bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Molecular evidence for the early evolution of photosynthetic water oxidation

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

[1]Department of Life Sciences, Imperial College London, London, UK

[2]School of Geographical Sciences, University of Bristol, Bristol, UK

[3]Plant Functional Biology and Climate Change Cluster, University of Technology Sydney, Sydney, 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 a Bayesian relaxed molecular clock analysis of the two core subunits of Photosystem II, D1 and D2, to study the origin of water oxidation. We timed the phylogeny of Type II reaction center proteins using geochemical constraints and calibrations derived from the fossil record of Cyanobacteria and photosynthetic eukaryotes. Age estimates were interpreted in the context of the structure and function of photochemical reaction centers. Firstly, we point out that the ancestral protein to D1 and D2 gave rise to a homodimeric photosystem that was capable of water oxidation, although less efficiently and by a different mechanism than the well-characterized oxygen evolving complex. 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 last common ancestor of Cyanobacteria. Thirdly, we show that Cyanobacteria did not obtain Type II reaction centers via horizontal gene transfer. Furthermore, the data suggests that the origin of photosynthesis in the early Archaean was necessarily followed by a rapid diversification of all the families of reaction center proteins, which included the divergence of D1 and D2. Our analysis suggests that primordial forms of water oxidation originated relatively soon after the emergence of photochemical reaction centers in the early Archaean.

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

The transition from anoxygenic to oxygenic photosynthesis initiated when an ancestral photochemical reaction center evolved the capacity to oxidize water to oxygen (Rutherford, 1989). Today, water oxidation is catalyzed in the Mn_4CaO_5 oxygen-evolving cluster of Photosystem II (PSII) of Cyanobacteria and photosynthetic eukaryotes. How and when Type II reaction centers diversified, and how and when one of these reaction centers evolved the capacity to oxidize water are questions that still remain to be answered. While there is agreement that by 3.5 Ga (billion years before the present) a form of anoxygenic photoautotrophy had already evolved (Butterfield, 2015; Nisbet & Fowler, 2014; Tice & Lowe, 2004), the sedimentological and isotopic evidence for the evolution of oxygenic photosynthesis ranges from 3.7 (Frei et al., 2016; Rosing & Frei, 2004) to the Great Oxidation Event (GOE) at ~2.4 Ga (Johnson et al., 2013). Similarly, molecular clock studies have generated a wider range of age estimates for the origin of Cyanobacteria between 3.5 (Falcon et al., 2010) to 2.0 Ga (Shih et al., 2017). There is thus great uncertainty and no consensus. For this reason, determining when Photosystem II evolved the capacity to oxidize water should greatly advance our understanding of the origin of oxygenic photosynthesis.

The evolution of Type II reaction center proteins has been described and discussed in some detail before (Beanland, 1990; Blankenship, 1992; Cardona, 2015, 2016; Nitschke & Rutherford, 1991; Rutherford & Nitschke, 1996; Sadekar et al., 2006) and it is presented and schematized in Fig. 1. Type II reaction centers can be divided in two major families: the *oxygenic* and the *anoxygenic* Type II reaction centers. The oxygenic Type II reaction center is also known as Photosystem II, and its electron transfer core is made of two homologous reaction center proteins, D1 and D2, exclusively found in Cyanobacteria and photosynthetic eukaryotes. The Mn_4CaO_5 cluster is located at the C-terminus of D1 (Ferreira et al., 2004). On the other hand, anoxygenic Type II reaction centers are found in phototrophic members of the phyla Proteobacteria, Chloroflexi, and Gemmatimonadetes, with the latter obtaining the reaction center via horizontal gene transfer from a gammaproteobacterium (Zeng et al., 2014). The core subunits of the anoxygenic Type II reaction centers are known as L and M and lack not only an oxygen evolving cluster but any indication that they are involved in the liberation of molecular oxygen.

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

There is no doubt that D1, D2, L, and M, share a common origin: Beanland (1990) was the first to record this but has been followed by many others (Cardona, 2015; Nitschke & Rutherford, 1991; Sadekar et al., 2006). That is to say that D1, D2, L and M, all descended from a single protein (denoted **II** in Fig. 1). The earliest event in the evolution of Type II reaction centers can be described as the divergence of this ancestral protein into two new forms, one ancestral to D1 and D2, the *oxygenic* branch; and a second one ancestral to L and M, the *anoxygenic* branch (Fig. 1). Hence, D1 and D2 originated from a gene duplication event and together make a monophyletic clade of Type II reaction center proteins, distinct from that which gave rise to L and M (Cardona, 2015, 2016b). The ancestral protein to D1 and D2 will be referred to as **D0** and the ancestral protein to L and M will be referred to as **K**.

As a result of the monophyletic relationship of D1 and D2 and the conserved structural and functional characteristics between these two proteins, it is possible to reconstruct 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 required for water oxidation and associated repair and protection processes (Cardona, 2016; 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 and 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, 2016). The small subunits are required for assembly and stability of the complex and none of them is present in anoxygenic Type II reaction centers (Dobakova et al., 2007; Komenda et al., 2012; Sugiura et al., 2015). The extrinsic polypeptides are important for the stability of

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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).

The strong inference that can be made from these shared 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 singlet oxygen and other reactive oxygen species.

Further evidence for water oxidation before the divergence of D1 and D2 is found 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 carboxylate ligands from glutamate residues at positions equivalent to 170 and 189 of standard D1, the carboxylate ligand at the C-terminus, a histidine at position equivalent to 337, and two additional ligands to the Mn_4CaO_5 cluster provided from the ancestral protein to CP43 and CP47 (Cardona, 2016; 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). A rational deduction from this is that the

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

ancestral photosystem made of **D0** oxidized water accompanied by the unavoidable production of reactive oxygen species, which then required the innovation of protective measures. From that it follows that the heterodimerization process of D1 and D2 was consequently a path to optimization of efficient water oxidation (Rutherford et al., 2012), including photoprotective mechanisms (Krieger-Liszkay et al., 2008), 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).

Several types of D1 can be distinguished phylogenetically (Cardona et al., 2015) and their evolution is schematized in Fig. 1C. The early evolving forms, referred to as atypical D1 forms (G0, G1, G2 in Fig. 1), are characterized by the absence of some of the ligands to the Mn_4CaO_5 cluster and have been recently found to be involved in the synthesis of chlorophyll *f* (M. Y. Ho et al., 2016) or the inactivation of PSII (Murray, 2012; Wegener et al., 2015). The late evolving forms, referred to as the standard D1 forms, are characterized by a complete set of ligands to the Mn_4CaO_5 cluster and are the main D1 used for water oxidation. Among the standard forms there are also several types, which have been roughly subdivided in two groups: the microaerobic forms of D1 (G3) and the dominant form of D1 (G4). The microaerobic forms are thought to be expressed only under low-oxygen conditions. The dominant form, G4, is the main D1 used for water oxidation by all Cyanobacteria and photosynthetic eukaryotes. Most Cyanobacteria carry in their genomes an array of different D1 types, yet every strain has at least one dominant form of D1 (G4). Therefore, all Cyanobacteria descended from a common ancestor that already had evolved efficient oxygenic photosynthesis, had a dominant form of D1, and was able to assemble a standard Photosystem II virtually indistinguishable from that of later evolving strains.

Based on the phylogeny of reaction center proteins 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. 1B); this is then followed by the divergence of the *anoxygenic* family (L/M) and the *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 standard form of D1. If **D0** was able to support water oxidation chemistry, then the

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

earliest stages in the evolution of oxygenic photosynthesis occurred between stages **2** and **3** as depicted in Fig. 1B. Therefore, the span of time between **D0** and the ancestral standard form of D1 (marked **8** in Fig. 1C) represents the duration of the evolutionary trajectory of Photosystem II from a simpler homodimeric water-splitting reaction center to the more complex enzyme inherited by all organisms capable of oxygenic photosynthesis. We denote this span of time by **ΔT**. If **ΔT** is small, such as a few million years or less for example, then the evolution of oxygenic photosynthesis was likely a sudden and fast process occurring shortly before the GOE as suggested by some recent analyses (Shih et al., 2017; Ward et al., 2016). On the other hand, if **ΔT** is large: several hundred million years or more for example, then water oxidation could significantly predate 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).

Here we report an in-depth evolutionary analysis of Type II reaction center proteins including Bayesian relaxed molecular clocks under various scenarios for the origin of photosynthesis. The data presented here indicate that the origin of light-driven water oxidation started early in the evolutionary history of photosynthesis, likely during the early Archaean and long before the last common ancestor of Cyanobacteria.

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*, and these are estimated to have diverged between 127.2 and 82.8 Ma (Clarke et al., 2011), see 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 fossils could push 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

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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). The trend in Fig. 3 indicates that 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, defined as the most recent common ancestor of the genus *Gloeobacter* and all other extant photosynthetic strains, 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 compared over the large time scale of the planet.

Let us reiterate that all the evidence suggests that 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 29% 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

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 29% sequence identity, this would suggest, 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 approach used above indicates that the divergence of D1 and D2 is likely placed well before the GOE, it is possible to make further inferences on the evolution of oxygenic photosynthesis by applying a Bayesian molecular clock approach to the phylogeny of Type II reaction center proteins and from the resulting rates of evolution. Figure 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). As an informed starting point we first 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. This is in agreement with the inferences made from the level of sequence identity as plotted in Fig. 3.

To test the effect of different root priors on our results we varied the age of the root and the standard deviation over a broad range. Table 1 lists estimates of divergence times of key ancestral Type II reaction center proteins and the respective **ΔT** value using different root priors. For example, under the assumption that Type II reaction centers had already evolved by 3.8 Ga ago (Czaja et al., 2013; Nisbet & Fowler, 2014; Rosing, 1999), **ΔT** is found to be 1.17 Ga. Similarly, if it is assumed to

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

be a late event centered at 3.2 Ga, though unlikely, **ΔT** is still 0.90 Ga. Furthermore, increasing the standard deviation on the root prior pushes the timing of the earliest events in the evolution of the Type II reaction center to even older ages rather than younger ages. Supplementary Table S3 lists the changes in **ΔT** by varying the standard deviation in the root prior. 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.

In Fig. 4 it can also be seen that the divergence of the L and M subunits occurs *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. These results highlight the conclusion that it is unlikely that Cyanobacteria obtained Type II reaction centers via horizontal gene transfer (HGT) of L and M from an early evolving representative of the phylum Chloroflexi or Proteobacteria (Cardona, 2015).

The Bayesian clock 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 achieved using 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

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

The inferences derived from Fig. 3 reveal that the rates of evolution had to be faster in the initial stages during the Archaean compared with the Proterozoic, even when **ΔT** is as large as 1.0 Ga. To gain 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. In Fig. 5A the rate of evolution (**ν**) of each node in the tree, expressed as amino acid substitutions per site per unit of time, is plotted 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. Clearly, faster rates are necessary to explain the origin and evolution of Type II reaction centers at any point in the Archaean and regardless of when exactly photosynthesis originated, as seen in Fig. 5B. The decrease in the rate of evolution is consistent with the observations derived from Fig. 3 and can be roughly fitted with a first-order exponential decay curve (fitting parameters are presented in 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 (stages **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 the same to occur 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 $\boldsymbol{\nu}_{\mathbf{max}}$. Fig. 5A shows that the rate of evolution flattens out to 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 $\boldsymbol{\nu}_{\mathbf{min}}$ and it is calculated as the average rate from each node in Group 4 D1 and D2. In Fig. 5A, $\boldsymbol{\nu}_{\mathbf{max}}$ is 5.03 ± 1.42 amino acid substitutions per site per Ga, while $\boldsymbol{\nu}_{\mathbf{min}}$ is 0.12 ± 0.04 substitutions per site per Ga. Based on these assumptions it is almost certain that reaction centers had

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

evolved by 3.5 Ga, to account for the divergence of D1 and D2 in one billion years (**ΔT**), since the initial rate of evolution had to be about 40 times larger than that observed since the last common ancestor of Cyanobacteria.

Table 2 lists the rates of evolution of a diverse number of proteins reported in independent studies in a broad range of organisms. It is found that the rate of evolution of the core subunits of PSII is similar to the rate of other proteins that are very ancient and highly conserved such as membrane proteins involved in bioenergetics or the ribosome. Our estimated rates fall well within the expected range of other cyanobacterial proteins, thus validating our calibration choices and consistent with the expected level of sequence identity as derived from Fig. 3.

Our Bayesian analyses suggests that the evolution and diversification of PSII is better explained by a long span of time since the gene duplication even of the homodimeric core until the emergence of standard PSII inherited by crown group Cyanobacteria. Notwithstanding, a fast rate of evolution at the earliest stage implies that $\mathbf{v_{max}}$ would increase if **ΔT** is considered to be smaller, as would be the case for an evolutionary scenario in which oxygenic photosynthesis evolves rapidly before the GOE after events of horizontal gene transfer. We illustrate this concept in Fig. 5B and 5C. These figures depict the effect on the rate of evolution as a function of **ΔT**. This manipulation of the molecular clock can only by accomplished computationally by changing the root prior to younger and younger ages. The increase of $\mathbf{v_{max}}$ with decreasing **ΔT** would obey a power law-function. For example, in the hypothetical case that **ΔT** is 1 million years; that is, if the divergence of D1 and D2 occurred in only 1 million years, then $\mathbf{v_{max}}$ would have to have been about 12500 amino acid changes per site per Ga, which is greater than the rate of evolution of some extant viruses, see Table 2.

Sensitivity analysis

To test the reliability 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: Supplementary Fig. S1 provides a comparison of estimated divergence times calculated with the CAT + GTR model (Z. Yang & Rannala, 2006) against divergence times calculated using the CAT model with a uniform (Poisson) model of equilibrium frequencies. The GTR model does not have a strong effect on the

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

calculated divergence times as the slope of the graph does not deviate from unity when paired with the uniform model (see supplementary Table S6 for linear regressions). 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. Supplementary Fig. S2 provides a comparison of the two calibration choices on the overall estimated times for flexible and non-flexible models. If the divergence times using both calibrations are plotted against each other, a linear relationship is obtained (supplementary 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 as shown in supplementary Fig. S3 and Table S8. Supplementary Fig. S3 presents a comparison of divergence time estimates of a tree calculated using a relaxed log-normal autocorrelated molecular clock with a tree calculated using an uncorrelated gamma model on the rates of evolution. The autocorrelated model assumes that neighboring 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. The molecular mechanism behind this difference could be related to the fact that photochemistry imposes a strong constraint 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 reactive oxygen species (Cardona et al.,

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2012; Rutherford et al., 2012). It seems reasonable then, that the rates of evolution of all Type II reaction center proteins should be correlated between closely related groups, thus corresponding to an autocorrelated model.

Discussion

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). Two main conclusions can be derived from this plot, which are also in agreement with the molecular clock analysis. Firstly, the three earliest stages in the evolution of photochemical reaction centers: the divergence of Type I and Type II reaction centers, the divergence of *anoxygenic* and *oxygenic* families of Type II reaction center proteins, and the divergence of D1 and D2, are 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 photosynthesis at 3.5 Ga (Butterfield, 2015; Nisbet & Fowler, 2014; Tice & Lowe, 2004), then all three stages could well predate this time. Secondly, the last common ancestor of crown group Cyanobacteria is more likely to have lived near the time of the GOE rather than shortly after the origin of photosynthesis in the early Archaean. This common ancestor must have had a standard PSII, which places it relatively far after the origin of photosynthesis. The early divergence of D1 and D2 means that the earliest stages in the evolution of photosynthetic water oxidation predates the last common ancestor of Cyanobacteria by about 1.0 Ga.

Bayesian relaxed molecular clock analysis

The application of a Bayesian molecular clock analysis to the phylogeny of Type II reaction centers is strongly constrained by three pieces of well-supported evidence: 1) evidence of photosynthesis at 3.5 Ga, 2) by strong evidence for the GOE at 2.4 Ga, and 3) by the relatively slow rate of evolution of the core proteins over the last 2.0 Ga. Under these constraints, the divergence between D1 and D2 is better explained by water oxidation starting early in the evolutionary history of photosynthesis, in the early Archaean, with the appearance of standard PSII occurring after a long period of evolutionary

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

innovation. We highlight this long period by introducing the concept of **ΔT**, which defines two specific points in time in the evolution of PSII: the time from the gene duplication event of the homodimeric core to the ancestral standard heterodimeric PSII retained by extant Cyanobacteria. The magnitude of **ΔT** is dictated by the large phylogenetic distance between D1 and D2 and the relatively slow rate of evolution determined from the geochemical and fossil record. Therefore, it is not surprising that under most models employed in this analysis **ΔT** is always in the range of 0.7 to 1.2 Ga.

We have considered two possible evolutionary scenarios that are both consistent with a large **ΔT** (Fig. 6). 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, assuming that the earliest events in the history of photosynthesis started about 3.5 Ga, the divergence of D1 and D2 is set at about 3.2 Ga.

In the second scenario, we considered 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 the present work, this scenario can be fitted most satisfactorily with the application of a relaxed molecular clock using an empirical amino acid substitution model (LG + Γ). 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; and this has the consequence of pushing the divergence of D1 and D2 closer to the root, and thus **D0** is set at about 3.4 Ga (Table 1, Fig. 6). This effect is due in part to the fact that the phylogenetic distance between D1 and D2 is invariable, and thus under any scenario the data is better explained by a long span of time separating **D0** and the standard heterodimeric PSII. 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.

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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. This scenario requires that the transferred Type II reaction center was still homodimeric occurring before the divergence of D1 and D2. The phylogeny of Type II reaction centers and our Bayesian analysis do not lend support to this scenario because the phylogenetic distance between D1 and D2 is too great and the rates of evolution of the core subunits are too slow for this transfer to have occurred not long before the GOE. Our analysis strengthens the notion that early-branching non-photosynthetic Cyanobacteria emerged after losing water oxidation catalysis in a process of adaptation to heterotrophic lifestyles. The loss of oxygenic photosynthesis is a relatively common process, several well-known examples are: the secondary loss of all PSII genes 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 parasitic Apicomplexa (Moore et al., 2008) and holoparasitic angiosperms (Ravin et al., 2016).

Rates of evolution

It should be observed that if a relatively constant rate of evolution 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 constrained 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 origin of Type II reaction centers at 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.

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

We are not aware of any systematic study of the change in the rate of evolution of proteins throughout geological time and such a study is outside the scope of this article. However, there are several families of proteins that are likely to follow a similar trend. For example, the phylogeny of rubisco is characterized by limited sequence divergence at the phylum level, but a much greater phylogenetic distance between orthologous groups (Kacar et al., 2017). Indeed, Kacar et al. (2017) suggested that the divergence of Group 3 rubisco of methanogenic Archaea and bacterial photosynthetic Group 1 rubisco occurred as the latter acquired adaptations to deal with oxygen and more oxidizing conditions. The divergence of Group 3 and Group 1 Rubisco also requires initially higher rates of evolution. A similar situation is expected for methanogenic and photosynthetic phosphoribulokinases, which also show a remarkable phylogenetic distance (Kono et al., 2017). Another similar case is found between the homologous enzymes Ni-sirohydrochlorin a,c-diamide reductive cyclase of methanogeneic Archaea, protochlorophyllide and clorophyllide reductase of phototrophic bacteria (Hu & Ribbe, 2015; Zheng et al., 2016), and nitrogenases. These enzymes show a very high level of sequence identity within phyla due to slow rates of evolution, but a great phylogenetic distance between orthologous groups.

The phenomenon of an initial fast rate of evolution followed by an exponential decrease demands an explanation. We can think of two mechanisms that may account for this observation, but there may be more. 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

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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.

Heterodimerization of ancestral Type II reaction centers

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

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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, which 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 bacteriochlorophyll and chlorophyll triplet-induced formation of radical oxygen species early during the evolutionary history of photosynthesis.

Concluding remarks

The evolution of Type II reaction centers highlights the long history of photosynthetic water oxidation. We showed that the span of time between the gene duplication event that led to D1 and D2 and the appearance of standard PSII could have been about a billion years, placing water oxidation in the early Archaean. So what happened during this long period of time? To begin with, water oxidation originated in a simpler, single-subunit, homodimeric photosystem in a completely anaerobic world. Therefore, the large increase in the structural complexity of PSII, PSI, and associated light harvesting complexes had to occur along this trajectory. This includes the acquisition of many peripheral and extrinsic protein subunits and the heterodimerization of D1 and D2, CP43 and CP47, and PsaA and PsaB. At the same time, the thermodynamic coupling between both photosystems and the retuning of the entire electron transport chain and all electron carriers to increasingly oxidizing conditions also had to occur. This expansion in complexity had to be coupled with the evolution of highly organized assembly and repair processes. Thus, the first water-oxidizing reaction centers may have been active

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

only for briefs amounts of time in the absence of efficient repair. Greater water oxidation efficiency also needed the innovation of photoprotective mechanisms acting at different time scales spanning several orders of magnitude, like dissipatory recombination pathways, non-photochemical quenching, state-transitions, or chromatic adaptation. Furthermore, the light reactions of photosynthesis had to be linked to carbon fixation and other downstream metabolic process. Signaling, feedback, and regulatory mechanisms had to be put in place to control photosynthesis under varying environmental conditions and across a diel cycle. Needless to say, all anaerobic reactions and processes inhibited by oxygen originally found in the earliest anaerobic water-oxidizing ancestors had to be separated from oxygen production or readapted to work under aerobic conditions. The link to carbon fixation is of particular importance. CO_2 levels in the atmosphere were higher than now (Nutman et al., 2017; Sheldon, 2006). However, limiting diffusion across boundary layers in mats and stromatolites would have restricted anoxygenic and early oxygenic phototrophs alike. The development of water oxidation would have opened up the way to faster photosynthetic rates, spurring on gross primary production rates later in the Archean, with the concomitant need for increases in nitrogen fixation.

Materials and Methods

Phylogeny of Type II reaction centers

Sequences were retrieved from the RefSeq NCBI database using PSI-BLAST restricted to Cyanobacteria, Proteobacteria, and Gemmatimonadetes. A total of 1703 complete sequences were downloaded and aligned 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 (ML) phylogenetic analysis was performed using PhyML 3.1 (Guindon et al., 2010) using the LG model of amino acid substitution. The amino acid equilibrium frequencies, proportion of invariable sites, and across site rate variation were allowed to be

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 ML tree using all sequences is shown in Fig. 1 and it replicates earlier evolutionary studies of Type II reaction centers that used simpler methods and fewer sequences (Beanland, 1990; Cardona, 2015).

Change in sequence identity as a function of time

To get a better understanding of the evolutionary trends of D1 and D2 reaction center proteins as a function of time, we compared the percentage of sequence identity of D1 and D2 proteins from species of photosynthetic eukaryotes with known or approximate divergence times. In total 23 pairs of sequences were compared and are listed in Supplementary Table 1. When two sequences were of different length, the longest was taken as 100%. Of these 23 pairs, the first 16 were based on the fossil calibrations recommended by Clarke et al. (2011) after their extensive review of the plant fossil record. Divergence times were taken as the average of the hard minimum and soft maximum fossil ages suggested by the authors. The last 7 comparisons are approximate dates taken from a recent molecular clock analysis of the evolution of red algae (E. C. Yang et al., 2016). The plot of sequence identity *vs* approximate divergence time was then fitted with a linear regression and the fitting parameters are shown in supplementary Table S2.

Bayesian relaxed molecular clock and fossil calibrations

A total of 54 bacterial sequences spanning the entire diversity of Type II reaction centers were selected for Bayesian molecular clock analysis, including atypical and standard forms of D1, Alpha-, Beta-, Gammaproteobacteria, Chloroflexales, and *Gemmatimonas phototrophica*. Furthermore, 10 D1 and 10 D2 sequences from photosynthetic eukaryotes from taxa with a well characterized fossil record were added to allocate calibration points. The phylogeny of Type II reaction centers was cross-calibrated on D1 and D2 as listed in Table 4 and calibration points were assigned as presented in Fig. 4, red dots.

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Dates for the *Arabidopsis*/*Populus* divergence, the divergence of the angiosperms (*Amborella*), gymnosperms (*Cycas*), and land plants (*Marchantia*) were implemented as suggested and discussed by Clarke et al. (2011), representing points 1 to 4 in Fig. 4. Three ages from eukaryotic algae were used too: an age of 190 Ma was assigned to the divergence of *Phaeodactylum trichornutum* and *Thalassiosira pseudonana*, based on fossil Jurassic diatoms from the Lias Group, reviewed by Sims et al. (2006). A minimum age of 600 Ma based on a Late Neoproterozoic Chinese multicellular red alga fossil (Xiao et al., 2004) was assigned to the split between the diatom sequences and the sequences from *Porphyra purpurea*, as a conservative estimate for the divergence of complex red algae, which predates this time (E. C. Yang et al., 2016). The earliest widely accepted 1.2 Ga red-algae fossil, *Bangiomorpha* (Butterfield, 2000; Knoll et al., 2013), was assigned as a minimum age to the divergence of the sequences from *Cyanidium caldarium*, a unicellular early-branching red algae, this was used as a conservative estimate for the most recent common ancestor of red algae. It is however quite likely that the last common ancestor of red algae significantly predates this age as suggested by recently described 1.6 Ga red algae fossils (Bengtson et al., 2017) and recent molecular clock analysis (Sánchez-Baracaldo et al., 2017; E. C. Yang et al., 2016).

Previously implemented Cyanobacterial fossils were also used to calibrate the clock (Blank & Sanchez-Baracaldo, 2010; Sanchez-Baracaldo, 2015; Sánchez-Baracaldo et al., 2017). A minimum age of 1.6 Ga was assigned to Nostocales because described akinetes of this age have been found in cherts from Siberia, China, and Australia (Golubic et al., 1995; Schirrmeister et al., 2016; Tomitani et al., 2006). Pleurocapsales are characterized by multiple fissions during cell division and fossils retaining this morphology have been described at 1.7 Ga (Golubic & Lee, 1999; Sergeev et al., 2002). The earliest well-assigned filamentous Cyanobacteria fossils of comparable size to those of the early-branching *Pseudanabaena* have been reported at 1.9 Ga (Golubic & Lee, 1999; Sánchez-Baracaldo et al., 2017; Schirrmeister et al., 2016; Sergeev et al., 2002), and this was assigned as a minimum age to the sequences from *Pseudanabaena biceps.*

The oldest calibration point, point 11, was selected to be the branching point of the D2 and the G4 D1 from *Gloeobacter violaceous*. This was set to be around the age for the GOE and was assigned as a minimum age of 2.45 Ga (Calibration 1) (Bekker et al., 2004). For comparison, a

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

calibration of 2.7 Ga was also used (Calibration 2) to test the effect on the estimated divergence times of an older age for crown group Cyanobacteria. Geological evidence suggests that oxygen 'whiffs' or 'oases' could significantly predating the GOE (Lyons et al., 2014) so this scenario is not entirely implausible.

The calibration points on D1 were allocated on Group 4 because this type of D1 is the only one retained in all Cyanobacteria with PSII, it is the only type of D1 inherited by photosynthetic eukaryotes, and it is the main D1 used to catalyze water oxidation under most growth conditions. 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.

It is well accepted that a form of anoxygenic photosynthesis had already evolved 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). Therefore, we used a root prior of 3.5 Ga as a conservative estimate for photoautrophy based on Mg-tetrapyrrole containing photochemical reaction centers. Nevertheless, because it is not yet known exactly when photosynthesis originated for the first time we also tested the effect of varying the root prior from 3.2 to 4.1 Ga on the estimated divergence time under restrictive and flexible scenarios.

A Bayesian Markov chain Monte Carlo approach was used to calculate the estimated divergence times. We used 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 (Z. Yang & Rannala, 2006). To understand the effect of different evolutionary models on the age estimates we compared the CAT + GTR + Γ model with 1) a LG + Γ model that sets less flexible boundaries on the calibration points, 2) a CAT + Γ model assuming a uniform (Poisson) distribution of amino acid equilibrium frequencies, or 3) an uncorrelated gamma model where the rates of

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

substitution can vary independently. The flexible bounds on the CAT + GTR + Γ model 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, which are necessary for the computation of divergence time from the phylogeny, were retrieved from the output files of Phylobayes. These rates are calculated by the software as described by the developers 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 sequence datasets and 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

Beanland, T. J. (1990). Evolutionary relationships between q-type photosynthetic reaction centers - hypothesis-testing using parsimony. *J Theor Biol*, *145*(4), 535-545. doi:Doi 10.1016/S0022-5193(05)80487-4

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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:DOI 10.1111/j.1472-4669.2009.00220.x

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

Bosak, T., Liang, B., Sim, M. S., & Petroff, A. P. (2009). Morphological record of oxygenic photosynthesis in conical stromatolites. *Proc Natl Acad Sci U S A, 106*(27), 10939-10943. doi:10.1073/pnas.0900885106

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. *Proc Natl Acad Sci U S A, 113*(43), 12144-12149. doi:10.1073/pnas.1608862113

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. *Proc Natl Acad Sci U S A, 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: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:DOI 10.1007/s00239-001-0020-2

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

Cardona, T. (2016). 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., Murray, J. W., & Rutherford, A. W. (2015). Origin and evolution of water oxidation before the last common ancestor of the cyanobacteria. *Mol Biol Evol, 32*(5), 1310-1328. doi: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. *Biochim Biophys Acta, 1817*(1), 26-43. doi:DOI 10.1016/j.bbabio.2011.07.012

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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:DOI 10.1111/j.1469-8137.2011.03794.x

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

Cogdell, R. J., Howard, T. D., Bittl, R., Schlodder, E., Geisenheimer, I., & Lubitz, W. (2000). How carotenoids protect bacterial photosynthesis. *Philos Trans R Soc Lond B Biol Sci, 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. *Nat Commun, 7*. doi:ARTN 11071
10.1038/ncomms11071

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 Planet Sci Lett, 363*, 192-203. doi:10.1016/j.epsl.2012.12.025

David, L. A., & Alm, E. J. (2011). Rapid evolutionary innovation during an archaean genetic expansion. *Nature, 469*(7328), 93-96. doi: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 Plant Sci, 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. *Biochim Biophys Acta, 1503*(1-2), 164-186. doi: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.

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 Physiol, 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. *Biochem Biophys Res Commun, 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. *Proc Natl Acad Sci U S A, 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:DOI 10.1126/science.1093087

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Franzen, L. G., Hansson, O., & Andreasson, L. E. (1985). The roles of the extrinsic subunits in photosystem ii as revealed by electron paramagnetic resonance. *Biochim Biophys Acta, 808*(1), 171-179. doi: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. *Sci Rep, 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 Evol Biol, 8*. doi:Artn 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: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: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). *Geol Soc Am Mem, 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. *Syst Biol, 59*(3), 307-321. doi:10.1093/sysbio/syq010

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

Hu, Y. L., & Ribbe, M. W. (2015). Nitrogenase and homologs. *Journal of Biological Inorganic Chemistry, 20*(2), 435-445. doi:10.1007/s00775-014-1225-3

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. *J Comp Biol, 11*(5), 787-799. doi:Doi 10.1089/1066527042432260

Johnson, J. E., Webb, S. M., Thomas, K., Ono, S., Kirschvink, J. L., & Fischer, W. W. (2013). Manganese-oxidizing photosynthesis before the rise of cyanobacteria. *Proc Natl Acad Sci U S A, 110*(28), 11238-11243. doi:DOI 10.1073/pnas.1305530110

Kacar, B., Hanson-Smith, V., Adam, Z. R., & Boekelheide, N. (2017). Constraining the timing of the great oxidation event within the rubisco phylogenetic tree. *Geobiology*, n/a-n/a. doi:10.1111/gbi.12243

Kagan, R. M., McFadden, H. J., McFadden, P. N., OConnor, C., & Clarke, S. (1997). Molecular phylogenetics of a protein repair methyltransferase. *Comp Biochem Physiol B Biochem Mol Biol 117*(3), 379-385. doi:Doi 10.1016/S0305-0491(96)00333-1

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Kay, E. H., & Hoekstra, H. E. (2008). Rodents. *Current Biology, 18*(10), R406-R410. doi: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. *Mol Biol Evol, 18*(3), 352-361.

Knoll, A. H. (2015). Paleobiological perspectives on early microbial evolution. *Cold Spring Harb Perspect Biol, 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

Kono, T., Mehrotra, S., Endo, C., Kizu, N., Matusda, M., Kimura, H., . . . Ashida, H. (2017). A rubisco-mediated carbon metabolic pathway in methanogenic archaea. *Nat Commun, 8*, 14007. doi:10.1038/ncomms14007

Krieger-Liszkay, A., Fufezan, C., & Trebst, A. (2008). Singlet oxygen production in photosystem ii and related protection mechanism. *Photosynth Res, 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. *Nat Plants, 1*, 15168. doi:10.1038/nplants.2015.168

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: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. *Mol Biol Evol, 21*(6), 1095-1109. doi:DOI 10.1093/molbev/msh112

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

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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. *Proc Natl Acad Sci U S A, 113*(29), 8194-8199. doi:10.1073/pnas.1607580113

Linke, K., & Ho, F. M. (2013). Water in photosystem ii: Structural, functional and mechanistic considerations. *Biochim Biophys 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

Marin, J., Battistuzzi, F. U., Brown, A. C., & Hedges, S. B. (2017). The timetree of prokaryotes: New insights into their evolution and speciation. *Mol Biol Evol, 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:Doi 10.1016/S1386-1425(00)00289-4

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

Murray, J. W. (2012). Sequence variation at the oxygen-evolving centre of photosystem ii: A new class of 'rogue' cyanobacterial d1 proteins. *Photosynth Res, 110*(3), 177-184. doi:10.1007/s11120-011-9714-5

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. *Proc Natl Acad Sci U S A, 111*(31), 11407-11412. doi:10.1073/pnas.1405222111

Nickelsen, J., & Rengstl, B. (2013). Photosystem ii assembly: From cyanobacteria to plants. *Annu Rev Plant Biol, 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: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., Michoux, F., Yu, J. F., Boehm, M., & Komenda, J. (2010). Recent advances in understanding the assembly and repair of photosystem ii. *Ann Bot, 106*(1), 1-16. doi:Doi 10.1093/Aob/Mcq059

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Nutman, A. P., Bennett, V. C., & Friend, C. R. L. (2017). Seeing through the magnetite: Reassessing eoarchean atmosphere composition from isua (greenland) ≥3.7 ga banded iron formations. *Geoscience Frontiers*. doi:10.1016/j.gsf.2017.02.008

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 Geosci, 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 Biol, 16*(Suppl 3), 238. doi:10.1186/s12870-016-0929-7

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

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: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 Planet Sci Lett 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:Doi 10.1016/0968-0004(89)90032-7

Rutherford, A. W., & Faller, P. (2003). Photosystem ii: Evolutionary perspectives. *Philos Trans R Soc Lond B Biol Sci, 358*(1429), 245-253. doi: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. *Biochim Biophys 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. *Mol Biol Evol, 23*(11), 2001-2007. doi:DOI 10.1093/molbev/msl079

Sanchez-Baracaldo, P. (2015). Origin of marine planktonic cyanobacteria. *Sci Rep, 5*, 17418. doi:10.1038/srep17418

Sánchez-Baracaldo, P., Raven, J. A., Pisani, D., & Knoll, A. H. (2017). Early photosynthetic eukaryotes inhabited low-salinity habitats. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1620089114

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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.

Schidlowski, M. (1988). A 3,800-million-year isotopic record of life from carbon in sedimentary-rocks. *Nature, 333*(6171), 313-318. doi: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. *Proc Natl Acad Sci U S A, 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. *Int J Astrobiol 15*, 187-204. doi:10.1017/S1473550415000579.

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

Sheldon, N. D. (2006). Precambrian paleosols and atmospheric co2 levels. *Precambrian Research, 147*(1-2), 148-155. doi:10.1016/j.precamres.2006.02.004

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

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. *Mol Syst Biol, 7*. doi:ARTN 539
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 Biol Evol, 6*(5), 1031-1045. doi:Doi 10.1093/Gbe/Evu073

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. *Biochim Biophys Acta, 1847*(2), 276-285. doi:10.1016/j.bbabio.2014.11.009

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

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Tomitani, A., Knoll, A. H., Cavanaugh, C. M., & Ohno, T. (2006). The evolutionary diversification of cyanobacteria: Molecular-phylogenetic and paleontological perspectives. *Proc Natl Acad Sci U S A, 103*(14), 5442-5447. doi:10.1073/pnas.0600999103

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. *J Biol Chem, 287*(8), 5720-5732. doi: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 Plant Sci, 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. *Orig Life Evol Biosph, 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. *J Biol Chem, 290*(6), 3764-3774. doi:10.1074/jbc.M114.604124

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. *Am J Bot, 91*(2), 214-227. doi:10.3732/ajb.91.2.214

Yang, E. C., Boo, S. M., Bhattacharya, D., Saunders, G. W., Knoll, A. H., Fredericq, S., . . . Yoon, H. S. (2016). Divergence time estimates and the evolution of major lineages in the florideophyte red algae. *Sci Rep, 6*, 21361. doi:10.1038/srep21361

Yang, Z., & Rannala, B. (2006). Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol Biol Evol, 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. *Proc Natl Acad Sci U S A, 111*(21), 7795-7800. doi:DOI 10.1073/pnas.1400295111

Zheng, K., Ngo, P. D., Owens, V. L., Yang, X. P., & Mansoorabadi, S. O. (2016). The biosynthetic pathway of coenzyme f430 in methanogenic and methanotrophic archaea. *Science, 354*(6310), 339-342. doi:10.1126/science.aag2947

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 2. Comparison of rates of protein evolution

	Rate (Amino acid substitutions per site per Ga)	**References**
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]
PsaA, Photosystem I core subunit (Cyanobacteria)	0.09	(Sanchez-Baracaldo, 2015)
AtpA, ATP Synthase CF1 Alpha Chain (Cyanobacteria)	0.08	(Sanchez-Baracaldo 2015
PetB, Cytochrome *b*6 (Cyanobacteria)	0.05	(Sanchez-Baracaldo 2015
S13 Ribosomal Protein (Cyanobacteria)	0.13	(Sanchez-Baracaldo 2015
L1 Ribosomal Protein (Cyanobacteria)	0.11	(Sanchez-Baracaldo 2015
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.

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 4. Calibration points

Node	Event	Maximum (Ma)	Minimum (Ma)
1	*Arabidopsis-Populus* divergence	127	82
2	Angiosperms	248	124
3	Gymnosperms	366	306
4	Land plants	-	475
5	Diatoms	-	190
6	Floridae	-	600
7	Last common ancestor of red algae	-	1200
8	Heterocystous Cyanobacteria	-	1600
9	Pleurocapsales	-	1700
10	Early-branching multicellular Cyanobacteria	-	1900
11	Last common ancestor of Cyanobacteria	-	2450/2700

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 1. Evolution of Type II reaction center proteins. (a) A Maximum Likelihood phylogeny of Type II reaction center proteins. (b) An schematic representation of the phylogeny shown in (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. (c) 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. The lack of fuchsia dot in the G0 sequence indicates the absence of some of the ligands to the Mn_4CaO_5 cluster. This sequence has acquired changes that are predicated to impair function and it is not known whether it is assembled into PSII. The pentagon in G1 indicates the possible binding of chlorophyll *a* as hypothesized by M. Y. 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

bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 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 there is no constant rate of evolutionary change 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 timing for the earliest events in the history of photosynthesis: the divergence of D1 and D2 (~29% 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

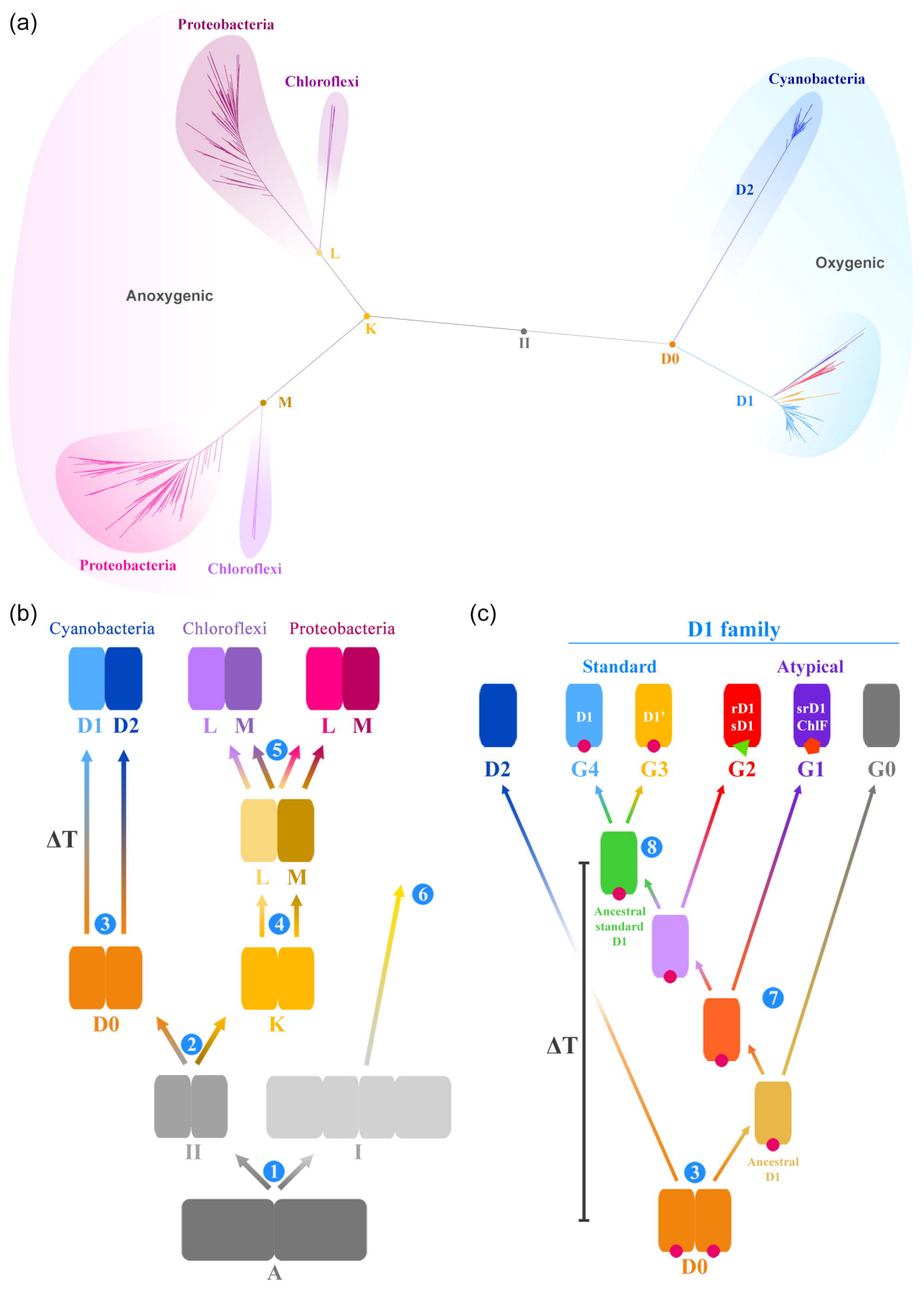
Figure 4. Relaxed molecular clock of Type II reaction center proteins. A log-normal autocorrelated relaxed clock is shown implementing a CAT + GTR + Γ non-parametric model with flexible boundaries on the calibration points. Red dots are calibration points as described in Materials and Methods. The gray dot denoted **II**, represents the ancestral Type II reaction center protein, as schematized in Fig. 1. 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

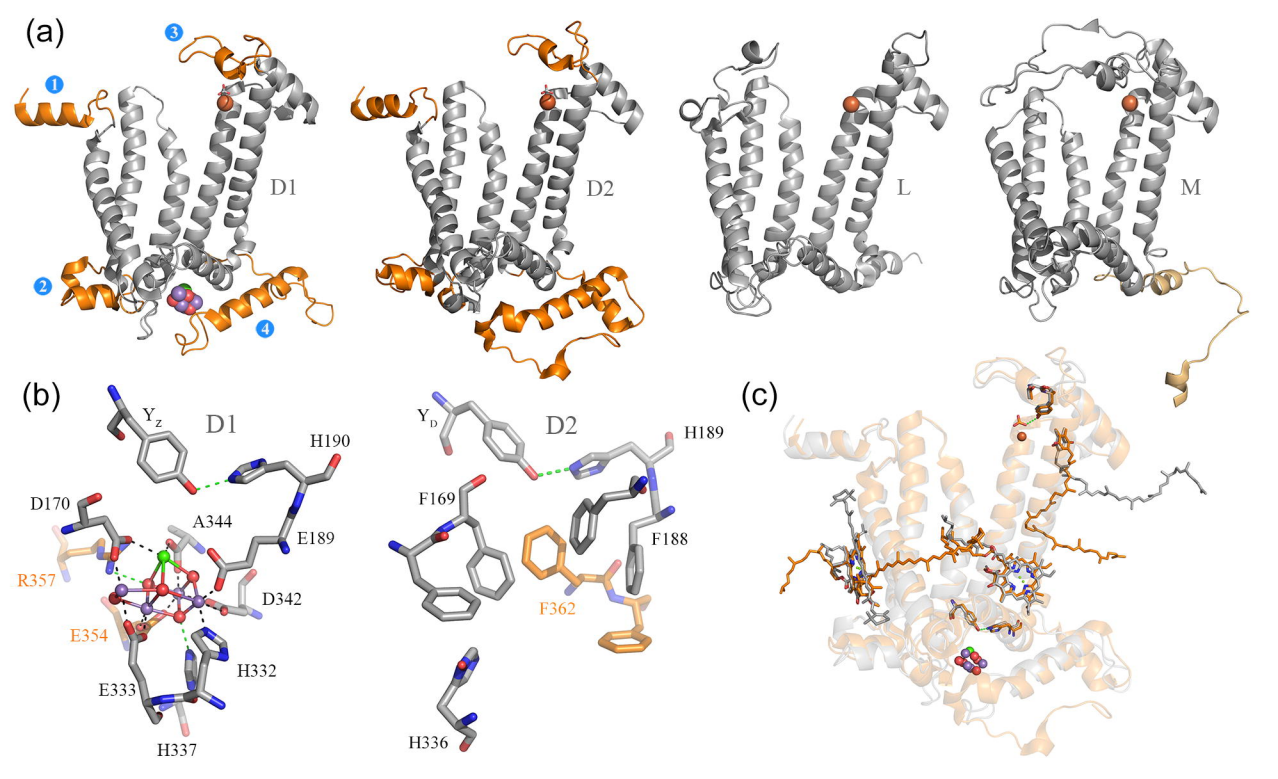
bioRxiv preprint doi: https://doi.org/10.1101/109447; this version posted September 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

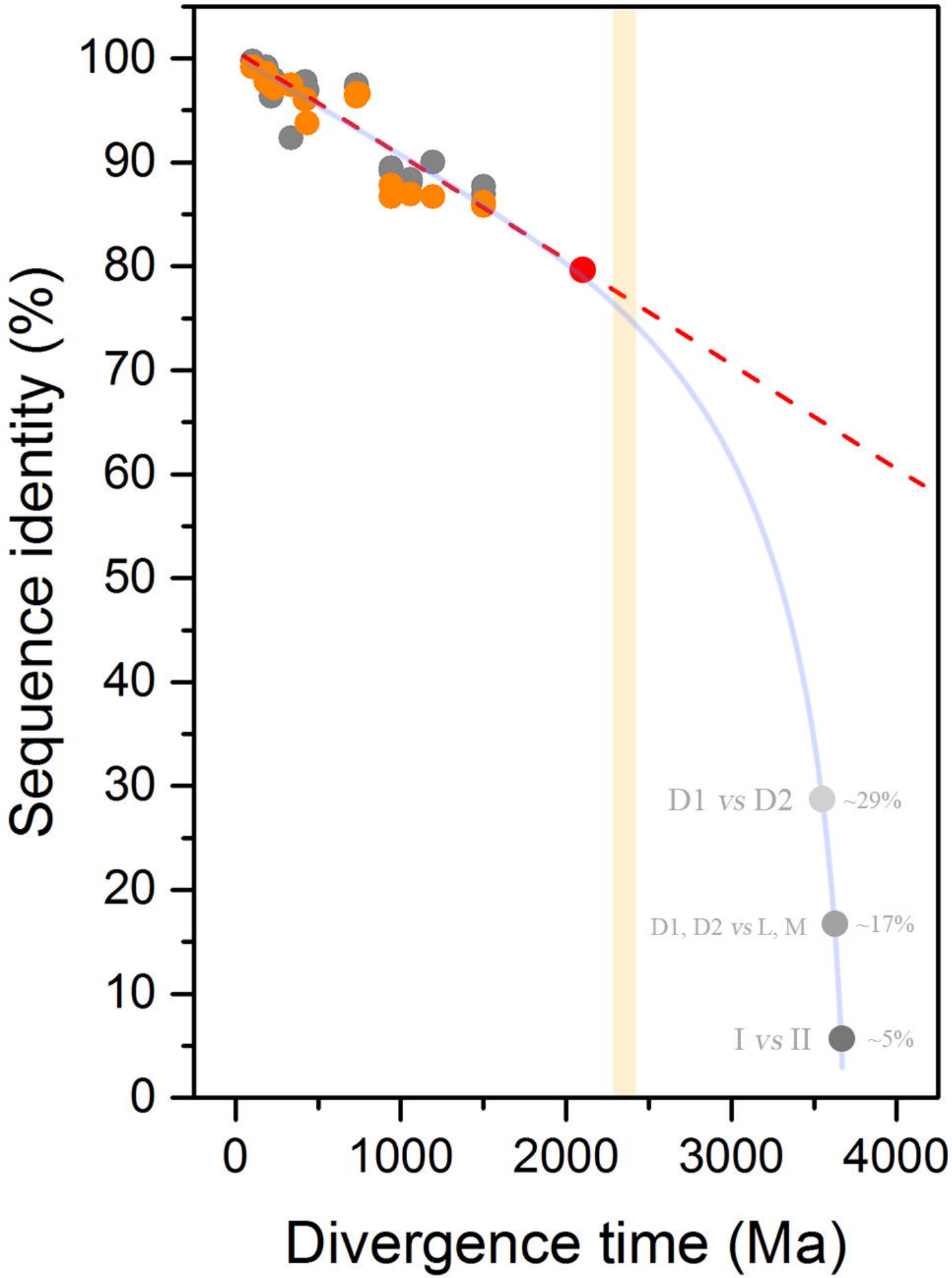
complete ligand sphere to the Mn_4CaO_5 cluster. The blue dot represents the origin of the dominant form of D1 inherited by all extant Cyanobacteria and photosynthetic eukaryotes. The gray bars represent the standard error of the estimated divergence times at the nodes. The orange bar shows the GOE

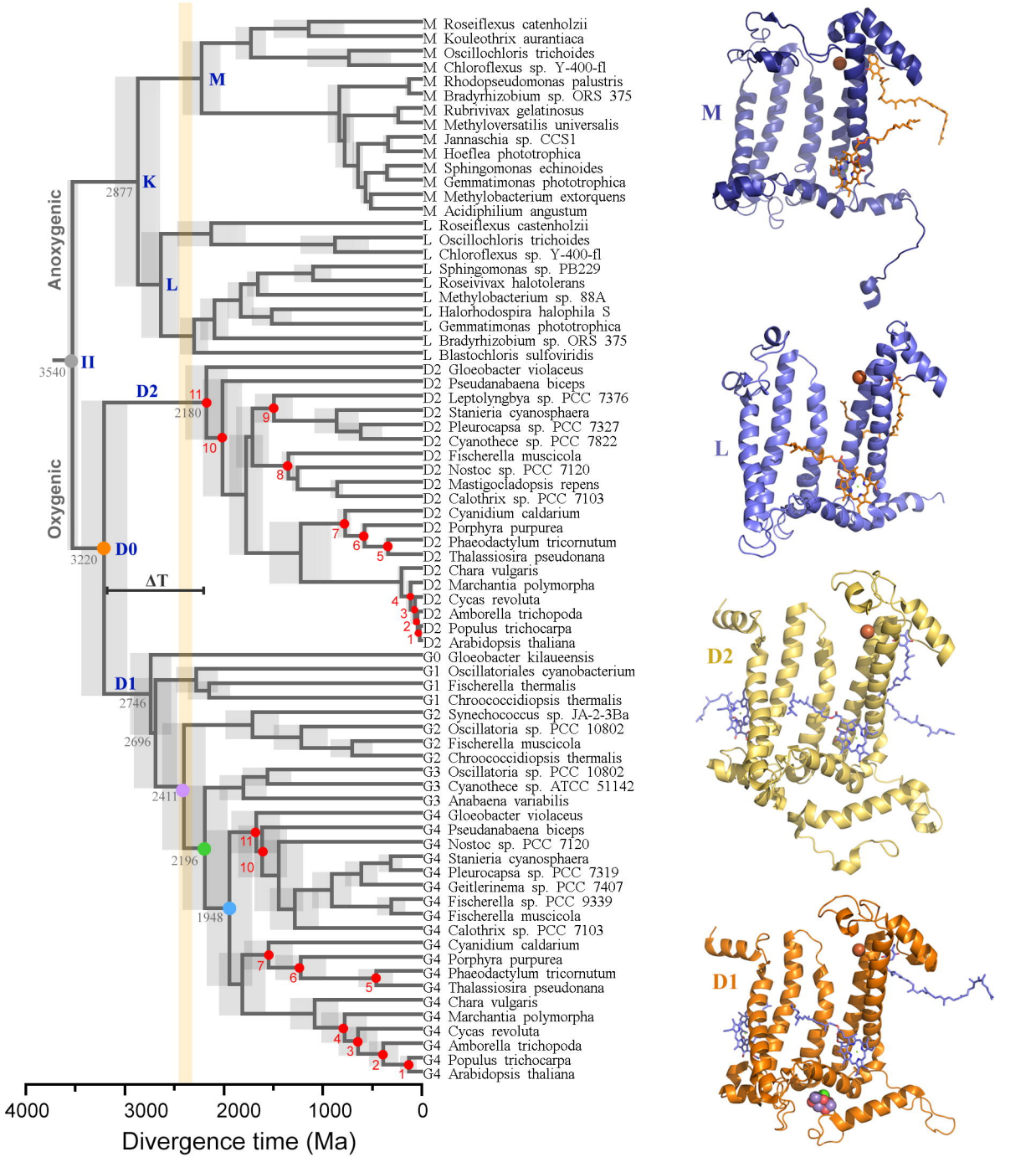
Figure 5. Rates of evolution as a function of time. (*a*) Change in the rate of evolution of oxygenic (gray) and anoxygenic (orange) Type II reaction center proteins. The rates correspond to the tree in Fig. 4, which was calculated using a root prior of 3.5 ± 0.05 Ga. The dashed lines represent a fit 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

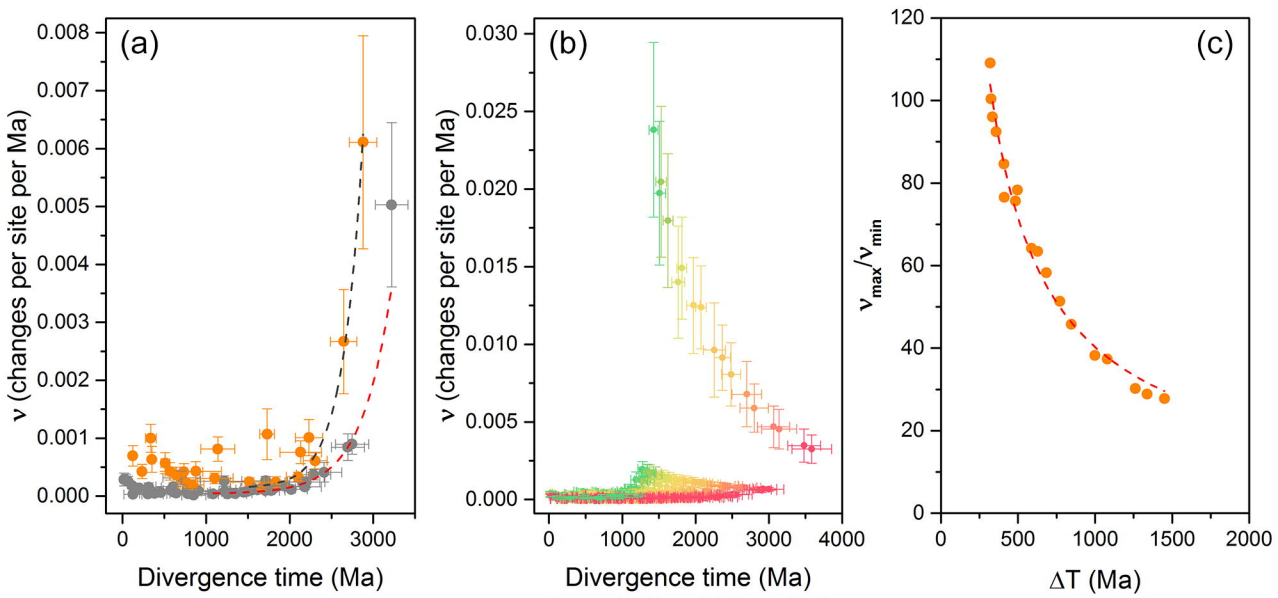
Figure 6. Scenarios for the evolution of Type II reaction centers. The rows of colored dots represent estimated divergence times calculated using the CAT + GTR + Γ and LG + Γ models and root priors of 3.8 or 3.5 ± 0.05 Ga as indicated in the illustration. Primordial forms of water oxidation emerged before the gene duplication event that led to D1 and D2 in an ancestral homodimeric photosystem (orange dot). The more ancient the ancestor of crown group Cyanobacteria is (green arrow), the closer the origin of water oxidation has to be, relative to the origin of the first photochemical reaction centers in the early Archaean (orange arrow). The yellow vertical bar marks the GOE

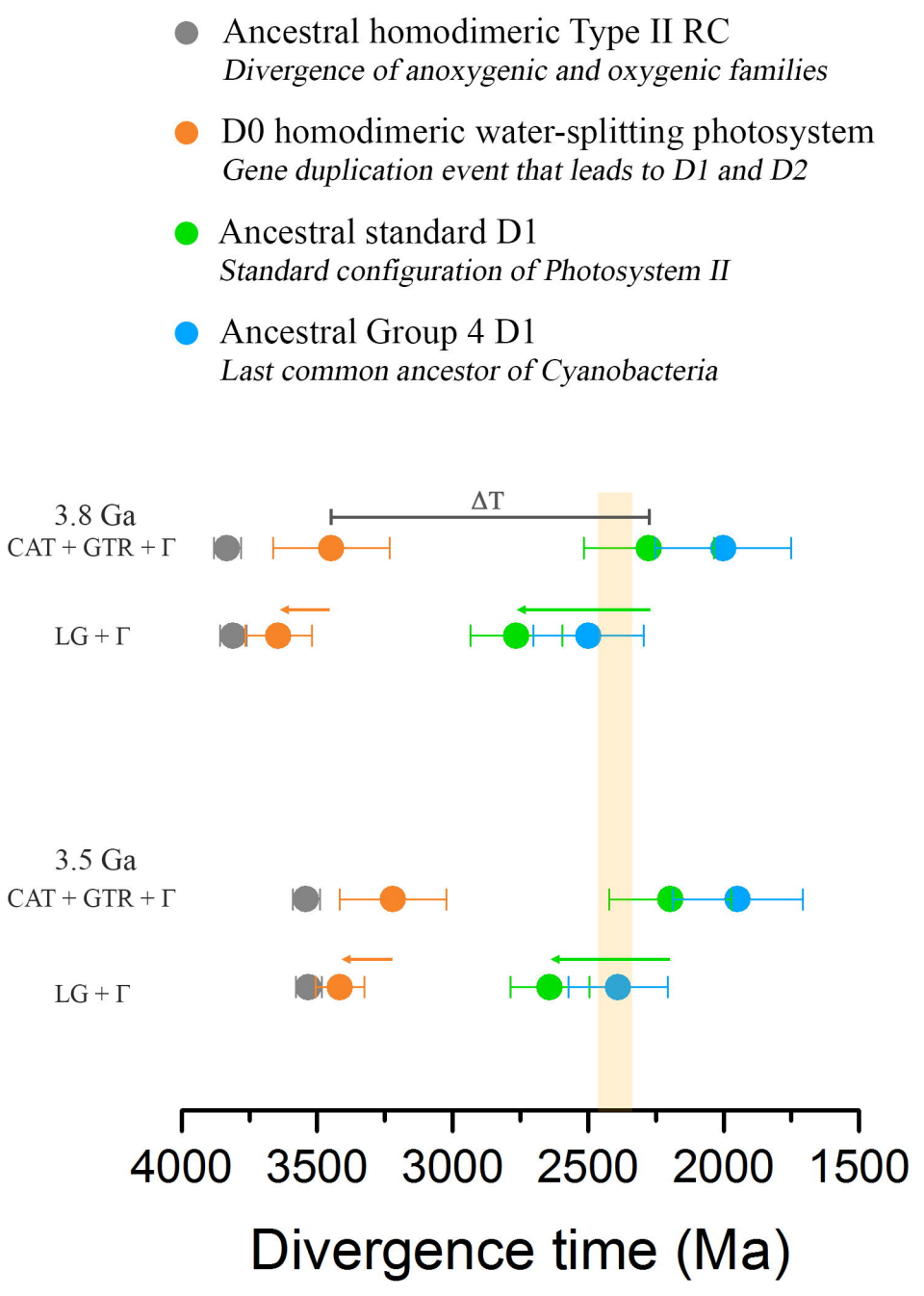

(a)
Proteobacteria
Chloroflexi
Cyanobacteria
D2
L
Oxygenic
Anoxygenic
K
II
D0
M
D1
Proteobacteria
Chloroflexi
(b)
Cyanobacteria
Chloroflexi
Proteobacteria
D1 D2
L M
L M
5
ΔT
L M
6
3
4
D0
K
2
II
I
1
A
(c)
D1 family
Standard
Atypical
D1
D1'
rD1
sD1
srD1
ChlF
D2
G4
G3
G2
G1
G0
8
Ancestral
standard
D1
7
ΔT
Ancestral
D1
3
D0


(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)


Sequence identity (%)
100
90
80
70
60
50
40
30
20
10
0
0
1000
2000
3000
4000
Divergence time (Ma)
D1 vs D2
~29%
D1, D2 vs L, M
~17%
I vs II
~5%


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
Anoxygenic
Oxygenic
M
K
L
II
D2
D0
D1
ΔT
2877
3540
2180
3220
2746
2696
2411
2196
1948
4000
3000
2000
1000
0
Divergence time (Ma)
M
L
D2
D1


(a)
ν (changes per site per Ma)
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
0
1000
2000
3000
Divergence time (Ma)
(b)
ν (changes per site per Ma)
0.030
0.025
0.020
0.015
0.010
0.005
0.000
0
1000
2000
3000
4000
Divergence time (Ma)
(c)
ν_{max}/ν_{min}
120
100
80
60
40
20
0
0
500
1000
1500
2000
ΔT (Ma)


Ancestral homodimeric Type II RC
Divergence of anoxygenic and oxygenic families
D0 homodimeric water-splitting photosystem
Gene duplication event that leads to D1 and D2
Ancestral standard D1
Standard configuration of Photosystem II
Ancestral Group 4 D1
Last common ancestor of Cyanobacteria
ΔT
3.8 Ga
CAT + GTR + Γ
LG + Γ
3.5 Ga
CAT + GTR + Γ
LG + Γ
4000
3500
3000
2500
2000
1500
Divergence time (Ma)