# Early Archean origin of heterodimeric Photosystem I

Tanai Cardona

Department of Life Sciences, Imperial College London, London, UK

Correspondence to: tanai.cardona@imperial.ac.uk

**Abstract**

When and how oxygenic photosynthesis originated remains controversial. Wide uncertainties exist for the earliest detection of biogenic oxygen in the geochemical record or the origin of water oxidation in ancestral lineages of the phylum Cyanobacteria. A unique trait of oxygenic photosynthesis is that the process uses a Type I reaction centre with a heterodimeric core, also known as Photosystem I, made of two distinct but homologous subunits, PsaA and PsaB. In contrast, all other known Type I reaction centres in anoxygenic phototrophs have a homodimeric core. A compelling hypothesis for the evolution of a heterodimeric Type I reaction centre is that the gene duplication that allowed the divergence of PsaA and PsaB was an adaptation to oxygenic photosynthesis to incorporate photoprotective mechanisms, therefore occurring after the origin of water oxidation. Here I show, using sequence comparisons and Bayesian relaxed molecular clocks that this gene duplication event may have occurred more than 3.4 billion years ago, long before the most recent common ancestor of Cyanobacteria and the Great Oxidation Event. If water oxidation predated this gene duplication event, then that would place primordial forms of oxygenic photosynthesis at a very early stage in the evolutionary history of life.

Keywords: Photosystem, Reaction Centre, Photosynthesis, Cyanobacteria, Chlorobi, Acidobacteria, Heliobacteria, Oxygenic, Anoxygenic, Water Oxidation

## Introduction

When and how oxygenic photosynthesis originated remains a highly debated subject with dates ranging from 3.8 billion years (Ga) (Rosing and Frei 2004, Frei *et al.* 2016) to shortly before 2.4 Ga (Ward *et al.* 2016, Shih *et al.* 2017), the onset of the Great Oxidation Event (GOE). Extensive geochemical data suggests that some oxygen was being produced from at least about 3.0 Ga (Lyons *et al.* 2014, Havig *et al.* 2017, Smit and Mezger 2017). Yet, it is well accepted that anoxygenic photosynthesis had already evolved 3.5 Ga ago and may predate 3.8 Ga (Tice and Lowe 2004, Nisbet and Fowler 2014, Butterfield 2015). If so, there could have been a gap of 600 million years (Ma) to more than 1.0 Ga between the origin of anoxygenic photosynthesis and the emergence of oxygenic photosynthesis. However, I have previously pointed out that the molecular evolution of photochemichal reaction centres is inconsistent with such a vast gap (Cardona 2016, Cardona 2017, Cardona *et al.* 2017). Instead, I have noted that the evolution of all reaction centre proteins, including those used in oxygenic photosynthesis, diversified early, rapidly, and in parallel (Cardona 2015). Consistent with this, Cardona *et al.* (2017) demonstrated that the divergence of the anoxygenic Type II reaction centres from water-oxidising Photosystem II (PSII) occurred soon after the emergence of photosynthesis as the result of one reaction centre gaining or losing the structural domains required to oxidise water to oxygen, still at a homodimeric stage. Cardona *et al.* (2017) also showed that one of the most ancient diversification events in the evolution of photosynthesis, after the divergence of Type I and Type II reaction centres, was the specific gene duplication of the ancestral water-splitting homodimeric core that permitted the heterodimerisation of PSII during the early Archean.

The core of Type I reaction centres in anoxygenic phototrophic bacteria is homodimeric, made of a single subunit, known as PshA in phototrophic Firmicutes, and PscA in phototrophic Chlorobi and Acidobacteria (Bryant 1994). Photosystem I (PSI) in Cyanobacteria is heterodimeric, made of two homologous core subunits named PsaA and PsaB (Figure 1). It is thought that the heterodimerisation process was an evolutionary response to oxygen (Ben-Shem *et al.* 2004, Hohmann-Marriott and Blankenship 2008, Rutherford *et al.* 2012). More exactly, it is proposed that the asymmetric fine-tuning of the redox cofactors allows back electron transfer reactions to occur safely, minimising the formation of triplet chlorophyll states that can lead to the production of reactive oxygen species (Rutherford *et al.*

2012). It can thus be postulated that the gene duplication event that led to the heterodimerisation of PSI occurred after the evolution of water oxidation and was driven by the presence of oxygen.

In this report I present the first attempt at estimating when the gene duplication event that led to the heterodimerisation of the core of PSI occurred. I show that this gene duplication event is likely to have happened well before 3.4 Ga, which is consistent with an early Archean origin of water oxidation.

**Results and discussion**

A first estimation of the time passed since the gene duplication event of the ancestral core subunit of PSI can be reached by comparing the level of sequence divergence between PsaA and PsaB in distantly related organisms. Overall, PsaA and PsaB share ~45% sequence identity (seq_id) (Cantrell and Bryant 1987, Liebl *et al.* 1993): in the particular case of *Gloeobacter kilaueensis*, belonging to the earliest branching genus of Cyanobacteria capable of oxygenic photosynthesis (Shih *et al.* 2013, Sanchez-Baracaldo 2015), PsaA and PsaB share 46.7% seq_id (Table 1). In *Arabidopsis thaliana*, a flowering plant, PsaA and PsaB share 45.6% seq_id. The similarity of seq_id between PsaA and PsaB in these two distant species suggests that in the very long time, likely more than 2.0 Ga ago, since the separation of the lineages leading to the genus *Gloeobacter* and the genus *Arabidopsis* (Sanchez-Baracaldo 2015, Shih *et al.* 2017), the rates of evolution of the core subunits have not varied drastically, resulting in a difference of seq_id of only 1.1%.

The PsaB subunit in *G. kilaueensis* and in *A. thaliana* share 72.8% seq_id and the PsaA subunit between these two species share 68.1% seq_id. Then, let us assume for the sake of simplicity that the most recent common ancestor (MRCA) of Cyanobacteria, at the divergence of the genus *Gloeobacter*, occurred around the time of the GOE: 2.4 Ga. This would mean that since the branching of *Gloeobacter*, at 2.4 Ga, PsaB subunits have only changed by a maximum of about 27% and PsaA by about 32%. Again for simplicity, let us average this to 30%. It follows then that since 2.4 Ga the core subunits of PSI, PsaA and PsaB, have changed roughly about 30% each. That would give an average rate of evolution of about 1% loss of seq_id per every 80 Ma.

Using this rate we can then predict that two species that diverged about 100 Ma ago should have a level of sequence identity of their PsaA and PsaB subunits of just below **99%**. *A. thaliana* and *Populus trichocarpa* are known to have diverged about 100 Ma ago, or sometime between 82 and 127 Ma as inferred

from fossil evidence (Clarke *et al.* 2011). The PsaA seq_id between *A. thaliana* and *P. trichocarpa* is **98.6%** and that of PsaB is **98.2%**, which is in agreement with our approximated rate.

The earliest widely accepted fossil of red algae is 1.2 Ga old (Butterfield 2000). There is also a recently characterised earlier fossil 1.6 Ga old (Bengtson *et al.* 2017). Using the approximated rate we can predict that the level of sequence divergence between *Cyanidioschyzon merolae* (a red algae) and *A. thaliana* should be between **85%** based on the younger fossil and **80%** based on the older fossil (1200/80 = **15%** or 1600/80 = **20%**). In agreement with the approximated rate of evolution, the PsaA seq_id between these species is **82.2%** and the PsaB seq_id is **80.9%**.

From these comparisons it can be concluded that the rate of evolution of PsaA and PsaB has not changed dramatically for more than 2.4 Ga and that the approximated rate is not likely to be too far off from the average rates of evolution found in the wild.

Now, using this approximated rate we can extrapolate back in time to estimate when the gene duplication event that led to the divergence of PsaA and PsaB started. PsaA and PsaB share about 46% seq_id, applying a rate of change of 1% loss of seq_id per 80 Ma gives a result of 4.3 Ga (54*80 = 4320 Ma). This is a very implausible date, but it is useful to illustrate a few key characteristics of the evolution of photosynthesis and reaction centre proteins as described in more detail below.

Let us assume instead that the MRCA of Cyanobacteria occurred at 3.0 Ga, 600 Ma before the GOE. That assumption would result in a slower rate of evolution of about 1% loss of seq_id for every 100 Ma. Consequently, it would set the gene duplication event at 5.30 Ga. In the other extreme case, if the MRCA postdate the GOE as suggested by Shih *et al.* (2017), at 2.0 Ga, that would result in a faster rate of evolution of 1% loss of seq_id for every 66 million years. If this rate is applied to the divergence of PsaA and PsaB, it would set the gene duplication event at 3.56 Ga (54*66 = 3564 Ma), still more than 1.0 Ga before the GOE. It becomes therefore apparent judging from a comparison of the level of seq_id alone that regardless of when the MRCA of Cyanobacteria existed (2.0 to 3.0 Ga), the gene duplication that led to PsaA and PsaB is more likely to be a tremendously ancient event, tracing back to the earliest stages in the evolutionary history of photosynthesis, than a late event occurring just before the GOE or after the GOE.

From these basic considerations, and even allowing for very large uncertainties, it can also be concluded that any scenario for the divergence of PsaA and PsaB occurring *after* 3.5 Ga would require faster rates of evolution at the earliest stages of divergence than any observed since the Proterozoic.

To validate the above observations I performed a Bayesian molecular clock analysis on Type I reaction centre proteins assuming a root of either 3.5 ± 0.05 or 3.8 ± 0.05 Ga (Figure 2). In other words, assuming that photochemical reaction centres originated in the early Archean, which is consistent with extensive isotopic and sedimentological evidence for photoautotrophic metabolism during this time (Schidlowski 1988, Rosing 1999, Rosing and Frei 2004, Tice and Lowe 2004, Grassineau *et al.* 2006, Czaja *et al.* 2013, Nisbet and Fowler 2014, Knoll 2015). The estimated divergence time of PsaA and PsaB when constraining the MRCA of Cyanobacteria before the GOE was in both cases the oldest node after the root. An estimated divergence time of 3.41 ± 0.09 Ga was computed for the younger root and 3.62 ± 0.14 Ga for the older root. To test the effect of the calibration point for the MRCA of Cyanobacteria (Figure 2, point 9) on the estimated divergence time for the split of PsaA and PsaB, I repeated the same analysis this time removing the constraint altogether or constraining it to a minimum age of 3.0 Ga. The plots shown in Figure 3 indicate that the calibration choice for the MRCA of Cyanobacteria had little or no effect on the timing of the gene duplication event of the core subunits of PSI.

Consistent with the conclusions derived from the comparison of the level of seq_id, even such ancient origins for the divergence of PsaA and PsaB still require high rates of evolution at the earliest stages. If the origin of photosynthesis happened about 3.5 Ga ago, that would require a rate at the time of the gene duplication event that led to PsaA and PsaB of 2.98 ± 0.79 amino acid changes per site per Ga, which is about 24 times larger than the average rate observed during the Proterozoic (0.12 ± 0.03 changes per site per Ga). If the origin of photosynthesis happened about 3.8 Ga ago, that would only slow down the initial rate to 2.76 ± 0.78 changes per site per Ga, with virtually no change observed for the average rate seen in the Proterozoic (Figure 4). This result implies that the phylogenetic distance between PsaA and PsaB is too large and the rate of evolution too slow, for the gene duplication event to have occurred within a few million years or a few hundred million years before the GOE. This is also in good agreement with the evolution of Type II reaction centres, which indicated that a water-splitting homodimeric photosystem is likely to have arisen in the early Archean (Cardona 2017, Cardona *et al.* 2017).

It has been suggested that Cyanobacteria could have obtained photosynthesis via horizontal gene transfer (HGT) from anoxygenic photosynthetic bacteria (Raymond 2009, Hohmann-Marriott and Blankenship 2011). A recent version of this hypothesis suggests that a non-photosynthetic ancestor of Cyanobacteria acquired photosynthesis via massive HGT right before the GOE (Shih *et al.* 2017, Soo *et al.* 2017). The molecular evolution of Type I reaction centres disproves this hypothesis. This is because the ancestral PSI had to be a homodimer and there are no anoxygenic bacteria with heterodimeric Type I reaction centres, which means that the hypothesised HGT had to occur before the gene duplication event. As we have seen above, it is unlikely that this duplication event occurred immediately before the GOE or after the GOE, given the fact that at this time PsaA and PsaB had already achieved a high degree of sequence divergence. These arguments also apply to Type II reaction centres as discussed in Cardona *et al.* (2017).

The high level of sequence divergence between the two core subunits of each photosystem, a trait shared by all Cyanobacteria, highlights the fact that their most recent common ancestor existed at an advanced stage in the evolution of oxygenic photosynthesis. This makes evident that a great diversity of oxygenic phototrophic bacteria spanning from the origin of water oxidation (using homodimeric reaction centres) to the emergence of crown group Cyanobacteria (using very sophisticated multiprotein photosystems with very divergent heterodimeric cores), either still remain undiscovered or have gone completely extinct. From the low level of sequence identity of the core subunits in addition to very slow rates of evolution it can be deduced with confidence that most of the diversity of oxygenic phototrophic bacteria actually predated crown group Cyanobacteria. Indeed, the evolutionary distance between the core subunits is so large that the early diversity of oxygenic phototrophic bacteria was likely not contained within a single ancestral phylum. Consequently, the possibility that Cyanobacteria are not the only existing phylum of bacteria to have descended from oxygenic phototrophs should be carefully considered in scenarios for the early evolution of bioenergetic processes.

In conclusion, the data presented here indicates that the gene duplication event that allowed the heterodimerisation of Photosystem I—in an ancestral lineage of bacteria that predated the most recent common ancestor of the phylum Cyanobacteria—likely occurred in the early Archean. If such event was

an adaptation to photosynthetic oxygen evolution it would set water oxidation more than a billion years before the Great Oxidation Event.

**Materials and Methods**

All sequences used in this study were retrieved from the NCBI refseq database and are provided in the supplementary data. The percentage of sequence identity between pair of sequences was estimated using BLAST with the longest sequence of the pair taken as 100%. Sequence alignments were performed using Clustal Omega employing 10 combined guide trees and Hidden Markov Model iterations (Sievers *et al.* 2011). Maximum Likelihood trees were calculated with PhyML 3.0 using the smart model selection option allowing the software to compute all parameters from the data based on the Bayesian Information Criterion (Lefort *et al.* 2017).

A total of 59 sequences spanning the entire diversity of reaction centre proteins, including 10 PsaA and 10 PsaB sequences from photosynthetic eukaryotes, were selected for molecular clock analysis. Calibration points from well-established fossils were assigned as shown in Figure 2 and age boundaries were allocated as listed in Table 2. Minimum and maximum divergence times for the *Arabidopsis*/*Populus* divergence, the appearance of the angiosperms, gymnosperms, and land plants were obtained from Clarke *et al.* (2011). A minimum age for diatoms were obtained from Jurassic fossils from the Lias Group, reviewed by Sims *et al.* (2006). A minimum age of 600 Ma was assigned to the node between the diatom sequences and the sequences from *Porphyra purpurea*. These were selected based on a Late Neoproterozoic Chinese multicellular red alga fossil (Xiao *et al.* 2004) and I used it as a conservative estimate for the divergence of complex red algae, which are probably older (Yang *et al.* 2016). The well-preserved 1.6 Ga red algae fossils recently reported by Bengtson *et al.* (2017) was used as a minimum age for the origin of photosynthetic eukaryotes, an age which is also conservative and in agreement with other recent molecular clock analysis (Yang *et al.* 2016, Sánchez-Baracaldo *et al.* 2017). In addition, a minimum age of 1.6 Ga was assigned to the divergence of heterocystous cyanobacteria as well-preserved akinetes of this age have been found in cherts from Siberia, China, and Australia (Golubic *et al.* 1995, Tomitani *et al.* 2006, Schirrmeister *et al.* 2016). A minimum age of 2.45 Ga was assigned to the node at the divergence of *Gloeobacter* (Point 9 in Figure 2) assuming that the most recent common ancestor of Cyanobacteria is at

least as old as the GOE (Bekker *et al.* 2004). In addition, the same analysis was also performed using an older minimum age for the MRCA of Cyanobacteria of 3.0 Ga, to consider the possibility that some of the earliest evidence for oxygen (Lyons *et al.* 2014) may have arisen from crown group Cyanobacteria. Alternatively, and to test the other extreme situation, I repeated the analysis removing this calibration point to release the oldest calibrated node from any time bias.

A Bayesian Markov chain Monte Carlo approach was used to infer estimated divergence times with the Phylobayes 3.3f software (Lartillot *et al.* 2009). A relaxed log-normal autocorrelated molecular clock was applied together with the CAT model (Lartillot and Philippe 2004, Lartillot *et al.* 2009) of equilibrium frequency profiles and applying a uniform model of amino acid substitutions. Flexible bounds on the calibration points were left as default, 2.5% tail probability falling outside each calibration boundary or 5% in the case of a single minimum boundary (Yang and Rannala 2006). All molecular clocks were computed using four discrete categories for the gamma distribution and four chains were run in parallel until convergence. Rates of evolution, expressed as amino acid changes per site per unit of time, are inferred by the software as described by the developers elsewhere (Kishino *et al.* 2001, Lepage *et al.* 2007).

**Acknowledgement**

The financial support of an Imperial College London Junior Research Fellowship and a Research Grant from the Leverhulme Trust (RPG-2017-223) is gratefully acknowledged.

**References**

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

Ben-Shem, A., F. Frolow and N. Nelson (2004). "Evolution of photosystem I - from symmetry through pseudosymmetry to asymmetry." FEBS Lett **564**(3): 274-280. DOI: 10.1016/S0014-5793(04)00360-6.

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

Bryant, D. A. (1994). "Gene nomenclature recommendations for green photosynthetic bacteria and heliobacteria." Photosynthesis Res **41**(1): 27-28. DOI: 10.1007/Bf02184142.

Butterfield, N. J. (2000). "*Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes." Paleobiology **26**(3): 386-404. DOI: 10.1666/0094-8373(2000)026<0386:Bpngns>2.0.Co;2.

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

Cantrell, A. and D. A. Bryant (1987). "Molecular cloning and nucleotide sequence of the *psaA* and *psaB* genes of the cyanobacterium *Synechococcus* sp. PCC 7002." Plant Mol Biol **9**(5): 453-468. DOI: 10.1007/BF00015877.

Cardona, T. (2015). "A fresh look at the evolution and diversification of photochemical reaction centers." Photosynthesis 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?" Front Plant Sci **7**: 257. DOI: 10.3389/fpls.2016.00257.

Cardona, T. (2017). "Photosystem II is a chimera of reaction centers." J Mol Evol **84**(2-3): 149-151. DOI: 10.1007/s00239-017-9784-x.

Cardona, T., P. Sanchez-Baracaldo, A. W. Rutherford and A. W. D. Larkum (2017). "Molecular evidence for the early evolution of photosynthetic water oxidation." BioRxiv **109447**. DOI: https://doi.org/10.1101/109447.

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

Czaja, A. D., C. M. Johnson, B. L. Beard, E. E. Roden, W. Q. Li and S. Moorbath (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.

Frei, R., S. A. Crowe, M. Bau, A. Polat, D. A. Fowle and L. N. Døssing (2016). "Oxidative elemental cycling under the low $O_2$ Eoarchean atmosphere." Sci Rep **6**: 21058. DOI: 10.1038/srep21058.

Golubic, S., V. N. Sergeev and A. H. Knoll (1995). "Mesoproterozoic Archaeoellipsoides: Akinetes of heterocystous cyanobacteria." Lethaia **28**(4): 285-298. DOI: 10.1111/j.1502-3931.1995.tb01817.x.

Grassineau, N. V., P. I. Abell, P. W. U. Appel, D. Lowry and E. G. Nisbet (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. DOI: https://doi.org/10.1130/2006.1198(02).

Havig, J. R., T. L. Hamilton, A. Bachan and L. R. Kump (2017). "Sulfur and carbon isotopic evidence for metabolic pathway evolution and a four-stepped Earth system progression across the Archean and Paleoproterozoic." Earth-Sci Rev **174**: 1-21. DOI: https://doi.org/10.1016/j.earscirev.2017.06.014.

Hohmann-Marriott, M. F. and R. E. Blankenship (2008). Anoxygenic Type-I photosystems and evolution of photosynthetic reaction centers. Photosynthetic Protein Complexes. P. Fromme, Wiley-VCH Verlag GmbH & Co. KGaA**:** 295-324.

Hohmann-Marriott, M. F. and R. E. Blankenship (2011). "Evolution of photosynthesis." Annu Rev Plant Biol **62**: 515-548. DOI: 10.1146/annurev-arplant-042110-103811.

Kishino, H., J. L. Thorne and W. J. Bruno (2001). "Performance of a divergence time estimation method under a probabilistic model of rate evolution." Mol Biol Evol **18**(3): 352-361. DOI: 10.1093/oxfordjournals.molbev.a003811.

Knoll, A. H. (2015). "Paleobiological perspectives on early microbial evolution." Cold Spring Harb Perspect Biol **7**(7): a018093. DOI: 10.1101/cshperspect.a018093.

Lartillot, N., T. Lepage and S. Blanquart (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. and H. Philippe (2004). "A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process." Mol Biol Evol **21**(6): 1095-1109. DOI: 10.1093/molbev/msh112.

Lefort, V., J. E. Longueville and O. Gascuel (2017). "SMS: smart model selection in PhyML." Mol Biol Evol **34**(9): 2422-2424. DOI: 10.1093/molbev/msx149.

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

Liebl, U., M. Mockensturm-Wilson, J. T. Trost, D. C. Brune, R. E. Blankenship and W. Vermaas (1993). "Single core polypeptide in the reaction center of the photosynthetic bacterium *Heliobacillus mobilis:* structural implications and relations to other photosystems." Proc Natl Acad Sci USA **90**(15): 7124-7128. DOI: 10.1073/pnas.90.15.7124.

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

Nisbet, E. G. and C. F. R. Fowler (2014). The early history of life. Treatise on Geochemistry. K. D. M. and W. H. Schlesinger. Amsterdam, Elsevier Science. **10:** 1-42.

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

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

Rosing, M. T. and R. Frei (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., A. Osyczka and F. Rappaport (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.

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

Sánchez-Baracaldo, P., J. A. Raven, D. Pisani and A. H. Knoll (2017). "Early photosynthetic eukaryotes inhabited low-salinity habitats." Proc Natl Acad Sci USA. DOI: 10.1073/pnas.1620089114.

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

Schirrmeister, B. E., P. Sanchez-Baracaldo and D. Wacey (2016). "Cyanobacterial evolution during the Precambrian." Int J Astrobiol **15**: 187-204. DOI: 10.1017/S1473550415000579.

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

Shih, P. M., D. Wu, A. Latifi, S. D. Axen, D. P. Fewer, E. Talla, A. Calteau, F. Cai, N. Tandeau de Marsac, R. Rippka, M. Herdman, K. Sivonen, T. Coursin, T. Laurent, L. Goodwin, M. Nolan, K. W. Davenport, C. S. Han, E. M. Rubin, J. A. Eisen, T. Woyke, M. Gugger and C. A. Kerfeld (2013). "Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing." Proc Natl Acad Sci USA **110**(3): 1053-1058. DOI: 10.1073/pnas.1217107110.

Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Z. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson and D. G. Higgins (2011). "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega." Mol Syst Biol **7**. DOI: 10.1038/msb.2011.75.

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

Smit, M. A. and K. Mezger (2017). "Earth's early $O_2$ cycle suppressed by primitive continents." Nat Geosci **10**(10): 788-792. DOI: 10.1038/Ngeo3030.

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

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

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

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

Xiao, S., A. H. Knoll, X. Yuan and C. M. Pueschel (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., S. M. Boo, D. Bhattacharya, G. W. Saunders, A. H. Knoll, S. Fredericq, L. Graf and H. S. Yoon (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. and B. Rannala (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.

**Table 1.** Percentage of sequence identity of PsaA and PsaB in various species

| Subunit comparisons | Amino acid identities | Sequence identity (%) |
|---|---|---|
| PsaA *G. kilaueensis* vs. PsaB *G. kilaueensis* | 370/791 | 46.7 |
| PsaA *A. thaliana* vs. PsaB *A. thaliana* | 342/750 | 45.6 |
| PsaA *G. kilaueensis* vs. PsaB *A. thaliana* | 337/791 | 42.6 |
| PsaB *G. kilaueensis* vs. PsaA *A. thaliana* | 339/750 | 45.2 |
| PsaA *G. kilaueensis* vs. PsaA *A. thaliana* | 535/734 | 72.8 |
| PsaB *G. kilaueensis* vs. PsaB *A. thaliana* | 539/790 | 68.1 |
| PsaA *A. thaliana* vs. PsaA *P. trichocarpa* | 740/750 | 98.6 |
| PsaB *A. thaliana* vs. PsaB *P. trichocarpa* | 721/734 | 98.2 |
| PsaA *A. thaliana* vs. PsaA *C. merolae* | 617/750 | 82.2 |
| PsaB *A. thaliana* vs. PsaB *C. merolae* | 594/734 | 80.9 |

**Table 2.** 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** | MRCA of photosynthetic eukaryotes | - | 1600 |
| **8** | Heterocystous Cyanobacteria | - | 1600 |
| **9** | MRCA of Cyanobacteria | - | - /2450/3000 |

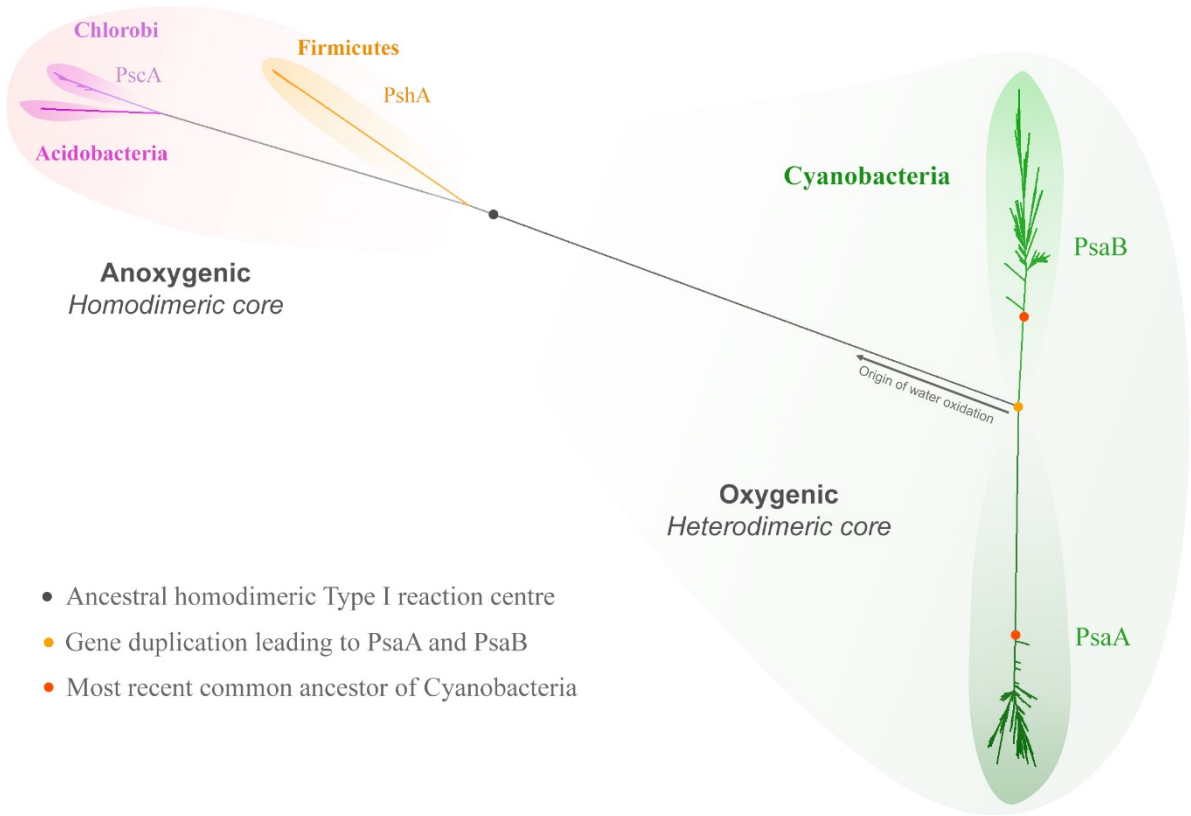


**Figure 1.** Maximum likelihood phylogeny of Type I reaction centre proteins. The tree is characterised by a deep split of reaction centre proteins, which separates those employed in anoxygenic phototrophic bacteria from those employed in oxygenic photosynthesis (grey spot). Cyanobacteria and photosynthetic eukaryotes are the only known phototrophs to have a heterodimeric Type I reaction centre made of two subunits known as PsaA and PsaB. All extant Cyanobacteria descended from a common ancestor that already had highly divergent PsaA and PsaB subunits (red spot). The gene duplication that led to PsaA and PsaB occurred at an earlier point in time (orange spot), which predated the most recent common ancestor of Cyanobacteria by an unknown period of time. It is hypothesised that the gene duplication that led to PsaA and PsaB occurred as a specialisation to oxygenic photosynthesis, therefore water oxidation should have originated before this gene duplication event (arrow).

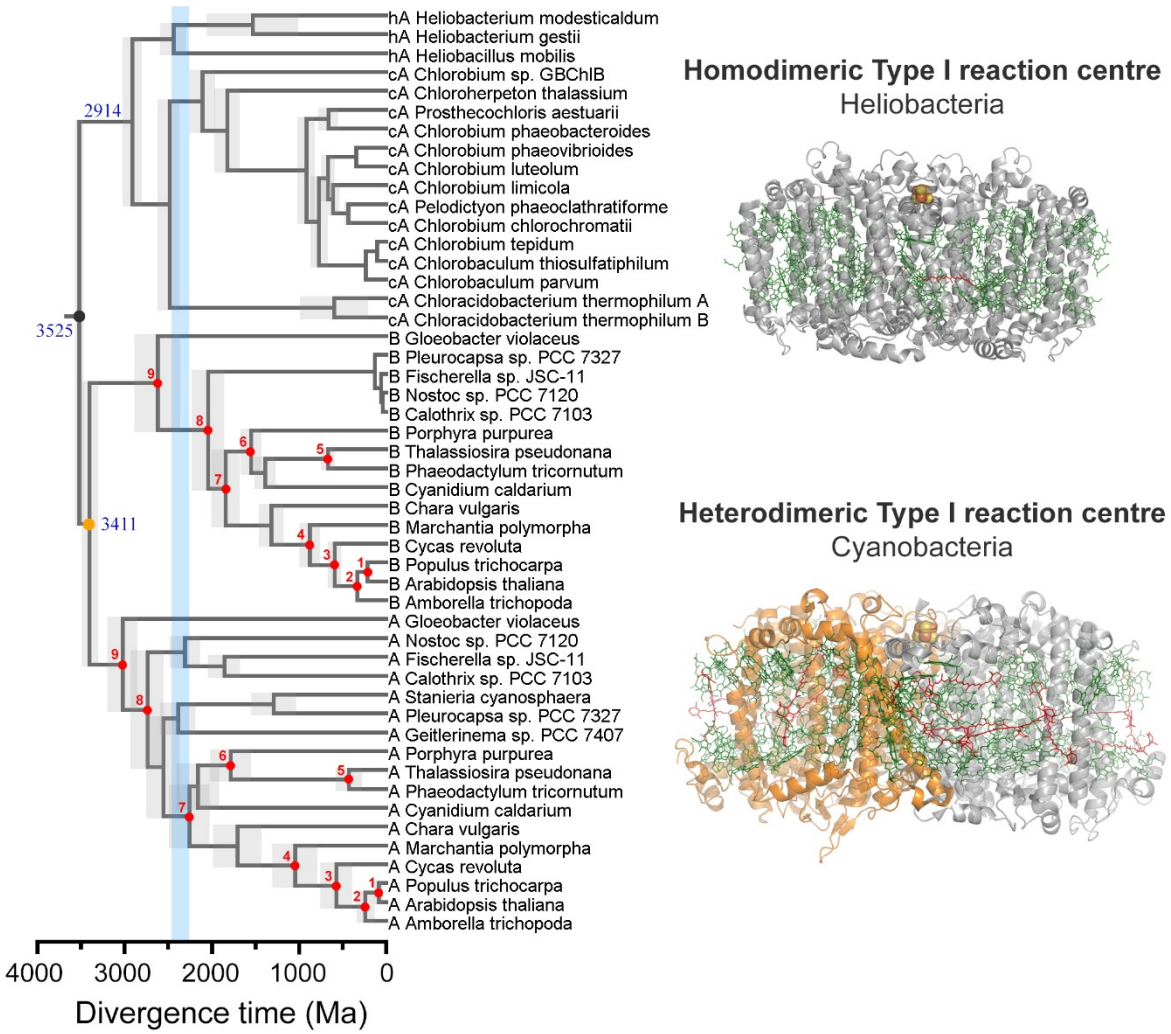


**Figure 2.** Bayesian relaxed molecular clock of Type I reaction centres. The tree was calculated assuming that Type I reaction centres had originated by 3.5 Ga (grey dot). The orange dot indicates the gene duplication event that allowed the divergence of PsaA and PsaB. Red dots highlight the nodes that were calibrated as described in Materials and Methods. The light grey bars along nodes show the standard error on the estimated divergence time. Sequences marked as hA denote those from strains of Heliobacteria, while those marked as cA from Chlorobi and Acidobacteria. Sequences marked as A and B represent PsaA and PsaB from Cyanobacterial and eukaryotic PSI. The blue bar marks the Great Oxidation Event.

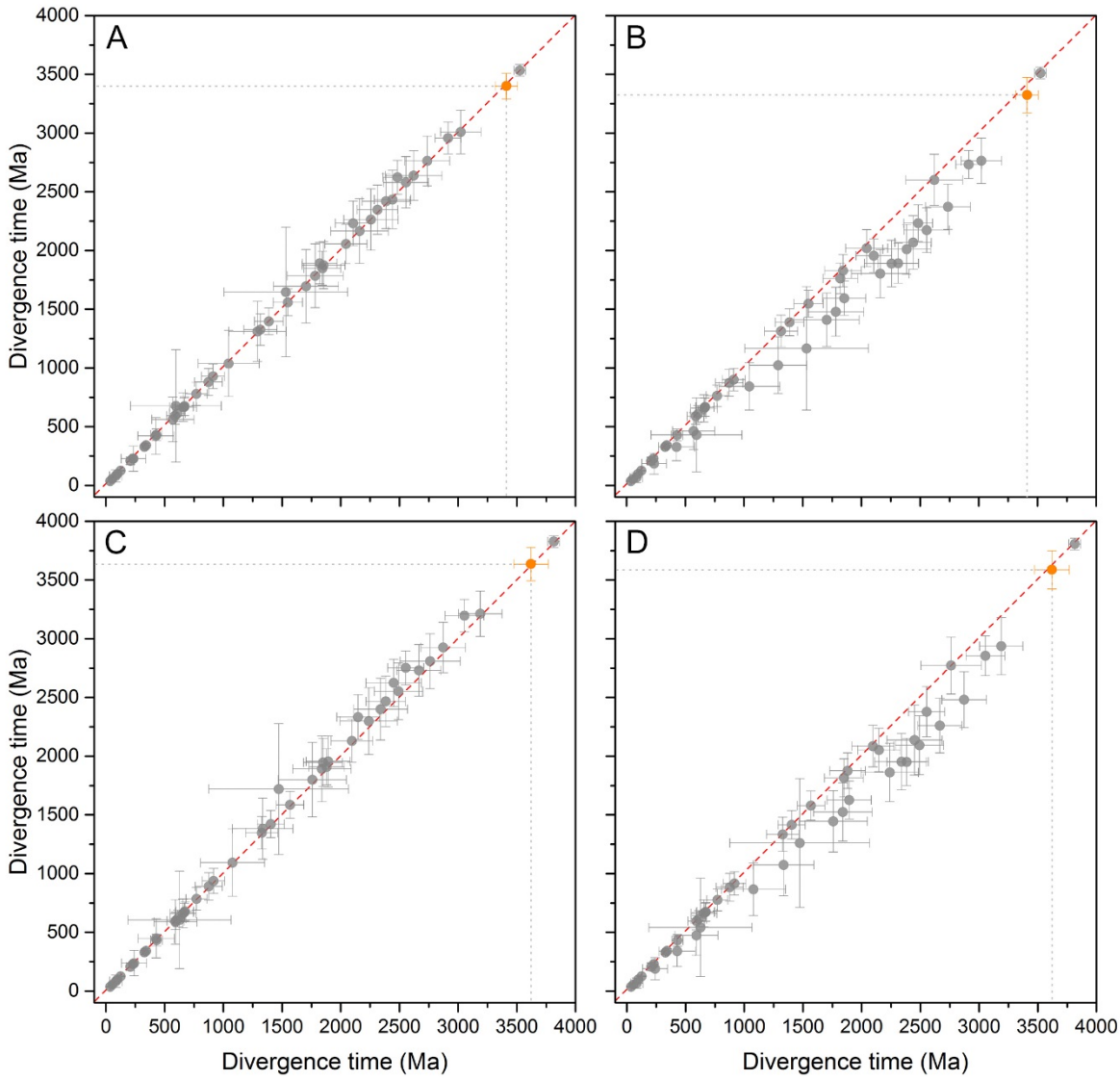


**Figure 3.** Effect of calibration choice on the estimated divergence time. (A) Comparison of divergence times of a tree calculated using a root prior of 3.5 ± 0.05 Ga and a minimum age for the MRCA of Cyanobacteria of 2.45 Ga (*x* axis) against a tree calculated using a minimum age for the MRCA of 3.00 Ga (*y* axis). There is no deviation from the slope (red dashed line) meaning that the calibration did not have a strong effect on the estimated divergence time. The orange dot represents the divergence time of the gene duplication of PsaA and PsaB. (B) Divergence times of a tree calculated using a root prior of 3.5 ± 0.05 Ga and a minimum age for the MRCA of Cyanobacteria of 2.45 Ga (*x* axis) against a tree calculated without a calibration on this point (*y* axis). There is a decrease in the estimated age on a number of nodes, however the age of the PsaA/PsaB duplication event (orange dot) remains virtually unchanged. (C) and

(D) were calculated as (A) and (B) respectively, but employing a root prior of 3.8 ± 0.05 Ga on the root. In all scenarios the gene duplication that led to PsaA and PsaB is the oldest node after the root.

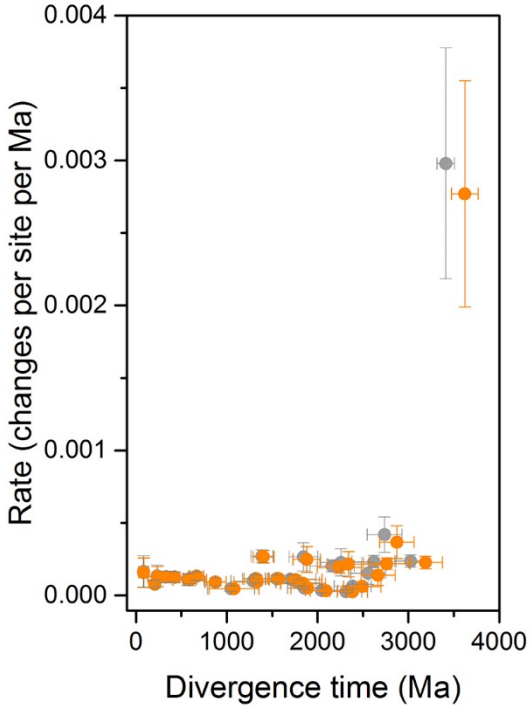


**Figure 4.** Rates of evolution as a function of divergence time. Grey dots represent divergence times of a tree calculated using a root prior of 3.5 ± 0.05 Ga, while orange dots represent those calculated using a root prior of 3.8 ± 0.05 Ga. The fastest rate of evolution occurs at the gene duplication event that leads to PsaA and PsaB. The plot shows that higher rates of evolution, measured as amino acid changes per site per million years, are necessary to explain the divergence of PsaA and PsaB at any point in time in the history of life.