004). U-rich archaean sea-floor sediments from Greenland - Indications of > 3700 Ma oxygenic photosynthesis. *Earth and Planetary Science Letters, 217*(3-4), 237-244. doi:10.1016/S0012-821x(03)00609-5

Rutherford, A. W. (1989). Photosystem II, the water-splitting enzyme. *Trends in Biochemical Sciences, 14*(6), 227-232. doi:10.1016/0968-0004(89)90032-7

Rutherford, A. W., & Faller, P. (2003). Photosystem II: Evolutionary perspectives. *Philosophical Transactions of the Royal Society B, 358*(1429), 245-253. doi:10.1098/rstb.2002.1186

Rutherford, A. W., & Nitschke, W. (1996). Photosystem II and the quinone–iron-containing reaction centers. In H. Baltscheffsky (Ed.), *Origin and evolution of biological energy conversion* (pp. 143–175). New York, N. Y.: VCH.

Rutherford, A. W., Osyczka, A., & Rappaport, F. (2012). Back-reactions, short-circuits, leaks and other energy wasteful reactions in biological electron transfer: Redox tuning to survive life in O_2. *FEBS Lett, 586*(5), 603-616. doi:10.1016/j.febslet.2011.12.039

Rutherford, A. W., Paterson, D. R., & Mullet, J. E. (1981). A light-induced spin-polarized triplet detected by epr in Photosystem II reaction centers. *Biochimica et Biophysica Acta, 635*(2), 205-214.

Sadekar, S., Raymond, J., & Blankenship, R. E. (2006). Conservation of distantly related membrane proteins: Photosynthetic reaction centers share a common structural core. *Molecular Biology and Evolution, 23*(11), 2001-2007. doi:10.1093/molbev/msl079

Sanchez-Baracaldo, P. (2015). Origin of marine planktonic Cyanobacteria. *Scientific Reports, 5*, 17418. doi:10.1038/srep17418

Sanchez-Baracaldo, P., Hayes, P. K., & Blank, C. E. (2005). Morphological and habitat evolution in the Cyanobacteria using a compartmentalization approach. *Geobiology, 3*(3), 145-165. doi:10.1111/j.1472-4669.2005.00050.x

Satkoski, A. M., Beukes, N. J., Li, W., Beard, B. L., & Johnson, C. M. (2015). A redox-stratified ocean 3.2 billion years ago. *Earth and Planetary Science Letters, 430*, 43-53.

Saw, J. H. W., Schatz, M., Brown, M. V., Kunkel, D. D., Foster, J. S., Shick, H., . . . Donachie, S. P. (2013). Cultivation and complete genome sequencing of *Gloeobacter kilaueensis* sp nov., from a lava cave in Kilauea caldera, Hawai'i. *PLoS One, 8*(10). doi:10.1371/journal.pone.0076376

Schidlowski, M. (1988). A 3,800-million-year isotopic record of life from carbon in sedimentary-rocks. *Nature, 333*(6171), 313-318. doi:10.1038/333313a0

Schirrmeister, B. E., de Vos, J. M., Antonelli, A., & Bagheri, H. C. (2013). Evolution of multicellularity coincided with increased diversification of Cyanobacteria and the Great Oxidation Event. *Proceedings of the National Academy of Sciences of the United States of America, 110*(5), 1791-1796. doi:10.1073/pnas.1209927110

Schirrmeister, B. E., Gugger, M., & Donoghue, P. C. J. (2015). Cyanobacteria and the Great Oxidation Event: Evidence from genes and fossils. *Palaeontology, 58*(5), 769-785. doi:10.1111/pala.12178

Schirrmeister, B. E., Sanchez-Baracaldo, P., & Wacey, D. (2016). Cyanobacterial evolution during the Precambrian. *International Journal of Astrobiology, 15*, 187-204. doi:10.1017/S1473550415000579.

Segata, N., Bornigen, D., Morgan, X. C., & Huttenhower, C. (2013). Phylophlan is a new method for improved phylogenetic and taxonomic placement of microbes. *Nature Communications, 4*. doi:10.1038/Ncomms3304

Sergeev, V. N., Gerasimenko, L. M., & Zavarzin, G. A. (2002). Proterozoic history and present state of Cyanobacteria. *Mikrobiologiia, 71*(6), 725-740.

Shih, P. M., Hemp, J., Ward, L. M., Matzke, N. J., & Fischer, W. W. (2017). Crown group Oxyphotobacteria postdate the rise of oxygen. *Geobiology, 15*(1), 19-29. doi:10.1111/gbi.12200

Shih, P. M., Wu, D., Latifi, A., Axen, S. D., Fewer, D. P., Talla, E., . . . Kerfeld, C. A. (2013). Improving the coverage of the Cyanobacterial phylum using diversity-driven genome sequencing. *Proceedings of the National Academy of Sciences of the United States of America, 110*(3), 1053-1058. doi:10.1073/pnas.1217107110

Sicora, C. I., Ho, F. M., Salminen, T., Styring, S., & Aro, E. M. (2009). Transcription of a "silent" Cyanobacterial *psbA* gene is induced by microaerobic conditions. *Biochimica et Biophysica Acta, 1787*(2), 105-112. doi:10.1016/j.bbabio.2008.12.002

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W. Z., . . . Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Molecular Systems Biology, 7*. doi:10.1038/msb.2011.75

Sims, P. A., Mann, D. G., & Medlin, L. K. (2006). Evolution of the diatoms: Insights from fossil, biological and molecular data. *Phycologia, 45*(4), 361-402. doi:10.2216/05-22.1

Soo, R. M., Hemp, J., Parks, D. H., Fischer, W. W., & Hugenholtz, P. (2017). On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. *Science, 355*(6332), 1436-1440. doi:10.1126/science.aal3794

Soo, R. M., Skennerton, C. T., Sekiguchi, Y., Imelfort, M., Paech, S. J., Dennis, P. G., . . . Hugenholtz, P. (2014). An expanded genomic representation of the phylum Cyanobacteria. *Genome Biology and Evolution, 6*(5), 1031-1045. doi:10.1093/Gbe/Evu073

Soo, R. M., Woodcroft, B. J., Parks, D. H., Tyson, G. W., & Hugenholtz, P. (2015). Back from the dead; the curious tale of the predatory cyanobacterium *Vampirovibrio chlorellavorus*. *PeerJ, 3*, e968. doi:10.7717/peerj.968

Sousa, F. L., Shavit-Grievink, L., Allen, J. F., & Martin, W. F. (2013). Chlorophyll biosynthesis gene evolution indicates photosystem gene duplication, not photosystem merger, at the origin of oxygenic photosynthesis. *Genome Biology and Evolution, 5*(1), 200-216. doi:10.1093/Gbe/Evs127

Suga, M., Akita, F., Hirata, K., Ueno, G., Murakami, H., Nakajima, Y., . . . Shen, J. R. (2015). Native structure of Photosystem II at 1.95 Angstrom resolution viewed by femtosecond X-ray pulses. *Nature, 517*(7532), 99-U265. doi:10.1038/nature13991

Sugiura, M., Nakamura, M., Koyama, K., & Boussac, A. (2015). Assembly of oxygen-evolving Photosystem II efficiently occurs with the apo-Cytb_{559} but the holo-Cytb_{559} accelerates the recovery of a functional enzyme upon photoinhibition. *Biochimica et Biophysica Acta, 1847*(2), 276-285. doi:10.1016/j.bbabio.2014.11.009

Sullivan, M. B., Lindell, D., Lee, J. A., Thompson, L. R., Bielawski, J. P., & Chisholm, S. W. (2006). Prevalence and evolution of core Photosystem II genes in marine Cyanobacterial viruses and their hosts. *PLoS Biology, 4*(8), 1344-1357. doi:10.1371/journal.pbio.0040234

Swingley, W. D., Blankenship, R. E., & Raymond, J. (2009). Evolutionary relationships among purple photosynthetic bacteria and the origin of proteobacterial photosynthetic systems. In C. N.

Hunter, F. Daldal, M. C. Thurnauer, & B. J. T. (Eds.), *The purple phototrophic bacteria* (pp. 17-29). Netherlands: Springer.

Tank, M., Thiel, V., & Imhoff, J. F. (2009). Phylogenetic relationship of phototrophic purple sulfur bacteria according to *pufL* and *pufM* genes. *International Microbiology, 12*(3), 175-185. doi:10.2436/20.1501.01.96

Thompson, A. W., Foster, R. A., Krupke, A., Carter, B. J., Musat, N., Vaulot, D., . . . Zehr, J. P. (2012). Unicellular cyanobacterium symbiotic with a single-celled eukaryotic alga. *Science, 337*(6101), 1546-1550. doi:10.1126/science.1222700

Tice, M. M., & Lowe, D. R. (2004). Photosynthetic microbial mats in the 3,416-myr-old ocean. *Nature, 431*(7008), 549-552. doi:10.1038/nature02888

Tsukatani, Y., Romberger, S. P., Golbeck, J. H., & Bryant, D. A. (2012). Isolation and characterization of homodimeric Type-I reaction center complex from *Candidatus* Chloracidobacterium thermophilum, an aerobic chlorophototroph. *Journal of Biological Chemistry, 287*(8), 5720-5732. doi:10.1074/jbc.M111.323329

Umena, Y., Kawakami, K., Shen, J. R., & Kamiya, N. (2011). Crystal structure of oxygen-evolving Photosystem II at a resolution of 1.9 A. *Nature, 473*(7345), 55-60. doi:10.1038/nature09913

Vass, I., & Cser, K. (2009). Janus-faced charge recombinations in Photosystem II photoinhibition. *Trends in Plant Science, 14*(4), 200-205. doi:10.1016/j.tplants.2009.01.009

Ward, L. M., Kirschvink, J. L., & Fischer, W. W. (2016). Timescales of oxygenation following the evolution of oxygenic photosynthesis. *Origins of Life and Evolution of Biospheres, 46*(1), 51-65. doi:10.1007/s11084-015-9460-3

Wegener, K. M., Nagarajan, A., & Pakrasi, H. B. (2015). An atypical PsbA gene encodes a sentinel D1 protein to form a physiologically relevant inactive Photosystem II complex in Cyanobacteria. *Journal of Biological Chemistry, 290*(6), 3764-3774. doi:10.1074/jbc.M114.604124

Weiss, M. C., Sousa, F. L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-Sathi, S., & Martin, W. F. (2016). The physiology and habitat of the Last Universal Common Ancestor. *Nature Microbiology, 1*(9), 16116. doi:10.1038/nmicrobiol.2016.116

Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N. N., . . . Eisen, J. A. (2009). A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature, 462*(7276), 1056-1060. doi:10.1038/nature08656

Xiao, S., Knoll, A. H., Yuan, X., & Pueschel, C. M. (2004). Phosphatized multicellular algae in the Neoproterozoic Doushantuo formation, China, and the early evolution of Florideophyte red algae. *American Journal of Botany, 91*(2), 214-227. doi:10.3732/ajb.91.2.214

Xiong, J., Inoue, K., & Bauer, C. E. (1998). Tracking molecular evolution of photosynthesis by characterization of a major photosynthesis gene cluster from *Heliobacillus mobilis*. *Proceedings of the National Academy of Sciences of the United States of America, 95*(25), 14851-14856. doi:10.1073/pnas.95.25.14851

Yang, Z., & Rannala, B. (2006). Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution, 23*(1), 212-226. doi:10.1093/molbev/msj024

Zeng, Y. H., Feng, F. Y., Medova, H., Dean, J., & Koblizek, M. (2014). Functional Type 2 photosynthetic reaction centers found in the rare bacterial phylum Gemmatimonadetes. *Proceedings of the National Academy of Sciences of the United States of America, 111*(21), 7795-7800. doi:10.1073/pnas.1400295111

Figure legends

Figure 1. Schematic representation of the evolution of Type II reaction centers. (a) All reaction centers have common ancestry and descended from a homodimeric reaction center, marked **A**. From **A**, two new reaction centers emerged, one ancestral to all Type II reaction centers (**II**) and a second ancestral to all Type I reaction centers. This is the earliest diversification event of reaction center proteins that can be inferred from sequence and structural data and it is marked **1**. The ancestral Type II reaction center protein (**II**) gave rise to two new proteins, one ancestral to D1 and D2, named **D0** and a second ancestral to L and M named **K**. The ancestral L and M subunits further diversify into Chloroflexi-type and Proteobacteria-type L and M subunits (**5**). Step **6** indicates that Type I reaction center proteins also diversified in parallel to Type II reaction center proteins. (b) Evolution of cyanobacterial D1 and D2, modified from Cardona et al. (2015). The conserved structural and functional traits between D1 and D2 indicate that the ancestral homodimeric photosystem was able to oxidize water, this is represented by a fuchsia spot in each **D0** monomer. However, ancestral reconstructions of **D0** suggest that the primordial catalytic cluster was not identical to that of standard D1, but may have been slightly modified at the first and second coordination spheres (Cardona et al 2015). The lack of fuchsia dot in the G0 sequence indicates the absence of ligands to the Mn_4CaO_5 cluster. This sequence has acquired changed that are predicated to impair function and it is not known whether it binds a cluster or cofactor if incorporated into a fully assembled PSII. The pentagon in G1 indicates the possible binding of chlorophyll *a*, as it is suggested that this D1 is necessary for the synthesis of chlorophyll *f* (Ho et al., 2016). The green triangle in the G2 sequences indicates the possibility of binding of a modified cluster as these sequences still retain some of the ligands to the Mn_4CaO_5 cluster. **ΔT** marks the span of time between **D0** and the appearance of the ancestral standard form of D1

Figure 2. Structural comparisons of Type II reaction center proteins. (*a*) Four conserved structural domains are shared by D1 and D2 not present in L and M: these are highlighted in orange and marked from 1 to 4 and their functions are described in the text. These four structural domains are required to support water oxidation, repair, assembly, and photoprotection. D1 and D2 are from the crystal

structure of PSII from *Thermosynechococcus vulcanus*, PDB: 3WU2 (Umena et al., 2011) and L and M from *Thermochromatium tepidum*, PDB: 3WMM (Niwa et al., 2014). (*b*) Structure of the Mn_4CaO_5 cluster in D1 and the corresponding site in D2. The residues colored in gray are provided by D1 and D2 protein and those colored in orange by CP43 and CP47. Notice that while a redox Y-H pair is conserved in D1 and D2, D2/CP47 evolved to stop the binding of metals by replacing ligands for phenylalanine residues creating a hydrophobic patch. Such a site is not found in any other Type II or Type I reaction center protein and it is unique to D2 (Cardona et al., 2015). (*c*) Overlap of D1 and D2 highlighting some of the conserved traits as described in the text

Figure 3. Decrease of sequence identity of D1 and D2 proteins as a function of divergence time. D1 subunits are shown in gray and D2 in orange. The divergence time between pairs of species is plotted against the level of sequence identity as tabulated in supplementary Table S1. The red circle, placed at 79.2% corresponds to the average of sequence identity of the three distinct Group 4 D1 sequences of *Gloeobacter violaceous* in comparison with that of *Cyanidioschyzon merolae*. The light orange bar marks the GOE. The dashed line is fitted from a linear function and shows that over a period of about 2.0 Ga no dramatic changes in the rates of evolution of D1 and D2 are observed. It also shows that a constant rate of evolution is unlikely across all times (red dashed line). Fit parameters for a linear regression are shown in supplementary Table S2. The gray dots around 3.5-3.8 Ga marks an approximated location for the earliest events in the history of photosynthesis, that is the divergence of D1 and D2 (~27% sequence identity), the divergence of anoxygenic (L/M) and oxygenic (D1/D2) reaction center proteins (~17%), and the divergence of Type I and Type II reaction center proteins (≤5%). The curved blue line highlights that any scenario for the diversification of reaction centers after the origin of life requires faster rates of evolution at the earliest stages in the evolution of photosynthesis

Figure 4. Molecular clock of Type II reaction center proteins. A log-normal autocorrelated relaxed clock is shown implementing a CAT + GTR + Γ non-parametric model with flexible boundaries on the calibration points. Red dots are calibration points as described in Materials and Methods. The gray

dot denoted **II**, represents the ancestral Type II reaction center protein, as schematized in Fig. 1A. The orange dot (**D0**) marks the initial divergence of D1 and D2. The violet dot marks the divergence point between G2 atypical D1 sequences and standard D1. The green dot marks the divergence point between the microaerobic D1 forms (G3) and the dominant form of D1 (G4). This point represents the last common ancestral protein to all standard D1 forms known to have a complete ligand sphere to the Mn_4CaO_5 cluster. The blue dot represents the origin of the dominant form of D1 inherited by all extant Cyanobacteria and photosynthetic eukaryotes. The gray bars represent the standard error of the estimated divergence times at the nodes. The orange bar shows the GOE

Figure 5. Rates of evolution as a function of time. (*a*) Change in the rate of evolution of oxygenic (gray) and anoxygenic (orange) Type II reaction center proteins. The rates were obtained from the tree in Fig. 4, which was calculated using a root prior of 3.5 ± 0.05 Ga. The dashed lines represent a fit of a single-component exponential decay and the rates are given as amino acid substitutions per site per million years. The fit components are shown in supplementary Table S4. (*b*) Rate of evolution of D1 and D2 nodes with changed root priors, the red curve farthest to the right was calculated using a root prior of 4.2 ± 0.05 Ga, while the green curve farthest to the left was calculated using a root prior of 0.8 ± 0.05 Ga. (*c*) Change in the rate of evolution as a function of **ΔT**, the dashed lines represents a fit to a power-law function. The fit components are shown in supplementary Table S5

Figure 6. Effect of evolutionary models on estimated divergence times. (*a*) Effect of equilibrium frequencies: the *y* axis represent estimated divergence times calculated with a CAT + GTR model while those in the *x* axis with a CAT model with uniform equilibrium frequencies (Poisson), see also supplementary Fig. S1 for additional model comparisons. (*b*) Effect of Calibration 1 and 2 on the estimated divergence time: the estimated divergence times in the *y* axis were calculated using a calibration point 11 with a minimum age of 2.45 Ga, while those in the *x* axis using a minimum age of 2.70 Ga (see Materials and Methods), both trees were ran under a CAT model with uniform equilibrium frequencies. See also supplementary Fig. S2 for additional model comparisons. (*c*) Effect of the model of evolutionary rate variation: the estimated divergence times in the *y* axis were

calculated using an uncorrelated gamma model, while those in the *x* axis using an autocorrelated log normal model, both trees were ran under a CAT model with uniform equilibrium frequencies. See also supplementary Fig. S3 for additional model comparisons. In each case the root prior was 3.5 ± 0.05 Ga

Tables

Table 1. Effect on **ΔT** assuming different ages for the most ancestral Type II reaction center protein

Root prior	II	D0	Ancestral standard D1	ΔT
		CAT + GTR + Γ		
3.2	3.25 ± 0.05	2.80 ± 0.16	1.99 ± 0.19	0.80
3.5	3.54 ± 0.05	3.22 ± 0.19	2.19 ± 0.24	1.02
3.8	3.83 ± 0.05	3.44 ± 0.21	2.27 ± 0.24	1.17
4.1	4.12 ± 0.05	3.71 ± 0.23	2.38 ± 0.25	1.32
		LG + Γ		
3.2	3.27 ± 0.05	3.19 ± 0.08	2.51 ± 0.13	0.68
3.5	3.53 ± 0.05	3.40 ± 0.09	2.64 ± 0.15	0.77
3.8	3.81 ± 0.05	3.64 ± 0.12	2.77 ± 0.17	0.88
4.1	4.10 ± 0.05	3.91 ± 0.14	2.90 ± 0.19	1.01

Table 2. Comparison of rates of protein evolution

	Rate	Citations
	Amino acid substitutions per site per Ga	
D0 (ν_{max})	5.03	This work[a]
Group 4 D1 and D2 (ν_{min})	0.12	This work[a]
K	6.11	This work[a]
L and M	0.40	This work[a]
ADP-glucose pyrophosphorylase large subunit (plants)	1.2	(Georgelis et al., 2008)
PRTB Protein[b] (humans)	0.13	(Matsunami et al., 2011)
Alcohol dehydrogenase (ascidians)	0.27	(Canestro et al., 2002)
Protein-L-isoaspartyl (D-aspartyl) *O*-methyltransferase (bacteria to humans[c])	0.39	(Kagan et al., 1997)
Hepatitis C Virus	3700	(Bukh et al., 2002)
Influenza virus type A (H1)	5800	(Carrat & Flahault, 2007)

[a]Estimated using a root prior of 3.5 Ga under a autocorrelated log-normal molecular clock as described in the text and Material and Methods.

[b]Proline-rich transcript overexpressed in the brain (PRTB). The human protein shares 99% sequence identity compared to that in mice. Rodents are estimated to have diverged about 74 Ma ago (Kay & Hoekstra, 2008).

[c]The authors pointed out that the rate of evolution of this methyltransferase has remain unchanged from bacteria to humans.

Table 3. Change in **ΔT** under different evolutionary models

Model	**Root prior (Ga)**	**Calibration (Ga)**	**ΔT (Ga)**
CAT + GTR (autoc.[a])	3.5	2.45	1.02
CAT + GTR (autoc.)	3.5	2.70	0.97
CAT + GTR (autoc.)	3.8	2.45	1.19
CAT + GTR (autoc.)	3.8	2.70	1.13
CAT + Pois. (autoc.)	3.5	2.45	1.02
CAT + Pois. (autoc.)	3.5	2.70	1.00
CAT + Pois. (autoc.)	3.8	2.45	1.17
CAT + Pois. (autoc.)	3.8	2.70	1.12
CAT + Pois. (uncor.[b])	3.5	2.45	1.94
CAT + Pois. (uncor.)	3.5	2.70	1.84
CAT + Pois. (uncor.)	3.8	2.45	1.67
CAT + Pois. (uncor.)	3.8	2.70	2.03

[a]Log normal autocorrelated clock model.

[b]Uncorrelated gamma model

Table 4. Calibration points

Node	Event	Maximum age (Ma)	Minimum age (Ma)
1	*Arabidopsis-Populus* divergence	127	82
2	Angiosperms	248	124
3	Gymnosperms	366	306
4	Land plants	-	475
5	Diatoms-Coscinodiscophyceae	-	190
6	Floridiophyceae	-	600
7	Rhodophyta	-	1222
8	Nostocales	-	1600
9	Pleurocapsales	-	1700
10	First occurrence of multicellularity	-	1900
11	*G. violaceous* divergence on D2 and on Group 4 D1	-	2450 (Cal. 1) 2700 (Cal. 2)

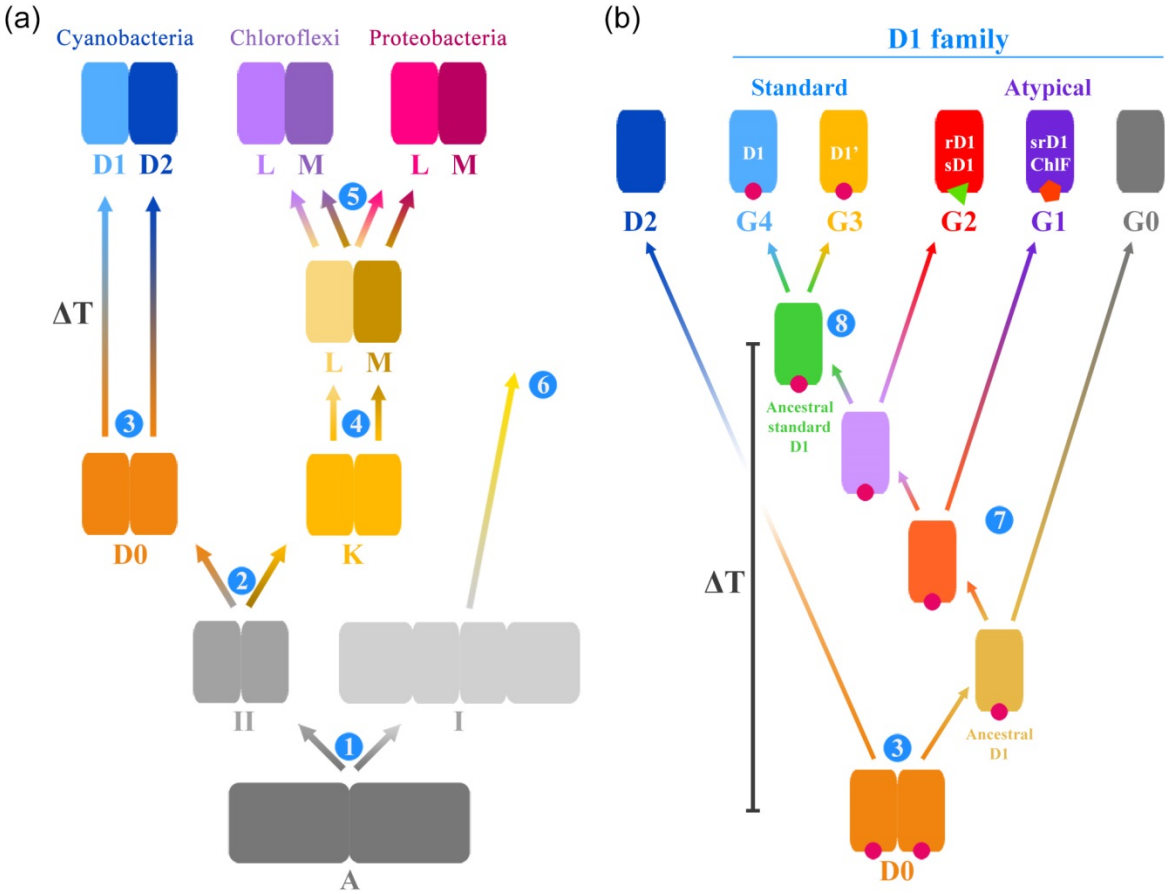
(a)
Cyanobacteria
Chloroflexi
Proteobacteria
D1 D2
L M
L M
ΔT
L M
D0
K
II
I
A
(b)
D1 family
Standard
Atypical
D1
D1'
rD1
sD1
srD1
ChlF
D2
G4
G3
G2
G1
G0
Ancestral
standard
D1
ΔT
Ancestral
D1
D0

Fig. 1

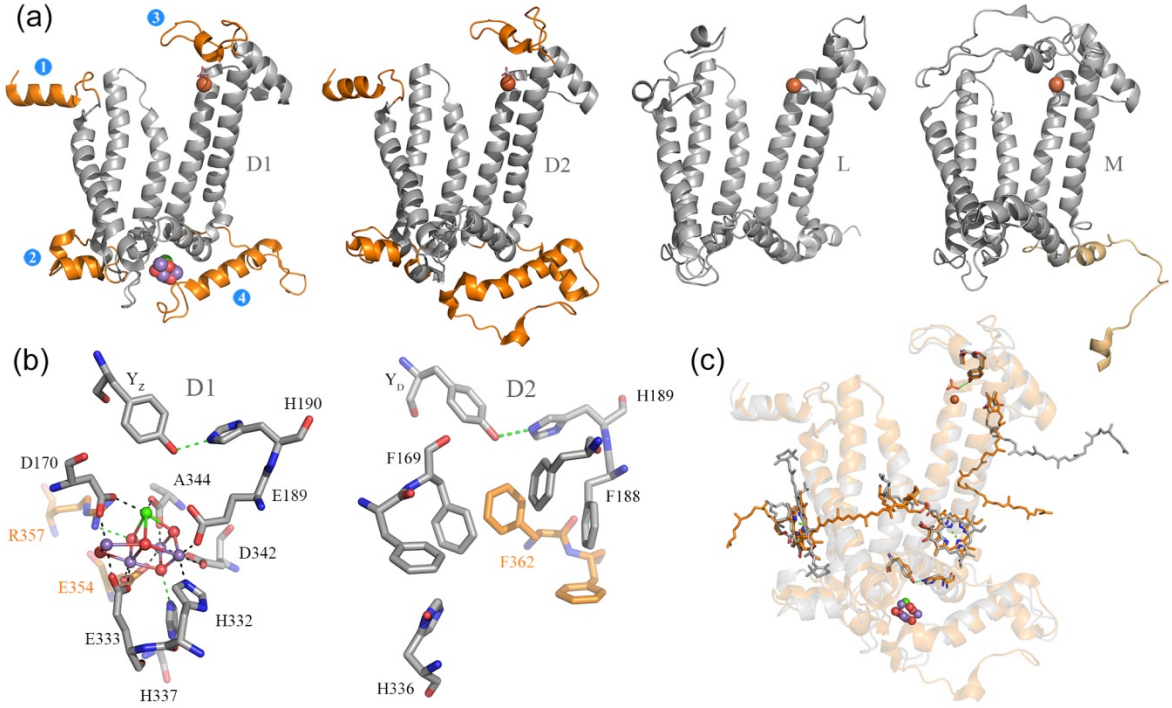
(a)
1
2
3
4
D1
D2
L
M
(b)
Y_Z
D1
H190
D170
A344
E189
R357
D342
E354
H332
E333
H337
Y_D
D2
H189
F169
F188
F362
H336
(c)

Fig. 2

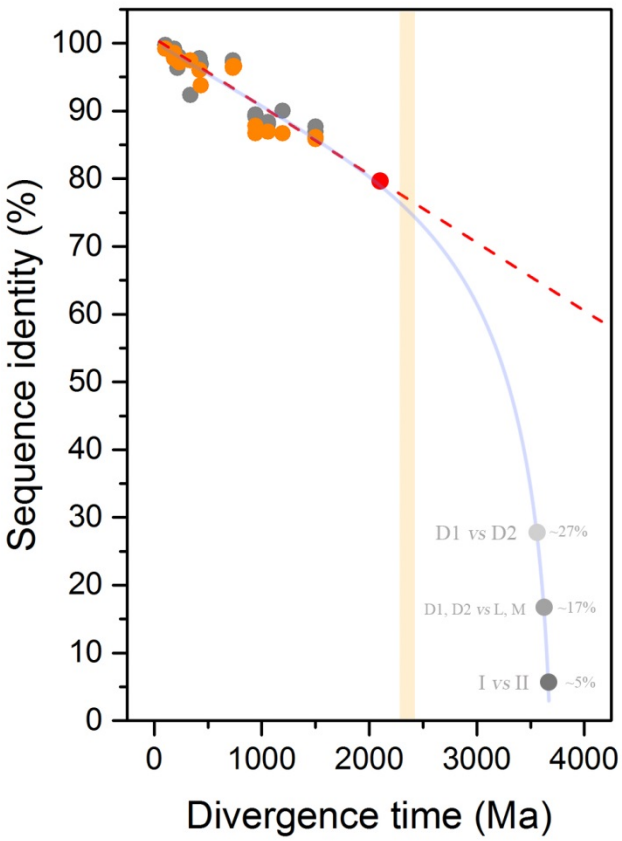
Sequence identity (%)
Divergence time (Ma)
D1 vs D2
~27%
D1, D2 vs L, M
~17%
I vs II
~5%

Fig. 3

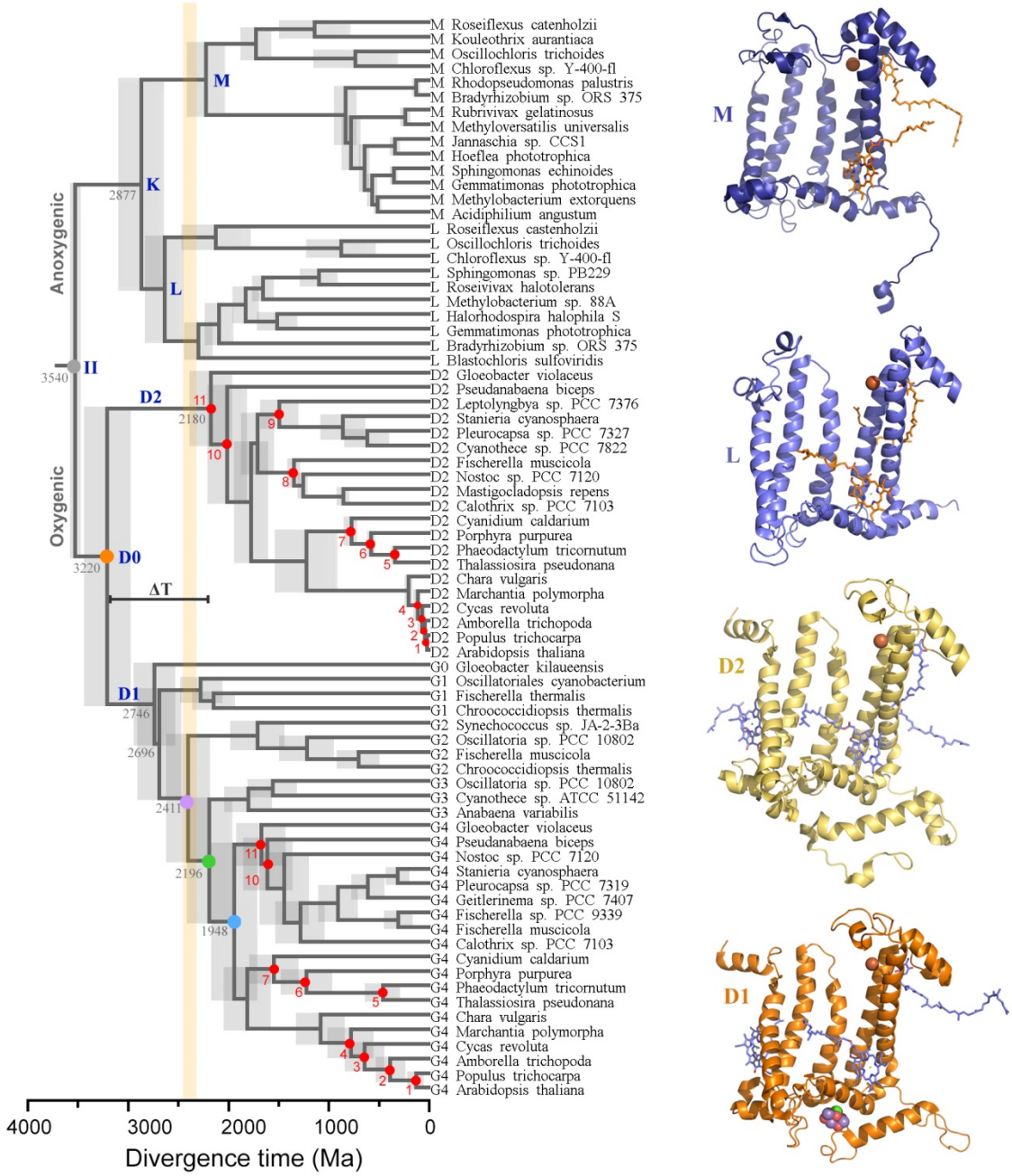
Anoxygenic
Oxygenic
M
K
L
II
D2
D0
D1
ΔT
2877
3540
2180
3220
2746
2696
2411
2196
1948
M Roseiflexus catenholzii
M Kouleothrix aurantiaca
M Oscillochloris trichoides
M Chloroflexus sp. Y-400-fl
M Rhodopseudomonas palustris
M Bradyrhizobium sp. ORS 375
M Rubrivivax gelatinosus
M Methyloversatilis universalis
M Jannaschia sp. CCS1
M Hoeflea phototrophica
M Sphingomonas echinoides
M Gemmatimonas phototrophica
M Methylobacterium extorquens
M Acidiphilium angustum
L Roseiflexus castenholzii
L Oscillochloris trichoides
L Chloroflexus sp. Y-400-fl
L Sphingomonas sp. PB229
L Roseivivax halotolerans
L Methylobacterium sp. 88A
L Halorhodospira halophila S
L Gemmatimonas phototrophica
L Bradyrhizobium sp. ORS 375
L Blastochloris sulfoviridis
D2 Gloeobacter violaceus
D2 Pseudanabaena biceps
D2 Leptolyngbya sp. PCC 7376
D2 Stanieria cyanosphaera
D2 Pleurocapsa sp. PCC 7327
D2 Cyanothece sp. PCC 7822
D2 Fischerella muscicola
D2 Nostoc sp. PCC 7120
D2 Mastigocladopsis repens
D2 Calothrix sp. PCC 7103
D2 Cyanidium caldarium
D2 Porphyra purpurea
D2 Phaeodactylum tricornutum
D2 Thalassiosira pseudonana
D2 Chara vulgaris
D2 Marchantia polymorpha
D2 Cycas revoluta
D2 Amborella trichopoda
D2 Populus trichocarpa
D2 Arabidopsis thaliana
G0 Gloeobacter kilaueensis
G1 Oscillatoriales cyanobacterium
G1 Fischerella thermalis
G1 Chroococcidiopsis thermalis
G2 Synechococcus sp. JA-2-3Ba
G2 Oscillatoria sp. PCC 10802
G2 Fischerella muscicola
G2 Chroococcidiopsis thermalis
G3 Oscillatoria sp. PCC 10802
G3 Cyanothece sp. ATCC 51142
G3 Anabaena variabilis
G4 Gloeobacter violaceus
G4 Pseudanabaena biceps
G4 Nostoc sp. PCC 7120
G4 Stanieria cyanosphaera
G4 Pleurocapsa sp. PCC 7319
G4 Geitlerinema sp. PCC 7407
G4 Fischerella sp. PCC 9339
G4 Fischerella muscicola
G4 Calothrix sp. PCC 7103
G4 Cyanidium caldarium
G4 Porphyra purpurea
G4 Phaeodactylum tricornutum
G4 Thalassiosira pseudonana
G4 Chara vulgaris
G4 Marchantia polymorpha
G4 Cycas revoluta
G4 Amborella trichopoda
G4 Populus trichocarpa
G4 Arabidopsis thaliana
4000
3000
2000
1000
0
Divergence time (Ma)
M
L
D2
D1

Fig. 4

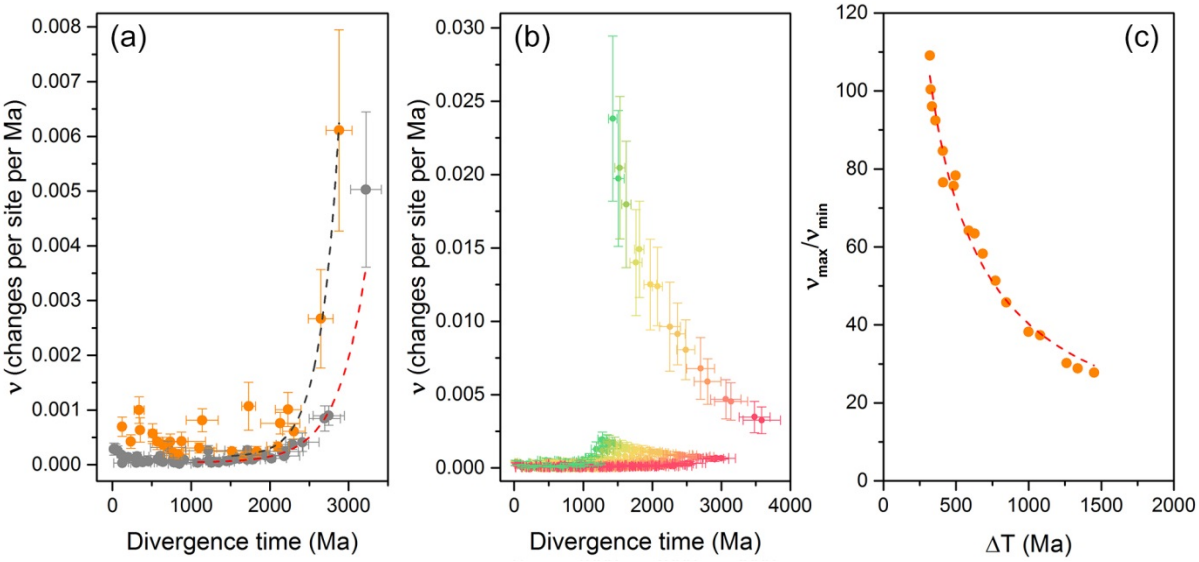
(a)
(b)
(c)
ν (changes per site per Ma)
Divergence time (Ma)
ΔT (Ma)
ν_{max}/ν_{min}

Fig. 5

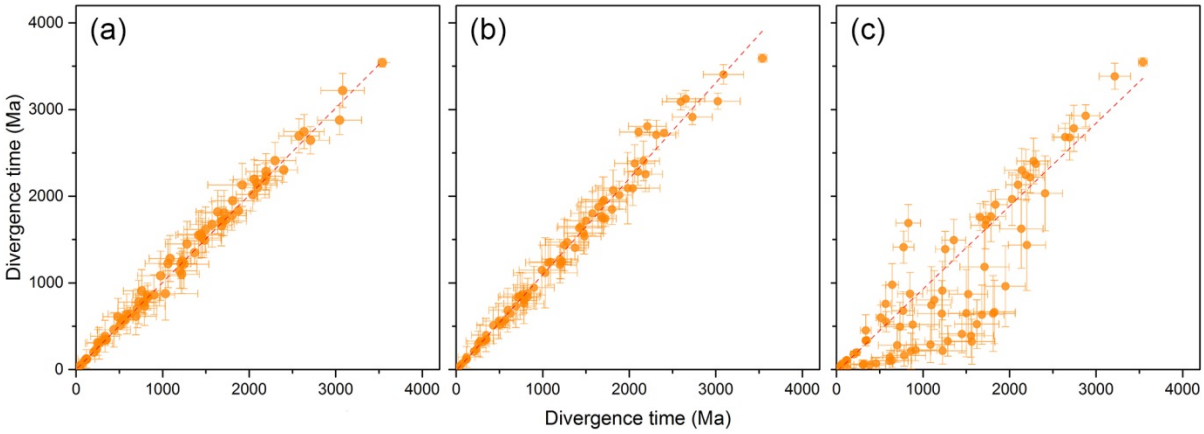
(a)
(b)
(c)
Divergence time (Ma)
Divergence time (Ma)
0
1000
2000
3000
4000

Fig. 6