Frequent transitions in self-assembly across the evolution of a central metabolic enzyme

Franziska L. Sendker[1]§, Tabea Schlotthauer[1]§, Christopher-Nils Mais[2]+, Yat Kei Lo[2]+, Mathias Girbig[1], Stefan Bohn[3], Thomas Heimerl[2], Daniel Schindler[2,4], Arielle Weinstein[5], Brain P. Metzger[5], Joseph W. Thornton[5,6], Arvind Pillai[5], Gert Bange[1,2,7], Jan M. Schuller[2,7], Georg K.A. Hochberg[1,2,7]*

1 Max-Planck-Institute for Terrestrial Microbiology; Karl-von-Frisch-Str. 10, 35043 Marburg, Germany

2 Center for Synthetic Microbiology (SYNMIKRO), Philipps-University Marburg; Karl-von-Frisch-Str. 14, 35043 Marburg, Germany

3 Institute of Structural Biology, Helmholtz Center Munich, Ingolstädter Landstraße 1 Neuherberg, Germany

4 MaxGENESYS Biofoundry, Max-Planck-Institute for Terrestrial Microbiology; Karl-von-Frisch-Str. 10, 35043 Marburg, Germany

5 Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA

6 Department of Human Genetics, University of Chicago, Chicago, IL, USA

7 Department of Chemistry, Philipps-University Marburg; Hans-Meerwein-Str. 4, 35043 Marburg, Germany

§ These authors contributed equally. + These authors contributed equally. * Correspondence to georg.hochberg@mpi-marburg.mpg.de

Abstract

Many enzymes assemble into homomeric protein complexes comprising multiple copies of one protein. Because structural form is usually assumed to follow function in biochemistry, these assemblies are thought to evolve because they provide some functional advantage. In many cases, however, no specific advantage is known and, in some cases, quaternary structure varies among orthologs. This has led to the proposition that self-assembly may instead vary neutrally within protein families. The extent of such variation has been difficult to ascertain because quaternary structure has until recently been difficult to measure on large scales. Here, we employ mass photometry, phylogenetics, and structural biology to interrogate the evolution of homo-oligomeric assembly across the entire phylogeny of prokaryotic citrate synthases – an enzyme with a highly conserved function. We discover a menagerie of different assembly types that come and go over the course of evolution, including cases of parallel evolution and reversions from complex to simple assemblies. Functional experiments in vitro and in vivo indicate that evolutionary transitions between different assemblies do not strongly influence enzyme catalysis. Our work suggests that enzymes can wander relatively freely through a large space of possible assemblies and demonstrates the power of characterizing structure-function relationships across entire phylogenies.

Main Text

Introduction

Proteins commonly fulfill their physiological functions not as an individual polypeptide chain but as multimeric complexes formed by two or more copies of the same protomer that assemble via non-covalent interactions[1–3]. Self-assembly into homo-oligomers requires specific interfaces and complementary interactions between many amino acids. Because these features seem unlikely to originate just by chance, self-assembly is often assumed to be functionally advantageous[4,5]. Proposed advantages of self-assembly include efficient encoding and easier folding of large structures, increasing productive encounters with substrate, reducing vulnerability to degradation, and particular forms of allosteric regulation[1]. But cases in which a functional benefit has been clearly demonstrated for homo-oligomeric assembly[6–8] are almost certainly outnumbered by those where none is known. Self-assembly can also vary between orthologs that at least in theory fulfill the same function[9–11], leading to the proposition that proteins could drift randomly in and out of different, but equally functional forms of self-assembly[12].

The scale of variation in self-assembly within protein families is not well understood, because until recently self-assembly state was difficult to measure. In most cases, our knowledge derives from X-ray or cryo-electron microscopy structures that are usually only available for at best a handful of orthologs from often closely related species[13]. Techniques to measure self-assembly states in solution like small-angle X-ray scattering or analytical size exclusion chromatography have relatively low resolution and can struggle to separate interconverting assemblies[14]. High-resolution native mass spectrometry is still low throughput and has very specific buffer requirements[15]. As a consequence, we have little data on the extent of variation in self-assembly

state and its connection to biochemical function across whole protein families. Two recent studies retraced the evolution of self-assembly across the phylogeny of geranylgeranylglyceryl phosphate synthase[16] and a subfamily of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco)[17]. In both cases, novel stoichiometries first evolved without a major impact on function, and were subsequently lost several times, consistent with a largely neutral acquisition and loss of different types of self-assembly.

Here we extend this approach and trace the evolution of self-assembly across the phylogenetic history of an entire protein family using mass photometry – a novel, fast and single-particle based technique to quantify protein assemblies[18] at physiologically relevant, nanomolar concentrations. We discover a whole range of previously unknown assemblies that interconvert frequently, including enzymes that populate several asymmetrical assemblies concurrently. Our results add to a growing set of observations that self-assembly into higher-order oligomers is a remarkably plastic trait, even in proteins with highly conserved functions.

Phylogenetic classification of citrate synthases

We chose to investigate the diversity in self-assembly in bacterial citrate synthases (CS), because it has a highly conserved function: CS catalyzes the first step of the citric acid cycle. Previous studies identified two different structural types of CS; type I dimers[19–21] and type II hexamers[22–24]. The type II CS from *E. coli* is known to be allosterically regulated by NADH via a mechanism that relies on hexamerization[25]. Other hexameric type II CS are not affected by NADH[23,24,26,27]. These observations imply at least one transition in homo-oligomeric assembly plus a regulatory novelty associated with assembly, making CS an appropriate protein family to study conservation and functional relevance of quaternary structure evolution.

We inferred a comprehensive phylogenetic tree of bacterial CS enzymes (Fig. 1a, Supplementary Fig. 1). Rooting of bacterial phylogenies is difficult because of the amount of horizontal transfer events and there being no consensus on overall phylogenetic relationships of bacterial phyla[28,29]. It was also not possible to use Archaea for outgroup-rooting because the archaeal CS did not form a monophyletic group and we decided not to root the CS phylogeny. A parsimonious inference would locate the position of the root somewhere within type I CS, which contain all archaeal CS sequences as well as CS from several major bacterial phyla considered to differentiate early in different phylogenetic analyses[30–32]. Consistent with previous studies[27,33,34] our tree clearly separates type I and type II CS (Fig. 1a) but overall indicates a great amount of horizontal transfer events between bacteria, archaea and even eukaryotes. The non-mitochondrial CS from eukaryotes (nmCS) which is part of the glyoxylate cycle[35,36] branch within the bacterial type II CS, for example. Between type II and type I CS, our tree features a distinct clade of enzymes from marine Gammaproteobacteria, Nitrospirota and Cyanobacteria. Enzymes from these groups had not been structurally characterized before but a recent functional analysis had indicated that these form a distinct group of CS[37]. We term this clade type III CS. We have recently described one specific cyanobacterial CS from that clade in detail which assembles into octadecameric complexes (Fig. 1a, *S. elongatus*); details on its evolution and function are described elsewhere[38]. The mitochondrial CS from eukaryotes differ strongly in their sequences forming a very distantly related clade to the bacterial and archaeal CS and were not included in this study.

Guided by our phylogenetic tree, we characterized the quaternary assemblies formed by 40 CS enzymes at low nanomolar concentrations across different evolutionary groups and found a surprising diversity beyond the known dimeric and hexameric homo-oligomeric complexes (Fig. 1c).

Hexameric assembly evolved independently in type II and III CS

We first purified enzymes from type III CS and found most of them to form hexamers (Fig. 2a-b, Supplementary Fig. 2). The CS from a *Myxococcota* species only assembles into dimers and is found in the first clade that branches off within type III CS (Fig. 2a-b). We found two additional oligomeric state transitions within type III enzymes, both in the group of Cyanobacteria: one is an octadecameric, fractal-like CS, which we described elsewhere[38], and the other one is the loss of the hexameric assembly step in the enzyme of *Cyanobium sp. PCC 7001* (Extended Data Fig. 2a).

We solved a crystal structure for the novel hexameric type III CS from *M. sulfidovorans* which revealed assembly into dihedral ring of three dimers (Fig. 2c, Extended Data Table 1). It presented the same type of symmetry as the known hexamers of type II CS and uses broadly the same surface of the dimeric subcomplexes to interact and form complexes (Fig. 2d). Closer examination of the interfaces revealed that the interfacing sites and molecular interactions differ for hexamers found in type III compared to type II CS (Extended Data Fig. 2b). The hexameric complex of *E.coli* (type II CS) is stabilized mostly by van der Waals interactions including the type II-specific JK-loop [39] contacting the opposing dimer (Fig. 2d). This loop is not present in type III CS (Extended Data Fig. 1) and the hexamer from *M. sulfidovorans* instead relies on a cation–π interaction (W150 → R74, K76) and a salt bridge (D151 → R80, Fig. 2d). Exchanging residues of either of these polar interactions disrupts assembly into hexamers resulting in the formation of only dimers (Extended Data Fig. 2c). Many other type III CS apparently employ one or two salt-bridges at the same sites instead of the cation–π interaction found in *M. sulfidovorans* (Extended Data Fig. 2d) indicating a certain flexibility and evolutionary divergence of this interface.

We next asked if type II and III hexamers emerged from a common hexameric ancestor and then diverged substantially or if they represent a case of parallel evolution. To test this, we used ancestral sequence reconstruction (ASR) to resurrect ancient CS that represent the last common ancestor of the two groups ($anc_{2/3}$) as well as subsequent ancestors within type III CS (anc_{3a}, anc_{3b}, Fig. 2b, Supplementary Fig. 3). The reconstructed sequences were synthesized, heterologously produced and their oligomeric states characterized. $Anc_{2/3}$ does not have the interfacial J/K loop of type II hexamers and is also confidently inferred to lack the interfacial residues of type III hexamers (Supplementary Fig. 2). Consistent with this, it assembled into mostly dimers but we detected a distribution of oligomers within the sample (Fig. 2e). These included predominantly multimers of dimers (tetramers, hexamers, octamers etc.) and also low amounts of oddmers (trimers, pentamers etc.). The larger oligomers are stable and could be isolated via size exclusion chromatography (Extended Data Fig. 2e). We assigned these oligomeric distributions as polydisperse assemblies. Anc_{3a} revealed a similar behavior as $anc_{2/3,}$ again without a preference for an assembly into hexamers (Fig. 2e). Anc_{3a}'s descendant Anc_{3b} formed relatively stable hexamers (Fig. 2e). From this we concluded that the type III hexamers evolved in the interval between anc_{3a} and anc_{3b}. Consistent with this, *Myxococcota* sequences, which descend from anc_{3a} and are sister to all other type III CS, do not have the necessary residues to form the type III interface (Extended Data Fig. 1). Further, the CS we purified from one *Myxococcota* species only forms dimers (Fig. 2b). Overall, our results thus indicate that hexamers evolved independently twice, once within type III CS and once along the lineage to type II CS.

We next wanted to understand how type III hexamers evolved out of a polydisperse ensemble. To evolve to predominantly populate hexamers, anc_{3b} had to stabilize this particular stoichiometry at the expense of the other stoichiometries its predecessor anc_{3a} could populate. To

understand how this occurred, we introduced individual historical substitutions into anc_{3a} that are in the vicinity of the novel interface (Fig. 2f). Five amino acid substitutions were sufficient to create a CS that lost most of the polydisperse behavior and formed moderate amounts of hexamers (anc_{3a+5} Fig. 2g, Extended Data Fig. 2e). Three of the substitutions result in the creation of two salt bridges between dimers (q145D → p78K, k90E → 80R, Fig.2h, Extended Data Fig. 2f; small and capital letters indicate the ancestral and derived amino acid, respectively). The e79Y substitution introduces an aromatic side chain which is highly conserved in type III enzymes and probably stacks against the backbone of a conserved lysine (K86) (Fig. 2f; Extended Data Fig. 1). The last substitution introduced an intramolecular interaction within (a76R→152Q) which was apparently necessary to stabilize hexamers. Interestingly, in the CS from *M. sulfidovorans* this arginine is recruited into the cation–π interaction that stabilizes the hexameric interaction (R74, Fig. 2d, Extended Data Fig.1).

Together, these results show that hexamers evolved through introduction of salt bridges and other polar interactions between dimers. Salt bridges alone usually do not confer strong binding energy because of the desolvation costs of their charged moieties[40]. But they are known to introduce specificity into interactions[41,42]. A plausible mechanism is thus these 5 substitutions single out the hexamers from the pre-exiting distribution, whilst relying on prior affinity between dimers in the polydisperse ensemble.

Interface turnover in type II citrate synthases

We next investigated the history of hexamers in type II CS (Fig. 3a). All type II CS have the JK-loop that is part of the interface that holds together dimers into hexamers in published type II CS structures (Fig. 2c). Most of the enzymes we characterized from different clades within type II CS assemble into hexamers (Fig. 3a+b). Additionally, we identified two extant type II CS that

only formed dimers (Fig. 1a *F. taffensis*, Fig. 3b *C. woesi*). This oligomeric state had been observed for the type II CS of *F. tularensis* and *M. tuberculosis* before (Fig. 3a, white circles). These dimers represent four independent reversions from hexamers, which occurred in distant parts of the type II phylogeny (Fig. 3a).

Using ASR, we could not identify the exact time point when the hexameric assembly first emerged in type II CS. We found the last common ancestor of all type II CS (anc_2) to be polydisperse, even though it contains the characteristic J/K loop of type II hexamers as well as its binding-site on the opposing dimer. We then characterized anc_2's two daughter nodes in the type II phylogeny. One (anc_{2b}) assembled into weak hexamers but not the other (anc_{2a}, Fig. 3c). More recent ancestral proteins descending from anc_{2a} were also ambiguous regarding their assembly into hexameric complexes (Extended Data Fig. 3a-b). Taken at face value, this would indicate that the specific hexamers emerged multiple times within type II CS using the same residues. Unlike our earlier inference about the LCA of type II and III CS, this is very unparsimonious with respect to the stoichiometries of extant CS: we found hexamers all across type II CSs, including in each descendant lineage of all deep type II ancestors with all using the JK-loop as part of their interface (Extended Data Fig. 3c). The network of interactions in the interface of type II hexamers diverges relatively strongly in different phyla (Fig. 2d, Extended Data Fig. 4). This makes accurate reconstruction of this interface potentially challenging: mis-matched pairs of interacting residues could potentially destabilize the interface and a reason why we only found the hexameric assembly in anc_{2b}. We therefore think hexamerization likely emerged early in type II CS evolution and was subsequently lost at least four times resulting in dimeric enzymes in several extant species.

The origin of eukaryotic nmCS was phylogenetically inferred to be in type II CS. Purification of two nmCS enzymes revealed the typical hexameric assembly and corroborated this inference (*C. fasciculata* Fig. 3d, *P. olsenii* Fig. 1d). All nmCS also have the JK-loop that forms the interface between dimers in type II CS (Extended Data Fig. 3c+4) but not all homologs retained hexameric assembly: the enzyme from the unicellular algae *Chlamydomonas reinhardtii* reverted to forming only dimers (Fig. 3d). We discovered another novel structural type of CS within nmCS: the enzyme from the land plant *Ananas comosus* mainly assembled into octamers (Fig. 3d). We solved a medium-resolution cryo-EM structure for the complex and identified the octamers to be a dihedral ring of four dimers (Fig. 3e, Extended Data Table 2, Supplementary Fig. 4). The interactions between dimers that form the octamer are different from type II or III hexamers. The interaction is made by the C-terminus of the protein, which is considerably elongated in nmCS compared to type I-III CS (Fig. 1d). It folds from one dimer over to an adjacent one (Fig. 3f), with no other apparent interaction between dimers. When we deleted the C-terminal extension (Δ487-513, Fig. 3g) the CS protein from *A. comosus* could only assemble into dimers. This implies that the contacts that hold other type II CS into hexamers are completely lost in this protein and replaced by a new interaction through the C-terminus. To understand how this occurred, we inferred an AlphaFold-Multimer prediction for the hexameric nmCS of *C. fasciculata.* We found that its elongated C-terminus could already connect to the adjacent chain while still having the type II interface contacts to form hexamers (Extended Data Fig. 3d). The switch in interfaces therefore emerged via an evolutionary intermediate that had two interfaces: One via the C-terminus and one via the JK loop. This redundancy then allowed the old interface to be lost somewhere in the Embryophytes, leading to the to the evolution of octamers held together exclusively by their C termini.

Assembly into a polydisperse distribution is an ancient trait of CS

We observed CS to assemble into a polydisperse distribution of oligomeric states in the ancestors of both type II (anc_2) and type III CS (anc_3) as well as their common ancestor ($anc_{2/3}$). To understand the evolutionary origin of this behavior and if it could be a transition state from dimeric to hexameric quaternary structure, we resurrected and characterized the ancestors of different clades of type I CS (Fig. 4a). All of these ancestral CS were also polydisperse ($anc_{1a\text{-}c}$, Fig. 4b) indicating that this type of assembly is probably very ancient since these represent precursors of phyla that are thought to have diverged early in microbial evolution[30,31]. To confirm that is not an artefact arising from incorrectly reconstructed sequences, we asked if it is also found in any early branching type I CS. We indeed found multiple extant type I CS that branch off early in their respective clades closely to these ancestors and also populate polydisperse ensembles (Fig. 4c, Extended Data Fig. 5a).

We next sought to better understand the organization of these oligomeric ensembles to test if they represent structural precursors out of which type II and type III hexamers emerged. Because the polydisperse nature of these proteins makes high-resolution structure determination challenging, we separated them by size via SEC to enrich for large stoichiometries. We were able to enrich anc_{1b} for hexamers but still found the sample to be highly heterogeneous when analyzing it via cryo-EM. This hindered obtaining high-resolution structural insights into this assembly but we could retrieve 2D class averages that appear to correspond to hexameric assemblies (Extended Data Fig. 5b-d). The dimeric subcomplexes apparently loosely connect to each other without forming closed circular assemblies, resulting in the formation of apparently non-symmetric higher-order oligomers.

Our data thus imply that the most ancient forms of CS were capable of forming polydisperse ensembles of potentially asymmetric larger order complexes. Most type I CS then lost this ability

and became monodisperse dimers. Similarly, type II and type III CS also became monodisperse, but settled on two different types of hexamers instead of dimers. We presently have no explanation for this repeated evolution of more monodisperse assemblies, but it may indicate either mutational or selective pressure (or a combination of both) towards monodispersity.

Oligomeric state transitions lead to minor changes in catalytic parameters

We have shown that there are many different quaternary structures that CS can assemble into, but all of these proteins catalyze the same enzymatic reaction. We therefore examined if changes in oligomeric state are connected to differences in catalytic parameters, which selection could in principle act on. First, we characterized the catalytic activity of the different oligomeric forms found within polydisperse CS. Here all complexes are built from the same sequence. We separated the polydisperse complexes of the CS from *D. pimensis* via SEC and isolated fractions with different oligomeric compositions (Fig. 5a, extended Data Fig. 5a). The fractions varied from almost exclusively dimeric to a majority of complexes with ≥ 8 subunits and their assemblies did not re-equilibrate for several hours. We then measured Michaelis-Menten-kinetics of the individual fractions. This revealed that all complexes are catalytically competent enzymes. There is a reduction in the turnover number k_{cat} for larger oligomeric complexes compared to the dimers (Fig. 5b), which may derive from steric hindrance of important catalytic motions. The K_m-values in contrast are very similar for all measured fractions (Extended Data Table 3). The results illustrate that different oligomeric species can have somewhat different catalytic properties, even if their sequences are identical. The catalytic parameters of the full equilibrium of the protein when not enriched for larger stoichiometries (Dp) was, however, very similar to the one of pure dimers. An evolutionary transition from polydisperse to dimeric state therefore

might not be visible to selection in this case, because the fraction of active sites in the larger oligomers is small enough not to have a major effect.

In another case, different oligomeric states appear not to differ in their catalytic activity: We discovered that we could shift the oligomeric state of the CS from *A. comosus* from octamers to roughly similar amounts hexamers and octamers with high quantities of acetyl-CoA (but not oxaloacetate) (Fig. 5c). In principle, this could mean that hexamers are the catalytically competent form. Further experiments contradict this: We compared the Michaelis-Menten-kinetics of the *A. comosus* Δ487-513 variant, which can only form dimers, to that of the wild-type enzyme and found them to be extremely similar (Fig. 5d, kinetic parameters for oxaloacetate were measured at saturated concentrations of 500 µM acetyl-CoA; extended Data Fig. 6a). It thus appears that this CS has evolved substrate-dependent oligomeric state changes that are not required for catalysis, though they could plausibly be important for allosteric regulations by small molecules, which we did not investigate here.

We next asked if the evolution of a new interface via changes in the sequence had effects on CS kinetic parameters. We decided to investigate the transition within type III enzymes from dimers to hexamers because here we have two ancestral enzymes bracketing this transition (anc_{3a}, anc_{3b}). We found the hexameric anc_{3b} to have a slightly lower k_{cat} and a reduced Km-value compared to the polydisperse anc_{3a} (Extended Data Fig. 6b, Extended Data Table 3). Both effects are very small compared to the variation in enzymatic parameters for bacterial CS[43]. We next asked if this change is causally linked to the substitutions that caused hexamers to evolve. We characterized the variant of anc_{3a} in which we transplanted five historical substitutions that were sufficient to trigger hexamer formation ($anc_{3a+5subs}$, Fig. 2f). In this variant roughly 45% of active sites reside in hexamers at assay concentrations (Fig. 2g). The k_{cat} of this variant was more than

twofold higher than both of the ancestors and the K_m-value was also elevated (Extended Data Fig. 6a, Extended Data Table 3). The mild decrease in activity between anc_{3a} and anc_{3b} is therefore likely not linked to the novel interface but to some of the other 24 substitutions that occurred in that interval. This demonstrates that a transition in quaternary structure can be accompanied immediately by a change in catalytic activity (in this case an increase in k_{cat}). But in this instance, this effect was almost immediately reversed by additional substitutions outside of the interface. In addition, all three enzymes (anc_{3a}, anc_{3+5}, anc_{3b}) have very similar rates of catalysis at substrate concentrations below < 25 µM substrate. If intracellular substrate concentrations are low, which for oxaloacetate is usually expected due to its instability[44], the changes in maximum turnover number (k_{cat}) would likely not be visible to selection.

Lastly, we investigated if the loss of an assembly step causes a change in catalytic activity. This may be the case if interfaces have accumulated substitutions that are harmless when the complex can assemble, but highly deleterious when it is prevented from forming. Such 'entrenchment' for example occurs if interfaces become so hydrophobic that they need to be shielded from solvent to prevent aggregation[45–47]. To test for this, we abolished interfaces of an extant oligomeric complex by mutation. We measured kinetic parameters of the type III hexamer from *M. sulfidovorans* and its dimeric interface mutant (W150A, Extended Data Fig. 2b) and found them to be very similar (Fig. 5f). This implies that the interface we abolished here has not accumulated entrenching substitutions. Together with the minor effect on catalysis that assembly appears to have, this may explain the relatively frequent losses of higher-order assembly we observed on our phylogeny.

Many different CS can substitute for an allosterically regulated hexamer CS in *E. coli*

Last we investigated the ability of different types and oligomeric forms of CS to complement its native function in vivo. We chose *E. coli* as host which natively holds a hexameric type II CS, that can be allosterically inhibited by NADH.[25] . To do this we constructed a scarless knock-out (KO) strain which lacked the native CS gene (*E. coli* BL21(DE3) ΔgltA). This strain can grow on complex media but is not viable if only glucose or acetate are available as sole carbon source because of the dependency on CS activity. We transformed this strain with plasmids encoding CS from different extant organisms or ancestral CS and recorded growth curves in minimal media (M9) supplemented with glucose (Supplementary Data Fig. 5). We found that CS function in *E. coli* can be complemented by most of our enzymes to some degree and that there is no obvious correlation to type or quaternary structure (Fig. 5g). All polydisperse and many dimeric type I CS produced growth rates comparable to complementation with the native (Fig. 5g). For type II and III CS, some complement well but others do not – but there is no indication that hexameric CS in general or specifically type II hexamers perform better in *E. coli* than other types of CS. All of the characterized nmCS complement poorly but this is likely connected to a low level of protein production (Extended Data Fig. 7a). In general, we found that the protein production of the different enzymes varied very strongly and that the CS that do not complement in growth curve experiments are often not efficiently produced (Extended Data Fig. 7a). If we increase the expression by addition of an inducer (IPTG) eventually all of the tested CS can promote growth of the KO-strain at least on solid media (Fig. 4h, Extended Data Fig. 7b). This demonstrates that all tested CS retain the ability to rescue CS function in *E. coli* independently of primary or quaternary structure at least under a strong selection in a laboratory setting. Self-assembly type is thus not a strong determinant of the adaptive value of any particular CS to *E.coli* even though its native CS is an allosterically regulatable hexamer.

Discussion

Here we have shown that even within the evolutionary history of only one protein family there is remarkable diversity in the assembly into homo-oligomeric complexes. This echoes previous findings from Form II Rubiscos, which also transitions between different oligomeric states[17]. In particular, we have discovered several types of assemblies that are unlikely to be well represented in the protein databank[48]: enzymes that can adopt multiple interconverting and likely asymmetric assemblies; complexes that form remarkable fractal-like structures[38], and one oligomer held together by very flexible terminal interactions. In addition, all these assembly types (with perhaps the exception of the octamer) have been lost at least once and in several cases multiple times. Our discovery rate of oligomeric state transitions was remarkably high for such a conserved enzyme, and it is entirely possible that more surprises lurk in the CS phylogeny.

Some generalizable principles emerge: First, all assemblies we have structures of are either dimers or rings of dimers, in which the active site is located on the outward-facing side of the dimer. This almost certainly reflects the constraint that any oligomeric assembly should not block access to the active site. A similar argument may also explain another bias: CS appears to have only evolved ring-like assemblies of at least three dimers. We never observed dihedral dimers of dimers. We can only speculate that such arrangements, which are very common in other families[49,50], never evolved in CS because they in some way sterically hinder CS catalytic motion[51,52], though sheer coincidence remains a plausible explanation. A third conspicuous pattern is the frequent loss of higher-order assembly. This implies that in many cases the interfaces that hold such assemblies together are not entrenched. Structurally, the reason may be

that these interfaces are in general quite small and relatively hydrophilic, especially when compared to larger and very hydrophobic dimer interface in CS[47].

Perhaps our most surprising discovery is that multiple CS on our tree can populate several stoichiometries at once and that this trait has been lost multiple times in favor of more monodisperse assemblies. We at present have no good explanation for why this is the case. One hypothesis would be that sequences of monodisperse quaternary structures are evolutionary more readily available than the ones of the polydisperse enzymes. Polydisperse assembly is not known to be common in other protein families[53], but that may result from a discovery bias: Most traditional techniques for measuring quaternary assemblies struggle to resolve such ensembles. In either case, these assemblies seem to influence kinetics only mildly and may thus represent a form of harm- and useless complexity.

Is all this stoichiometric variation functionally meaningless? This is difficult to answer without detailed experiments in the organisms that host these enzymes. We did this recently for the fractal-like CS from the cyanobacterium *S.elongatus* and could not detect any functional advantage of this bizarre assembly over a simpler enzyme in vivo[38]. We thus think it at least plausible that some of these stoichiometries have no particular function. On the other hand, even initially useless evolutionary inventions can be the seed of future adaptive functions that are built on top of them. One example is likely the allosteric regulation via NADH in type II CS, which is found only in a subset of type II hexamers. Overall, we suspect that many proteins and specifically enzymes can wander quite freely through the space of possible multimeric states. This observation agrees with recent computational inferences[54] and raises the question of what our null hypothesis should be when we discover a new type of quaternary assembly. Based on the work presented here, we advocate to assume it is useless until it is proven not to be.

References

1. Goodsell, D. S. & Olson, A. J. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153 (2000).
2. Levy, E. D. & Teichmann, S. Structural, evolutionary, and assembly principles of protein oligomerization. in *Progress in Molecular Biology and Translational Science* vol. 117 25–51 (Elsevier B.V., 2013).
3. Matthews, J. M. & Sunde, M. Dimers, oligomers, everywhere. *Advances in Experimental Medicine and Biology* vol. 747 1–18 at https://doi.org/10.1007/978-1-4614-3229-6_1 (2012).
4. Perica, T. *et al.* The emergence of protein complexes: quaternary structure, dynamics and allostery. *Biochem. Soc. Trans.* **40**, 475–491 (2012).
5. Marianayagam NJ, Sunde M, M. J. The power of two: protein dimerization in biology. *Trends Biochem. Sci.* **29**, 618–625 (2004).
6. Bershtein, S., Wu, W. & Shakhnovich, E. I. Soluble oligomerization provides a beneficial fitness effect on destabilizing mutations. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 4857–4862 (2012).
7. Li, Y. *et al.* Oligomeric interactions maintain active-site structure in a noncooperative enzyme family. *EMBO J.* (2022) doi:10.15252/embj.2021108368.
8. Kleiner, D. *et al.* Evolution of homo-oligomerization of methionine S-adenosyltransferases is replete with structure–function constrains. *Protein Sci.* **31**, e4352 (2022).
9. Hashimoto, K., Madej, T., Bryant, S. H. & Panchenko, A. R. Functional states of homooligomers: Insights from the evolution of glycosyltransferases. *J. Mol. Biol.* **399**, 196–206 (2010).
10. Lansky, S. *et al.* A pentameric TRPV3 channel with a dilated pore. *Nature* **621**, 206–214 (2023).
11. Naidoo, N. *et al.* Crystal Structure of Lsm3 Octamer from Saccharomyces cerevisiae: Implications for Lsm Ring Organisation and Recruitment. *J. Mol. Biol.* **377**, 1357–1371 (2008).
12. Lynch, M. Evolutionary diversification of the multimeric states of proteins. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 2821–2828 (2013).
13. Dey, S., Ritchie, D. W. & Levy, E. D. PDB-wide identification of biological assemblies from conserved quaternary structure geometry. *Nat. Methods* **15**, 67–72 (2018).
14. Marciano, S. *et al.* Protein quaternary structures in solution are a mixture of multiple forms. *Chem. Sci.* **13**, 11680–11695 (2022).
15. Barth, M. & Schmidt, C. Native mass spectrometry—A valuable tool in structural biology. *J. Mass Spectrom.* **55**, e4578 (2020).

16. Kropp, C., Straub, K., Linde, M. & Babinger, P. Hexamerization and thermostability emerged very early during geranylgeranylglyceryl phosphate synthase evolution. *Protein Sci.* **30**, 583–596 (2021).

17. Liu, A. K. *et al.* Structural plasticity enables evolution and innovation of RuBisCO assemblies. *Sci. Adv.* **8**, (2022).

18. Young, G. *et al.* Quantitative mass imaging of single biological macromolecules. *Science* **360**, 423–427 (2018).

19. Russell, R. J., Hough, D. W., Danson, M. J. & Taylor, G. L. The crystal structure of citrate synthase from the thermophilic archaeon, Thermoplasma acidophilum. *Structure* **2**, 1157–1167 (1994).

20. Russell, R. J. M., Gerike, U., Danson, M. J., Hough, D. W. & Taylor, G. L. Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium. *Structure* **6**, 351–361 (1998).

21. McEvily, A. J. & Harrison, J. H. Subunit equilibria of porcine heart citrate synthase. Effects of enzyme concentration, pH, and substrates. *J. Biol. Chem.* **261**, 2593–2598 (1986).

22. Nguyen, N. T. *et al.* Comparative Analysis of Folding and Substrate Binding Sites between Regulated Hexameric Type II Citrate Synthases and Unregulated Dimeric Type I Enzymes,. *Biochemistry* **40**, 13177–13187 (2001).

23. Sievers, M., Stöckli, M. & Teuber, M. Purification and properties of citrate synthase from Acetobacter europaeus. *FEMS Microbiol. Lett.* **146**, 53–58 (1997).

24. Park, S.-H. *et al.* Structural basis of the cooperative activation of type II citrate synthase (HyCS) from Hymenobacter sp. PAMC 26554. *Int. J. Biol. Macromol.* **183**, 213–221 (2021).

25. Stokell, D. J. *et al.* Probing the Roles of Key Residues in the Unique Regulatory NADH Binding Site of Type II Citrate Synthase of Escherichia coli. *J. Biol. Chem.* **278**, 35435–35443 (2003).

26. Francois, J. A. *et al.* Structure of a NADH-insensitive hexameric citrate synthase that resists acid inactivation. *Biochemistry* **45**, 13487–13499 (2006).

27. Weitzman, P. D. J. & Jones, D. Regulation of Citrate Synthase and Microbial Taxonomy. *Nat. 1968 2195151* **219**, 270–272 (1968).

28. Galtier, N. A Model of Horizontal Gene Transfer and the Bacterial Phylogeny Problem. *Syst. Biol.* **56**, 633–642 (2007).

29. Arnold, B. J., Huang, I. T. & Hanage, W. P. Horizontal gene transfer and adaptive evolution in bacteria. *Nat. Rev. Microbiol. 2021 204* **20**, 206–218 (2021).

30. Hug, L. A. *et al.* A new view of the tree of life. *Nat. Microbiol. 2016 15* **1**, 1–6 (2016).

31. Khaledian, E., Brayton, K. A. & Broschat, S. L. A Systematic Approach to Bacterial Phylogeny Using Order Level Sampling and Identification of HGT Using Network Science. *Microorg. 2020, Vol. 8, Page 312* **8**, 312 (2020).

32. Zhu, Q. *et al.* Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. *Nat. Commun. 2019 101* **10**, 1–14 (2019).

33. Mukai, A. & Endoh, H. Presence of a bacterial-like citrate synthase gene in Tetrahymena thermophila: Recent lateral gene transfers (LGT) or multiple gene losses subsequent to a single ancient LGT? *J. Mol. Evol.* **58**, 540–549 (2004).

34. Schnarrenberger, C. & Martin, W. Evolution of the enzymes of the citric acid cycle and the glyoxylate cycle of higher plants. *Eur. J. Biochem.* **269**, 868–883 (2002).

35. Kondrashov, F. A., Koonin, E. V., Morgunov, I. G., Finogenova, T. V. & Kondrashova, M. N. Evolution of glyoxylate cycle enzymes in Metazoa: evidence of multiple horizontal transfer events and pseudogene formation. *Biol. Direct* **1**, 31 (2006).

36. Kunze, M., Pracharoenwattana, I., Smith, S. M. & Hartig, A. A central role for the peroxisomal membrane in glyoxylate cycle function. *Biochim. Biophys. Acta - Mol. Cell Res.* **1763**, 1441–1452 (2006).

37. Ito, S., Koyama, N. & Osanai, T. Citrate synthase from Synechocystis is a distinct class of bacterial citrate synthase. *Sci. Rep.* **9**, 6038 (2019).

38. Manuscript found in Supplementary Data. Article currently in review - Emergence of fractal geometries in the evolution of a metabolic enzyme.

39. Nguyen, N. T. *et al.* Comparative Analysis of Folding and Substrate Binding Sites between Regulated Hexameric Type II Citrate Synthases and Unregulated Dimeric Type I Enzymes,. *Biochemistry* **40**, 13177–13187 (2001).

40. Hendsch, Z. S. & Tidor, B. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci.* **3**, 211–226 (1994).

41. Donald, J. E., Kulp, D. W. & DeGrado, W. F. Salt Bridges: Geometrically Specific, Designable Interactions. *Proteins* **79**, 898 (2011).

42. Fersht, A. R. *et al.* Hydrogen bonding and biological specificity analysed by protein engineering. *Nat. 1985 3146008* **314**, 235–238 (1985).

43. Chang, A. *et al.* BRENDA, the ELIXIR core data resource in 2021: New developments and updates. *Nucleic Acids Res.* **49**, D498–D508 (2021).

44. Krebs, H. A. The effect of inorganic salts on the ketone decomposition of oxaloacetic acid. *Biochem. J.* **36**, 303 (1942).

45. Hochberg, G. K. A. *et al.* A hydrophobic ratchet entrenches molecular complexes. *Nature* (2020) doi:10.1038/s41586-020-3021-2.

46. Finnigan, G. C., Hanson-smith, V., Stevens, T. H. & Thornton, J. W. Evolution of increased complexity in a molecular machine. *Nature* (2012) doi:10.1038/nature10724.

47. Abrusán, G. & Foguet, C. An Assessment of Quaternary Structure Functionality in Homomer Protein Complexes. *Mol. Biol. Evol.* **40**, (2023).

48. Ahnert, S. E., Marsh, J. A., Hernández, H., Robinson, C. V & Teichmann, S. A. Principles of assembly reveal a periodic table of protein complexes. *Science (80-.).* **350**, aaa2245

(2015).

49. Levy, E. D., Erba, E. B., Robinson, C. V. & Teichmann, S. A. Assembly reflects evolution of protein complexes. *Nat. 2008 4537199* **453**, 1262–1265 (2008).

50. Powers, E. T. & Powers, D. L. A Perspective on Mechanisms of Protein Tetramer Formation. *Biophys. J.* **85**, 3587 (2003).

51. Wiegand, G. & Remington, S. J. Citrate synthase: structure, control, and mechanism. *Annu. Rev. Biophys. Biophys. Chem.* **15**, 97–117 (1986).

52. Kanamori, E., Kawaguchi, S.-I., Kuramitsu, S., Kouyama, T. & Murakami, M. Structural comparison between the open and closed forms of citrate synthase from Thermus thermophilus HB8. *Biophys. physicobiology* **12**, 47–56 (2015).

53. Aquilina, J. A., Benesch, J. L. P., Bateman, O. A., Slingsby, C. & Robinson, C. V. Polydispersity of a mammalian chaperone: Mass spectrometry reveals the population of oligomers in αB-crystallin. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 10611–10616 (2003).

54. Schweke, H. *et al.* An atlas of protein homo-oligomerization across domains of life. *bioRxiv* 2023.06.09.544317 (2023) doi:10.1101/2023.06.09.544317.

55. Yang, J., Zhang, Z., Zhang, X. A. & Luo, Q. A ligation-independent cloning method using nicking DNA endonuclease. *Biotechniques* **49**, 817–821 (2010).

56. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

57. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

58. Le, S. Q. & Gascuel, O. An Improved General Amino Acid Replacement Matrix. *Mol. Biol. Evol.* **25**, 1307–1320 (2008).

59. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).

60. Anisimova, M. & Gascuel, O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* **55**, 539–552 (2006).

61. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

62. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, (2014).

63. Ellman, G. L. Tissue sulfhydryl groups. *Arch. Biochem. Biophys.* **82**, 70–77 (1959).

64. Srere, P. A., Brazil, H., Gonen, L. & Takahashi, M. The Citrate Condensing Enzyme of Pigeon Breast Muscle and Moth Flight Muscle. *Acta Chem. Scand.* **17 supl.**, 129–134 (1963).

65. Oscarsson, M. *et al.* MXCuBE2: the dawn of MXCuBE Collaboration. *urn:issn:1600-5775* **26**, 393–405 (2019).

66. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *urn:issn:0907-4449* **66**, 133–144 (2010).

67. McCoy, A. J. & IUCr. Solving structures of protein complexes by molecular replacement with Phaser. *urn:issn:0907-4449* **63**, 32–41 (2006).

68. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).

69. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

70. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. Sect. D, Struct. Biol.* **75**, 861–877 (2019).

71. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

72. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

73. Bepler, T., Kelley, K., Noble, A. J. & Berger, B. Topaz-Denoise: general deep denoising models for cryoEM and cryoET. *Nat. Commun.* **11**, (2020).

74. Zi Tan, Y. *et al.* Addressing preferred specimen orientation in single-particle cryo-EMthrough tilting. *Nat. Methods* **14**, 793–796 (2017).

75. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).

76. Reisch, C. R. & Prather, K. L. J. The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in Escherichia coli. *Sci. Reports 2015 51* **5**, 1–12 (2015).

77. Baranyi, J. & Roberts, T. A. A dynamic approach to predicting bacterial growth in food. *Int. J. Food Microbiol.* **23**, 277–294 (1994).

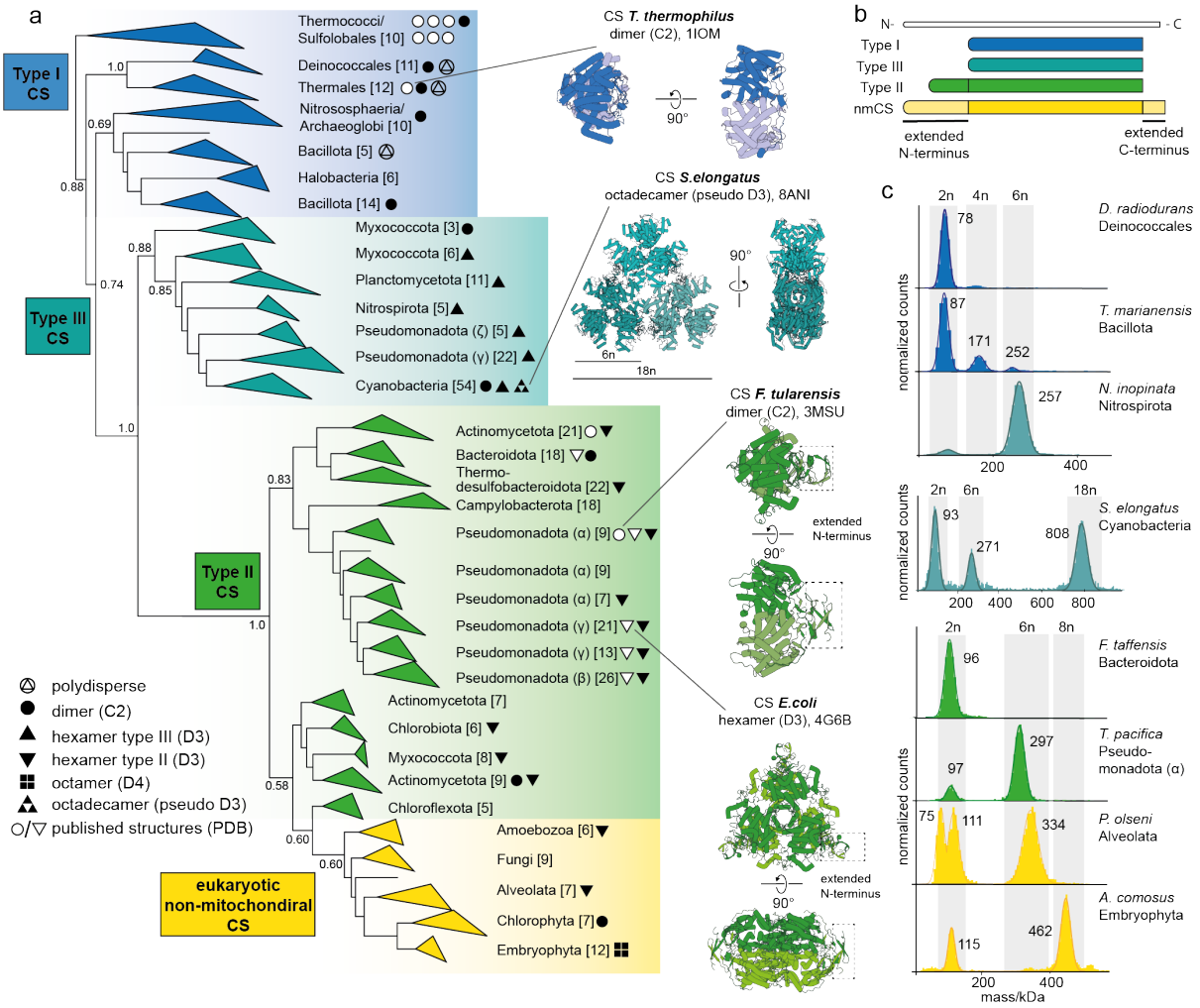


Fig. 1. Assembly of CS across the phylogenetic tree (a) Phylogenetic tree of CS in Bacteria, Archaea and eukaryotic nmCS (full phylogeny in supplementary Fig. 1) and classification into type I, II, III, and nmCS. Brackets indicate the number of sequences within each clade. Branch supports values are shown for important nodes as Felsenstein's bootstrap values. The transfer into Eukaryotes is well supported by phylogenetic and experimental results (see below) but the branching order of major eukaryotic lineages is poorly supported and disagrees with known relationships. Symbols (●, ▲, ▼etc.) indicate the quaternary structure of characterized CS within the respective clade: White symbols represent solved structures that have been deposited to the PDB and black symbols correspond to assemblies that were characterized in this study. Representative structures of known CS assemblies are shown. (b) Cartoon representation of the

amino acid sequence structure of type I-III and nmCS. (c) Mass photometry (MP) measurements of purified CS displaying different forms of assembly.

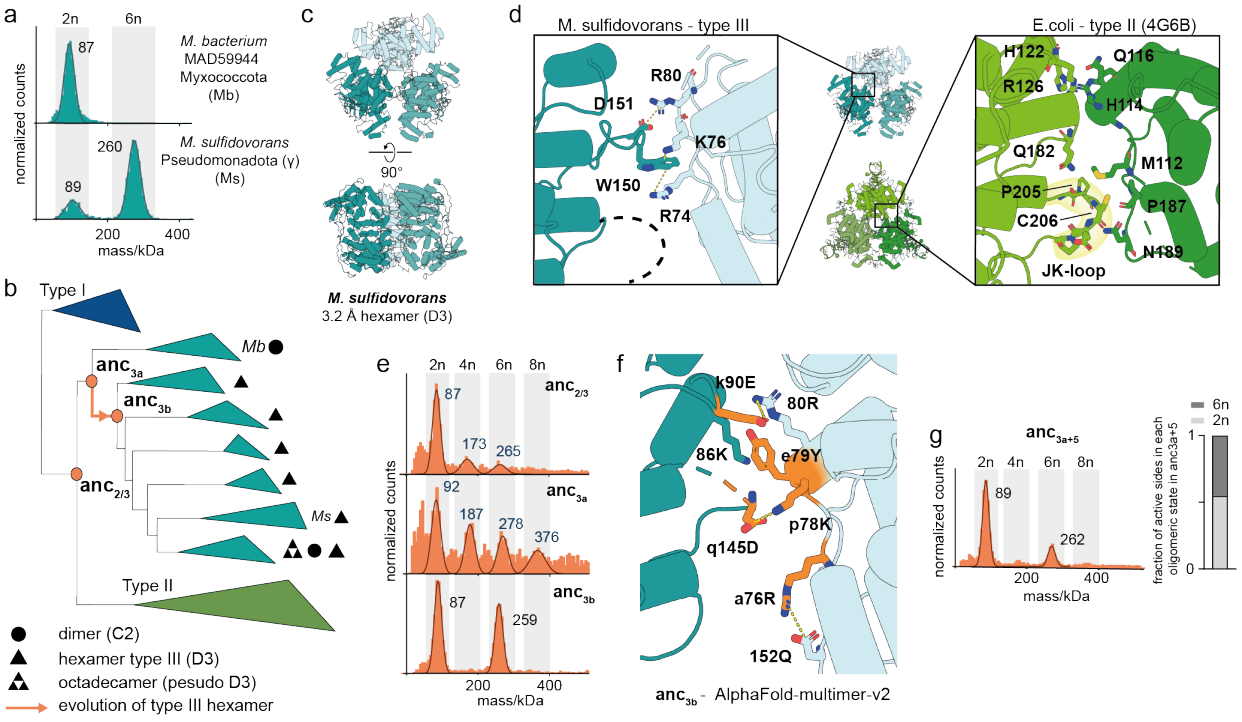


Fig. 2. Parallel evolution of hexameric CS (a) MP measurements of two type III CS. (b) Schematic representation of the CS phylogeny displaying the quaternary structures of characterized type III CS. All MP spectra and species names for the characterized quaternary structures are found in supplementary Fig 1+2. Nodes corresponding to resurrected ancestral CS are indicated. (c) X-ray structure of hexameric type III CS from *M. sulfidovorans*. (d) Comparison of the interface area that connects dimers into hexamers in the type III structure and the type II CS from *E. coli.* (e) MP measurements of resurrected ancestral CS along the evolutionary trajectory of hexameric assembly within type III enzymes. (f) Location of historical substitutions (orange sites) within the type III interface in a modelled structure of anc_{3b} using AlphaFold-multimer-v2. (g) MP measurement a variant of anc_{3a} with a set of 5 historical substitutions shown in (f) that are sufficient to trigger formation of hexamers and loss of

polydisperse behavior (anc_{3+5}). Bar graph displays the fraction of all active sites in dimers vs hexamers for anc_{3+5}.

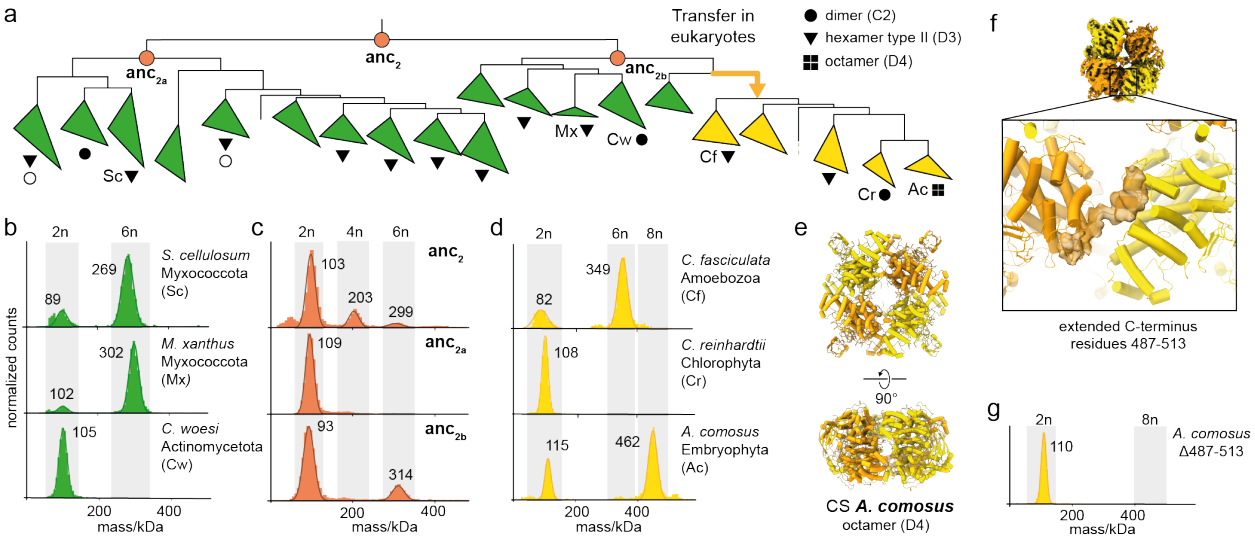


Fig. 3 Widespread hexameric assembly in type II CS and evolution of novel octamers (a) Schematic representation of part of the CS phylogeny displaying the quaternary structures of characterized type II CS and nmCS. All MP spectra and species names for the characterized quaternary structures are found in supplementary Fig. 1 and 2. Nodes corresponding to resurrected ancestral CS are indicated. (b-d) MP measurements of extant type II CS (b), ancestral type II CS (c) and nmCS (d). (e) Cryo-EM structure of octameric nmCS from *A. comosus*. (f) Focus on the extended C-terminus of nmCS within the Cryo-EM density of *A. comosus* which folds over from one dimeric subcomplex to an adjacent one. (g) MP measurement of a variant of *A. comosus* CS in which the extended C-terminus was cut off (Δ487-513).

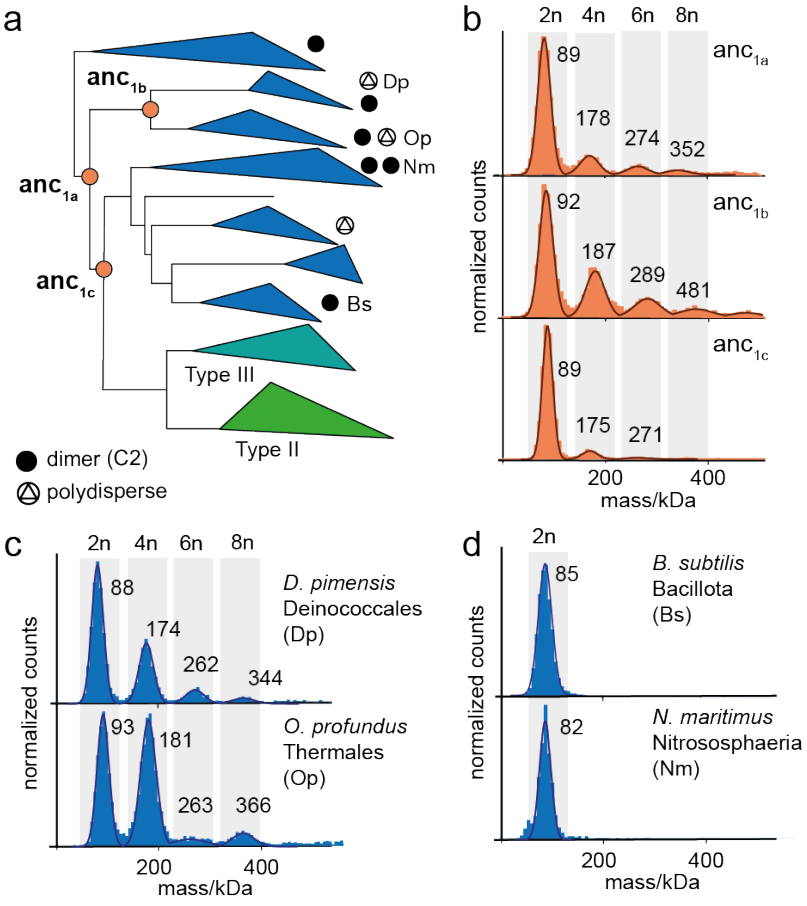


Fig. 4 Polydisperse assemblies early in the evolution of CS (a) Schematic representation of the CS phylogeny displaying the quaternary structures of characterized type I CS. All MP spectra and species names for the characterized quaternary structures are found in supplementary Fig. 1 and 2. Nodes corresponding to resurrected ancestral CS are indicated. (b-d) MP measurements of ancestral CS (b), dimeric (c) and polydisperse (d) type I CS.

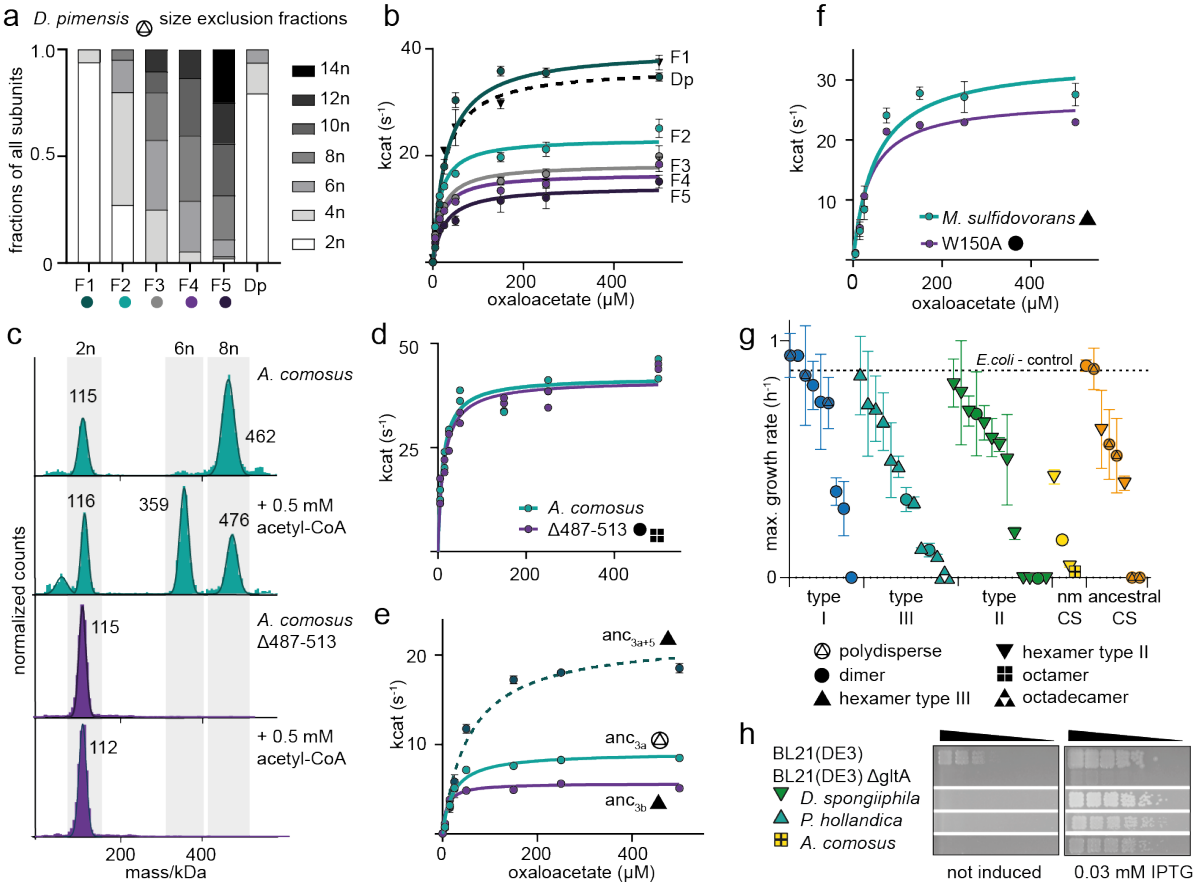


Fig. 5 Effects on catalytic function by changes in quaternary structure (a) Fraction of CS subunits within the different oligomeric states for each SEC-fraction determined by MP for the CS from *D. pimensis* (Dp) (F1-F5, see also Extended Data Fig. 5a). (b) Michaelis-Menten kinetics of the different fractions shown in (a) and the unseparated sample of *D. pimensis* CS (Dp). Error bars = SD, n = 3 technical replicates. (c) MP measurements of CS from *A. comosus* and its truncation variant Δ487-513 in the absence and presence of acetyl-CoA. (d) Michaelis-Menten kinetics of CS from *A. comosus* and its Δ487-513 variant. Error bars = SD, n=2 technical replicates. (e) Michaelis-Menten kinetics of ancestral CS bracketing the emergence of hexamers within type III enzymes (anc_{3a}, anc_{3b}) and the minimal substitution construct to yield hexameric complexes (anc_{3a+5}). Error bars = SD, n = 3 technical replicates. (f) Michaelis-Menten kinetics of extant CS from *M. sulfidovorans* and a variant that disrupts the interface between dimers (W150A). Error bars = SD, n = 3 technical replicates. (g) Maximum growth rates on M9 media

of an *E. coli* strain lacking the native CS gene (BL21(DE3) ΔgltA) complemented with a plasmid encoding for CS genes from all characterized extant and ancestral enzyme, error bars = SD, n = 3 biological replicates. The positive control is the KO-strain complemented with the native CS from *E. coli* which displayed the same growth behaviour as the wild-type BL21(DE3) strain. (h) Spot assays on solid M9 media using either leaky expression or induction by IPTG, shown for strains that did not complement in (g), showing that increased CS production leads to complementation. Cultures were spotted in a five-step serial dilution using a ratio of 1:5 for each step and incubated on M9-solid media supplemented with glucose and different IPTG concentrations. One representative plate is shown for each experiment, out of a total of three replicates for each plate. Full data for all strains shown in Extended Data Fig. 7.

Methods

Molecular cloning

The genes encoding the CS from *Bacillus subtilis*, *Escherichia coli* and *Myxococcus xanthus* were amplified from genomic DNA by PCR (Q5® High-Fidelity 2X Master Mix, New England Biolabs, USA-MA) and introduced into the pLIC expression vector[55] by Gibson cloning (Gibson Assembly Master Mix, New England Biolabs, USA-MA). All other extant and ancestral CS sequences were obtained as gene fragments from Twist Bioscience (USA-CA) or Integrated DNA Technologies (USA-IA) and introduced in the same expression vector by Gibson cloning. All CS sequences were tagged with a C-terminal polyhistidine-tag for purification (tag-sequence: LE-HHHHHH-Stop). For single-site mutants and deletions of the CS-sequences the KLD Enzyme Mix (New England Biolabs, US-MA) was used. Mutagenesis primers were designed with NEBasechanger (nebasechanger.neb.com) and used to PCR-amplify the vector encoding for the gene that was to be changed. Resulting PCR products were added to the KLD enzyme mix and subsequently transformed. All cloned genes were verified by Sanger-sequencing (Microsynth, Germany).

The DNA sequences of all purified proteins and the NCBI identifier of all extant sequences can be found in Supplementary Table S1.

Protein purification

For heterologous overexpression the vectors with the gene of interest were transformed into chemically competent *E. coli* BL21 (DE3) cells. Transformed colonies were used to inoculate expression cultures (500 mL) made from LB-medium supplemented with 12.5 g/L lactose (Fisher chemical, USA-MA). The cultures were incubated overnight at 30 °C and 200 rpm. Cells were harvested by centrifugation (4.500 g, 15 min, 4 °C), resuspended in Buffer A (20 mM Tris,

300 mM NaCl, 20 mM imidazole, pH 8) and freshly supplemented with DNAse I (3 Units/µL, Applichem GmbH, Germany). Cells were disrupted using a Microfluidizer® (Microfluidics International Corporation, USA-MA) in three cycles at 15.000 psi and centrifuged to spin down cell debris and aggregates (30.000 g, 30 min, 4 °C). The clarified lysate was loaded with a peristaltic pump (Hei-FLOW 06, Heidolph, Germany) on prepacked nickel-NTA columns (5 mL Nuvia IMAC Ni-Charged, Biorad, USA-CA) that were equilibrated before with buffer A. The loaded column was first washed with buffer A for seven column volumes and then with 10% (v/v) buffer B (20 mM Tris, 300 mM NaCl, 500 mM imidazole, pH 8) in buffer A for seven column volumes. The bound protein was eluted with buffer B and either buffer exchanged with PD-10 desalting columns (Cytiva, USA-MA) into PBS or 20 mM Tris, 200 mM NaCl, pH 7.5 or further purified by size-exclusion chromatography (SEC). For SEC the protein was injected on an ENrich SEC 650 column (Biorad, USA-CA) with PBS as running buffer using a NGC Chromatography System (Biorad, USA-CA). To separate the different oligomers of polydisperse assemblies fractions of 250 µL were collected into a 96-well plate and analyzed via MP afterwards. Purity of the proteins was analyzed by SDS-PAGE. After either buffer exchange or SEC, the purified proteins were flash-frozen with liquid nitrogen and stored at -20 °C before further use.

Phylogenetic analysis and ancestral sequence reconstruction

Amino acid sequences of 418 CS genes from bacteria, archaea and eukaryotes were collected from the NCBI Reference Sequence Database and aligned via MUSCLE v3.8.31[56]. The maximum likelihood (ML) phylogeny was inferred from the multiple sequence alignment (MSA) using raxML v8.2.10[57]. The LG substitution matrix[58] was used as determined by automatic best-fit model selection as well as fixed base frequencies and a gamma-model of rate heterogeneity.

The robustness of the ML tree topology was assessed by inferring 100 non-parametric bootstrap trees with raxML, from which Felsenstein's and transfer bootstrap values were derived using BOOSTER (booster.pasteur.fr). Using PhyML 3.0[59], we also inferred approximate likelihood-ratio test (aLRT)[60] for branches to statistically evaluate the branch support in the phylogeny.

Based on the CS tree and the MSA, ancestral sequences were inferred using the codeML package within PAML v4.9[61]. To adjust for gaps and the different length of the N-termini of the CS sequences, their ancestral state was determined using parsimony inference in PAUP 4.0a based on a binary version of the MSA (1=amino acid, 0=gap, no residue). The state assignment for each node in the tree (amino acid or gap) was then applied to the inferred ancestral sequences. Alignment figures were created using ESPrit3.0[62].

Mass photometry

Measurements were performed on a OneMP or a TwoMP mass photometer (Refeyn Ltd, UK). Reusable silicone gaskets (CultureWellTM, CW-50R-1.0, 50-3mm diameter x 1 mm depth) were set up on a cleaned microscopic cover slip (1.5 H, 24 x 60 mm, Carl Roth, Germany) and mounted on the stage of the mass photometer using immersion oil (IMMOIL-F30CC, Olympus, Japan). The gasket was filled with 19 µL buffer (PBS or 20 mM Tris, 200 mM NaCl pH 7.5) to focus the instrument. Then, 1 µL of prediluted protein solution (1 µM) was added to the buffer droplet and mixed thoroughly. Final concentration of the proteins during measurement was between 12.5-50 nM. Data was acquired for 60 s at 100 frames per second using AcquireMP (Refeyn Ltd, v1.2.1). The resulting movies were processed and analyzed using DiscoverMP (Refeyn Ltd, v2022 R1). The instrument was calibrated at least once during each measuring session using a homemade calibration standard of a protein mixture with known sizes (86-430 kDa).

For measurements with substrates/effector molecules the prediluted protein sample (2 µM) was incubated for 10 min with the respective effector concentration. The same substrate concentration was also included in the buffer in the gasket that was used for focusing.

Kinetic enzyme assays

For the CS kinetic assays the colorimetric quantification of thiol-groups was used based on 5,5′-dithiobis-(2-nitrobenzoic acid) (DTNB)[63,64]. The photospectrometric reactions were carried out in 50 mM Tris pH 7.5, 10 mM KCl, 0.1 mg/mL DTNB and 25 nM protein concentration at 25 °C. To measure K_m values, one substrate was saturated and added to the reaction mix (1 mM oxaloacetate or 0.5 mM acetyl-CoA respectively). The other substrate was varied in concentration and added last to start the reaction. Reaction progress was followed by measuring the appearance of 2-nitro-5-thiobenzoate at 412 nm (Extinction coefficient 14.150 M^{-1} cm^{-1}) in a plate reader (Infinite M Nano+, Tecan, Switzerland). Data analysis and determination of enzyme kinetic parameters was done with GraphPad Prism (Version 8.4.3 (686)).

Crystallography and structure determination

A solution of 20 mg/mL CS from *M. sulfidovorans* was incubated with 5mM oxaloacetic acid. Crystallization was then performed by the hanging-drop method at 20 °C in 1µL drops consisting of equal parts of protein and precipitation solutions. The crystallization condition consisted of 1 M LiCl, 0.1 M citric acid and 10% PEG 6000 at a pH of 5.0. For data collection, a cryo solution consisting of 70% mother liquor and 30% glycerol was added and the crystal was flash-frozen in liquid nitrogen. Data were collected at 100 K at Deutsches Elektronen-Synchrotron (Hamburg, Germany). MxCube2 was used for data collection[65]. Data were processed with XDS (version 06/2023) and scaled with XSCALE[66]. The structure was initially determined by molecular replacement with PHASER[67], utilizing a monomer generated *in silico* by AlphaFold2[68]. The

structure was then iteratively built in WinCoot (version 0.9.6)[69] and refined with PHENIX (version 1.19)[70].

Cryo-electron microscopy

For cryo-EM sample preparation of CS from *A. comosus*, 4.5µl of the purified protein at 1 mg/mL were applied to glow discharged Quantifoil 2/1 grids, blotted for 3.5s with force 4 in a Vitrobot Mark IV (Thermo Fisher) at 100% humidity and 4°C, and plunge frozen in liquid ethane, cooled by liquid nitrogen. Cryo-EM data was acquired with a FEI Titan Krios transmission electron microscope using SerialEM software[71]. Movie frames were recorded at a nominal magnification of 29,000X using a K3 direct electron detector (Gatan). The total electron dose of ~55 electrons per Å2 was distributed over 30 frames at a calibrated physical pixel size of 1.09 Å. Micrographs were recorded in a defocus range of -0.5 to -3.0 µm.

For each sample of polydisperse anc_{1b} at 0.25 mg/mL 4 µl of protein suspension was applied onto glow discharged Quantifoil grids (R 1.2/1.3 Cu 200). All samples were plunge frozen in a propane/ethane (63%/37%) mixture using a ThermoFisher Vitrobot Mark IV at 4 °C, 100% humidity and a blotting time of 5 seconds. For the first data set blot force was set to a value of 4 whereas for data set two and three a value of 6 was used. Cryo-EM data of the polydisperse anc_{1b} sample was collected on a CRYO ARM 200 (JEOL) transmission electron microscope (TEM) operated at 200 kEV and equipped with a K2 direct detector (Gatan). Three datasets were recorded at a magnification of 60,000× corresponding to a pixel size of 0.85 Å/pixel and with a total dose of 50 e/Å^2 for dataset 1 (2,646 movies, 50 frames), 40 e/Å^2 for both dataset 2 (5,258 movies, 40 frames), and dataset 3 (1,569 movies, 25 frames). SerialEM[71] was used for automated data acquisition and micrographs were pre-processed with CryoSPARC Live[72].

Image processing, classification and refinement

For the CS from *A. comosus* all processing steps were carried out in cryoSPARC v3.1.0[72]. A total of 4,294 movies were aligned using the patch motion correction tool and contrast transfer function (CTF) parameters were determined by the patch CTF tool. 1,600 Micrographs of estimated CTF resolution < 3.5 Å were selected for particle picking. A Topaz convolutional neural network particle picking model[73] was generated from initial blob picking and 2D classifications, followed by several rounds of Topaz Train. 587,742 particles were extracted in a box size of 200 by 200 pixels at a pixel size of 1.09 Å using the Topaz extract tool. After 2D classification, 372,284 particles were selected to generate five initial models by running the ab-initio reconstruction tool. The model corresponded to the octamer (33.9% particles) was further refined by non-uniform refinement, reaching a final resolution of 4.15 Å (GSFSC = 0.143) which was used for model building. Local-resolution and 3D-FSC plots were calculated using the local resolution tool and the "Remote 3DFSC Processing Server" web interface[74], respectively. The initial model was built based on a dimeric prediction from AlphaFold-Multimer and refined using WinCoot[69], in which the C-termini (residues 488-513) were modelled by alanine. The model was subjected to real-space refinements against the respective density maps using phenix.real_space_refine implemented in PHENIX v1.19.2[75].

The polydisperse anc_{1b} datasets were processed with CryoSPARC. Particles were initially picked with the blob picker tool, extracted with a box size of 408 pixels, and subjected to iterative rounds of 2D classification. The cleaned 2D class averages were used to train a neuronal network (ResNet8) for particle picking via Topaz[73] and processed via 2D classification and the ab-initio reconstruction tool to further clean-up the data. Particles from the three datasets were first

processed separately and later combined. The 2D classification job, shown in Extended Data Fig. 5c, was performed with enforced non-negativity and solvent-clamping.

KO-strain construction

To facilitate the characterization of CS, we deleted the native gltA coding region from the E. coli genome. To do so, we used the no-SCAR protocol[76]. We obtained plasmids from Addgene and transformed the pCas9-CR4 (ID: 62655) plasmid into E. coli BL21 cells. Plasmid pKDsgRNA-ack (ID: 62645) was then used as a PCR template for gRNA insertion (PCR1: GGTGTGTTCACCTTTGACCCGTTTTAGAGCTAGAAATAGCAAG; TTTATAACCTCCTTAGAGCTCGA; PCR2: GGGTCAAAGGTGAACACACCGTGCTCAGTATCTCTATCACTGA, CCAATTGTCCATATTGCATCA). These two fragments were then combined into a single plasmid via CPEC using Q5 polymerase. This plasmid was transformed into the pCas9-CR4 containing strain. Lambda-red recombination was then induced with arabinose and the strain transformed with 5 ul of a 100 µM ssDNA fragment (T*A*A*G*TTCCGGCAGTCTTACGCAATAAGGCG-CTAAGGAGACCTTAATGATTGATTGCTAAGCCGTTTACTTTCCGGACCCGCCTTTAATAG) that contained 45 bp of homology to the regions immediate 5' and 3' of the native gltA coding region. Successful deletion was confirmed via Sanger sequencing. The pKDsgRNA-ack plasmid containing the gRNA was removed by growing the resulting strain at 37C. The pCas9-CR4 plasmid was removed by transforming in the pKDsgRNA-p15 (ID: 62656) plasmid, which targets the pCas9-CR4 plasmid. Finally, the pKDsgRNA-p15 plasmid was removed by growing the strain at 37 °C.

Complementation assays, Data analysis and Western Blots

The KO-strain *E.coli* BL21(DE3)ΔgltA was transformed with IPTG-inducible expression plasmids that encoded for the different CS (pLIC-backbone, same as used for protein

purification). We tested growth of the transformed strains in LB-medium, where there is no selection pressure on CS activity, to test if the heterologous CS-genes had toxic effects. CS genes that impaired growth under non-selective conditions were excluded from further analysis. For growth rate experiments under CS selective conditions precultures of the complemented strains were grown overnight in LB-medium at 37 °C, 240 rpm. Cells from these overnight cultures were harvested by centrifugation at 15.900 xg for 3 min and washed two times with PBS. The washed cells were used to inoculate new cultures in 96-well plates in minimal medium (M9+glucose) to an OD_{600}=0.05. Each strain was set up in triplicates (180 µL/ well) and incubated in a microplate reader (Infinite M Nano+, Tecan, Switzerland) at 37 °C, 220 rpm for 40 h. No inducer was added during the growth experiment and CS production relied solely on leaky expression. To prevent evaporation the lid was placed onto the 96-well plates and sealed with parafilm. As a negative control we used the untransformed KO-strain *E.coli* BL21(DE3) ΔgltA and as positive control we complemented the KO-strain with the native CS from E.coli BL21(DE3) on the plasmid. Bacterial growth was monitored by absorbance measurements at 600 nm every 15 min and converted to OD_{600}-values via a standard curve. We used the growth rate package v0.8.2 within R studio to fit parametric growth models to the collected data and infer maximum growth rates[77].

We tested CS complementation at different protein production rates with spot assays in which the strains were grown with different concentrations of the inducer IPTG. Overnight cultures in LB were diluted in distilled water to OD_{600}=1.0. Then, a five-step serial dilution of the cell suspension was performed using a ratio of 1:5 for each step. The diluted cells were spotted on solid M9 + glucose medium with different IPTG concentrations (0, 0.01, 0.02, 0.03, 0.04, 0.05 mM) using a Singer Instruments Rotor HDA+ screening robot. Spotting was performed in a 7 x 7 grid with revisit of the source plate for each transfer. The plates were set up in triplicates and

incubated for 72 h at 37 °C. Images of the plates were taken every 24 h using the Singer Instruments PhenoBooth+.

Protein production of the different CS homologs was analyzed via western blot. Complemented strains were grown in LB (50 mL) for 15 h, 37 °C and 240 rpm. Cells were harvested by centrifugation at 4,000 rpm for 20 min and resuspended in 1 mL PBS. The cells were lysed mechanically by addition of 0.1 mm glass beads and using the Fastprep24 (MP Biomedicals, USA-CA) at a strength of 6.0 for 30 s for four cycles. Cells were pelleted afterwards by centrifugation at 13,300 rpm for 20 min. The supernatant was transferred and the cell pellet was resuspended in 1 mL PBS. The protein concentration of the samples was measured using a Bradford standard curve. SDS gel was loaded with 10 µg of protein for each sample and 30 µg of resuspended cell pellet. Resolved proteins were afterwards transferred onto a nitrocellulose membrane via western blotting. The membrane was blocked overnight in 5% milk, and then incubated with an Anti-histag antibody conjugated with a horseradish peroxidase for 1 hTBS-T (tris-buffered saline with tween) with 5% milk. The CS proteins were detected using chemiluminescence with a ChemiDoc imaging system (Bio-Rad, USA-CA).

Modelling of homo-oligomeric complexes with AlphaFold2 Multimer

Structural models were generated for anc3a or the extant CS from *N. inopinata, A. comosus, P. americanus* and *C. fasciculata* using the AlphaFold2 Multimer ColabFold server[68] with default settings. Modelled structures are deposited in the source data. Data were rendered and visualized with PyMol (v.2.4.0).

Data availability: Atomic structures reported in this paper are deposited to the Protein Data Bank under accession code 8QWB and 8QZP. The cryo-EM data was deposited to the Electron Microscopy Data Bank under EMD-18779. All raw data for MP spectra and kinetic traces as well as phylogenetic trees, alignments, and ancestral sequences will be deposited on Edmond, the Open Research Data Repository of the Max Planck Society for permanent public access.

Acknowledgments: FLS, TS, GB, DS, and GKAH are generously supported by the Max-Planck Society. JWT acknowledges support from the NIH (R01GM131128). Cofounded by the European Union (ERC, EVOCATION, 101040472). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. The authors acknowledge support from the EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany). We are grateful for support from Sriram Garg who inferred models of oligomeric CS via Alphafold Multimer v2.

Author contributions: FLS, AP, JWT and GKAH conceived the project. FLS and GKAH analyzed data and planned experiments. FLS performed phylogenetics, enzyme kinetic measurements and analysis of the structural data. FLS and TS inferred ancestral sequences, performed protein purification and MP measurements. AW and BPM constructed the CS knockout strain. TS performed complementation assays together with DS. CNM collected, solved, and refined the X-ray structure with supervision from GB. SB, MG, TH and YKL collected and processed the cryo-EM data sets. YKL refined the cryo-EM structure. FLS and GKAH wrote the manuscript with contributions and comments from all authors.

Competing interests: The authors declare no other competing interests.

Supplementary Figures

S1 – Full phylogenetic tree of CS

S2 – MP measurements of additional characterized extant CS

S3 – Sequences, confidence and robustness of reconstructed ancestral CS sequences

S4 – Cryo-EM data processing

S5 – Growth data of complemented strains with fitted growth curves

Supplementary Tables

T1 – List and sequences of DNA sequences of proteins used in this study

Correspondence and requests for materials should be addressed to georg.hochberg@mpi-marburg.mpg.de

Extended Data Table 1 Data collection and refinement statistics for the crystal structure of *M. sulfidovorans* CS

	Citrate synthase from *M.sulfidovorans* PDB 8QWB
Wavelength	0.9762
Resolution range	48.31 - 3.201 (3.316 - 3.201)
Space group	P 1 21 1
Unit cell	155.69 96.61 203.1 90 110.41 90
Total reflections	636593 (60352)
Unique reflections	92644 (9175)
Multiplicity	6.9 (6.6)
Completeness (%)	98.85 (97.95)
Mean I/sigma(I)	10.32 (0.88)
Wilson B-factor	114.97
R-merge	0.1319 (2.263)
R-meas	0.1428 (2.457)
R-pim	0.05402 (0.9447)
CC1/2	0.998 (0.467)
CC*	1 (0.798)
Reflections used in refinement	92635 (9161)
Reflections used for R-free	4635 (459)
R-work	0.3252 (0.4478)
R-free	0.3245 (0.4490)
CC (work)	0.890 (0.463)
CC (free)	0.883 (0.481)
Number of non-hydrogen atoms	30067
macromolecules	30067
Protein residues	3816
RMS (bonds)	0.012
RMS (angles)	1.35
Ramachandran favored (%)	93.70
Ramachandran allowed (%)	6.22
Ramachandran outliers (%)	0.08
Rotamer outliers (%)	2.84
Clashscore	28.54
Average B-factor	134.04
macromolecules	134.04

Statistics for the highest-resolution shell are shown in parentheses.

Extended Data Table 2 Cryo-EM data collection, refinement, and validation statistics of nmCS from *A. comosus*

	CS *A. comosus* **PDB 8QZP** **EMD-18779**
Data collection and Processing	
Microscope	Titan Krios
Voltage (keV)	300
Camera	Gatan K3
Magnification	29,000X
Pixel size at detector (Å/pixel)	1.09
Total electron exposure ($e^-/Å^2$)	55
Exposure rate (e^-/pixel/sec)	4.5
Number of frames collected during exposure	30
Defocus range (μm)	-3.0 to -0.5
Automation software	SerialEM
Micrographs collected (no.)	4,294
Micrographs used (no.)	1,600
Total extracted particles (no.)	587,742
Reconstruction	
Refined particles (no.)	372,284
Final particles (no.)	108,304
Symmetry imposed	C1
Resolution (global, Å)	
FSC 0.143 (unmasked/masked)	4.1/4.8
Resolution range (local, Å)	53.9 to 2.3
Map sharpening *B* factor ($Å^2$)	-222.5
Map sharpening methods	Global B factor
Model composition	
Protein	CS A. comosus
Model Refinement	
Real space refinement software	PHENIX v1.19.2
Model-Map scores (CC_mask)	0.75
B factors ($Å^2$)	(min/max/mean)
Protein	30.00/329.44/169.74
R.m.s. deviations from ideal values	
Bond lengths (Å)	0.005
Bond angles (°)	0.797
Validation	
MolProbity score	2.41
CaBLAM outliers	3.76
Clashscore	22.20
Poor rotamers (%)	1.23
Cβ outliers (%)	0.00
EMRinger score	0.30
Ramachandran plot	
Favored (%)	91.49
Allowed (%)	7.98
Outliers (%)	0.52

Extended data Table 3 Kinetic enzyme parameters of *D. pimensis* CS and isolated SEC-fractions containing different oligomeric state distributions

	K_{cat} (s^{-1})	$K_m$oxaloacetate (µM)
D. pimensis **– non-fractionated**	36.6 ± 1.1	25.4 ± 3.1
F1	39.9 ± 1.3	29.6 ± 3.9
F2	23.2 ± 0.6	15.0 ± 1.8
F3	16.6 ± 0.5	17.7 ± 2.5
F4	18.4 ± 0.6	18.4 ± 2.5
F5	14.2 ± 0.7	25.9 ± 5.0

Measurements were performed at 25 °C. N=3 technical replicates, errors = standard error

Extended Data Table 2 Kinetic characterization for different extant and ancestral CS and their variants

	aceytlCoA		oxaloacetate	
	K_{cat} (s^{-1})	$K_m$acetyl-CoA (μM)	K_{cat} (s^{-1})	K_m oxaloacetate (μM)
anc_{3a}	9.1 ± 0.3	22.4 ± 2.6	10.3 ± 0.1	23.9 ± 1.1
anc_{3b}	5.6 ± 0.2	8.6 ± 1.9	7.5 ± 0.2	13.2 ± 2.2
anc_{3a+5}	21.9 ± 0.8	56.5 ± 6.8	23.3 ± 61.1	61.1 ± 10.2
M. sulfidovorans WT	42.4 ± 1.0	127.4 ± 7.7	33.3 ± 1.8	50.5 ± 5.2
M. sulfidovorans W150A	34.1 ± 1.0	104.7 ± 8.9	26.7 ± 1.1	36.7 ± 5.7
A. comosus WT	41.1 ± 1.5	13.2 ± 2.3	49.2 ± 2.2	72.1 ± 10.1
A. comosus Δ487-513	41.8 ± 1.5	11.3 ± 2.0	58.9 ± 2.6	87.2 ± 11.7

Measurements were performed at 25 °C. N=3 technical replicates, errors = standard error

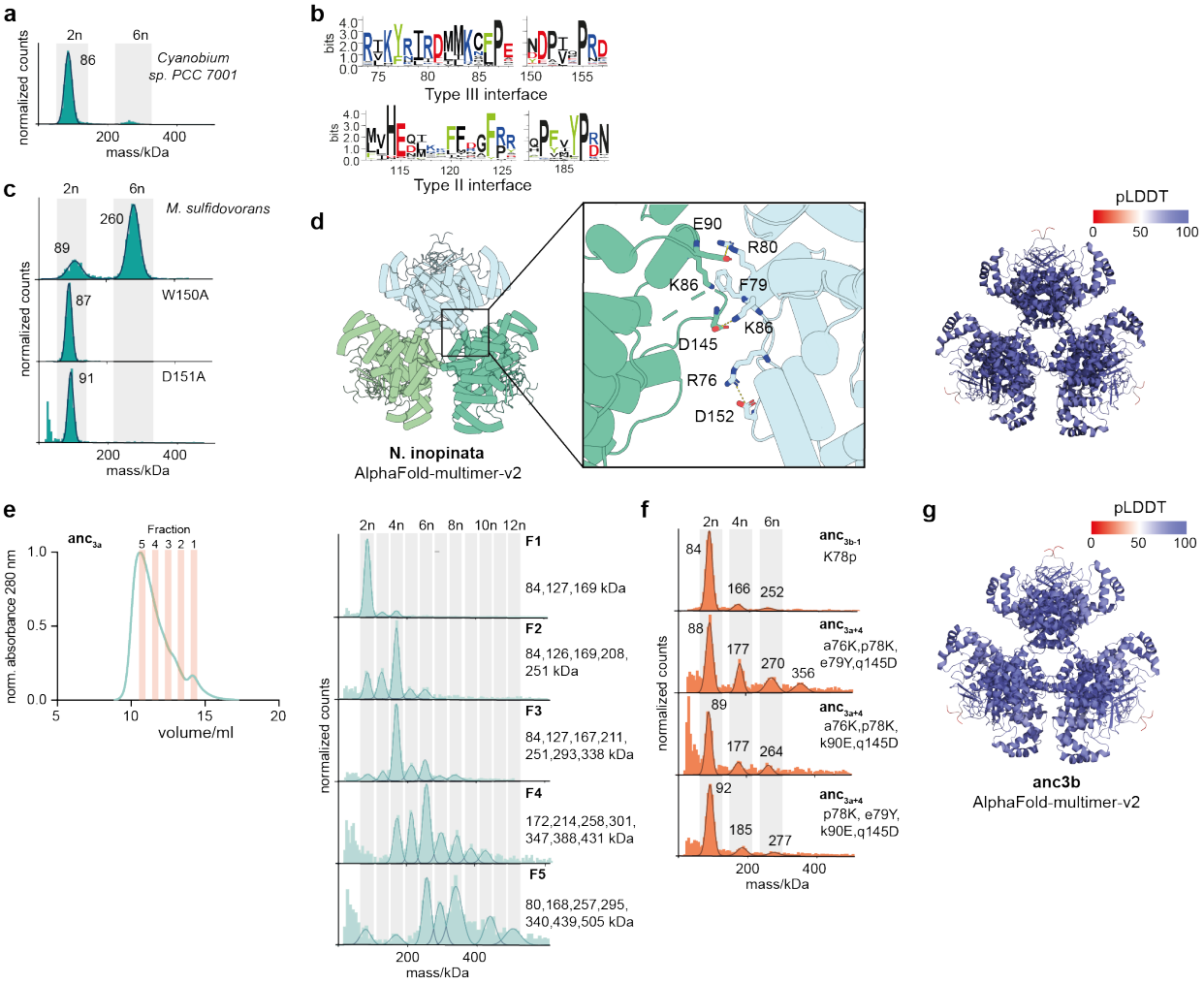
Extended Data Fig. 1 Conservation of residues inducing hexameric assembly in type III CS.

Alignment of characterized type III CS, ancestors anc_{3a}, anc_{3b} and the type II CS from *E. coli*.

Residues marked in teal are part of the hexamer interface in type III enzymes. Connecting lines

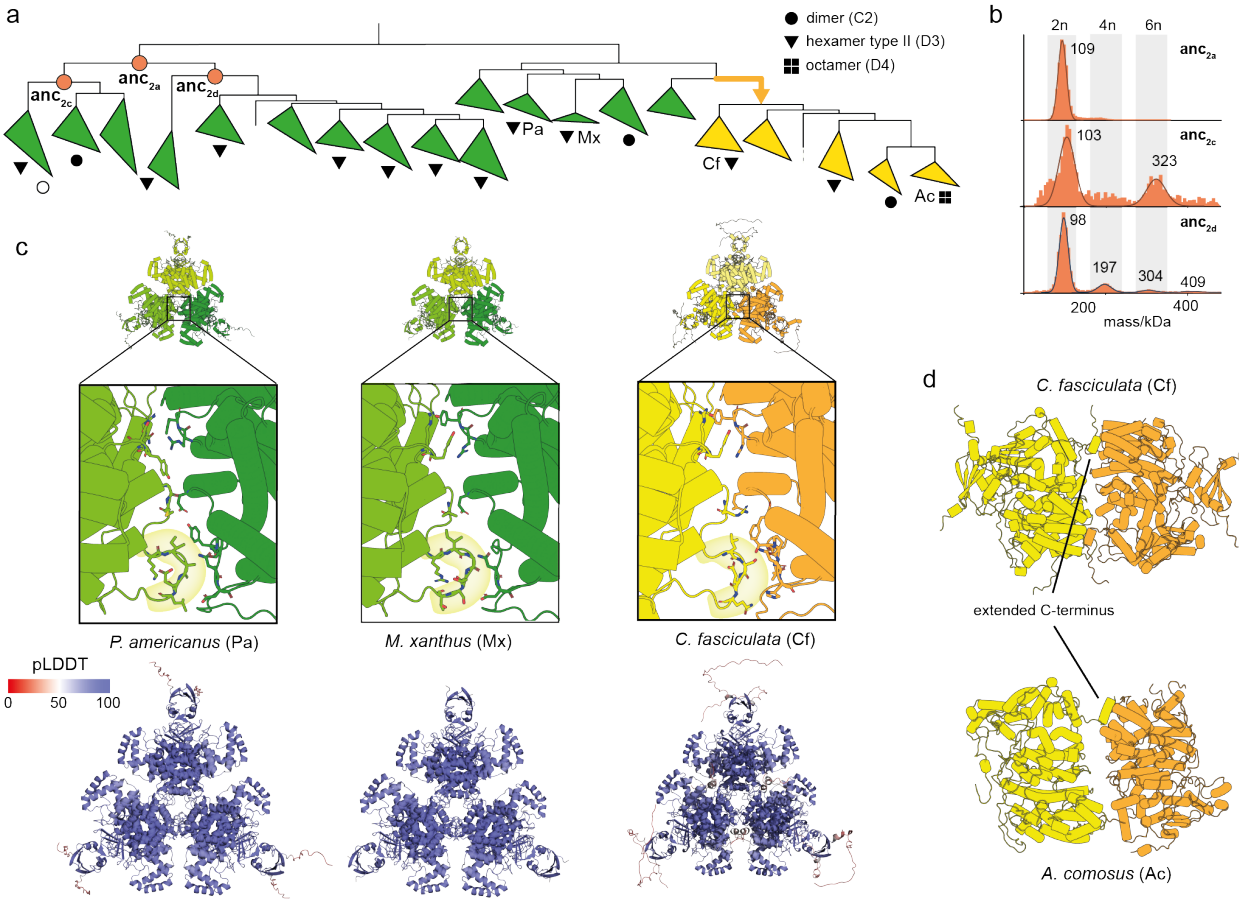
indicate direct interactions between residues in structures (teal=*M. sulfidovorans*, orange=anc_{3b}).

Residues that take part in the hexamer formation in *E. coli* CS are colored in green.



Extended Data Fig. 2 Important residues for type III hexamer formation (a) MP measurement of the dimeric type III CS from *Cyanobium* sp. PCC 7001. (b) Sequence logo of amino acid residues that are found within the interface of type II or III CS, demonstrating strong differences. Homologous sites are aligned, shift in site numbers results from a longer N-terminus in type II CS. (c) MP measurement of interface variants of the type III CS from *M. sulfidovorans*. (d) Model of the hexameric type III CS from *N. inopinata* inferred with AlphaFold-multimer-v2 with close up of the interface between dimers. Right: structural model colored according to the predicted local distance difference test (pLDDT) (e) SEC trace of the polydisperse anc_{3a} and MP measurements of the isolated fractions. (f) MP measurements of variants of anc_{3a} with different combinations of substitutions that emerged in the interval to anc_{3b} (small and capital letters

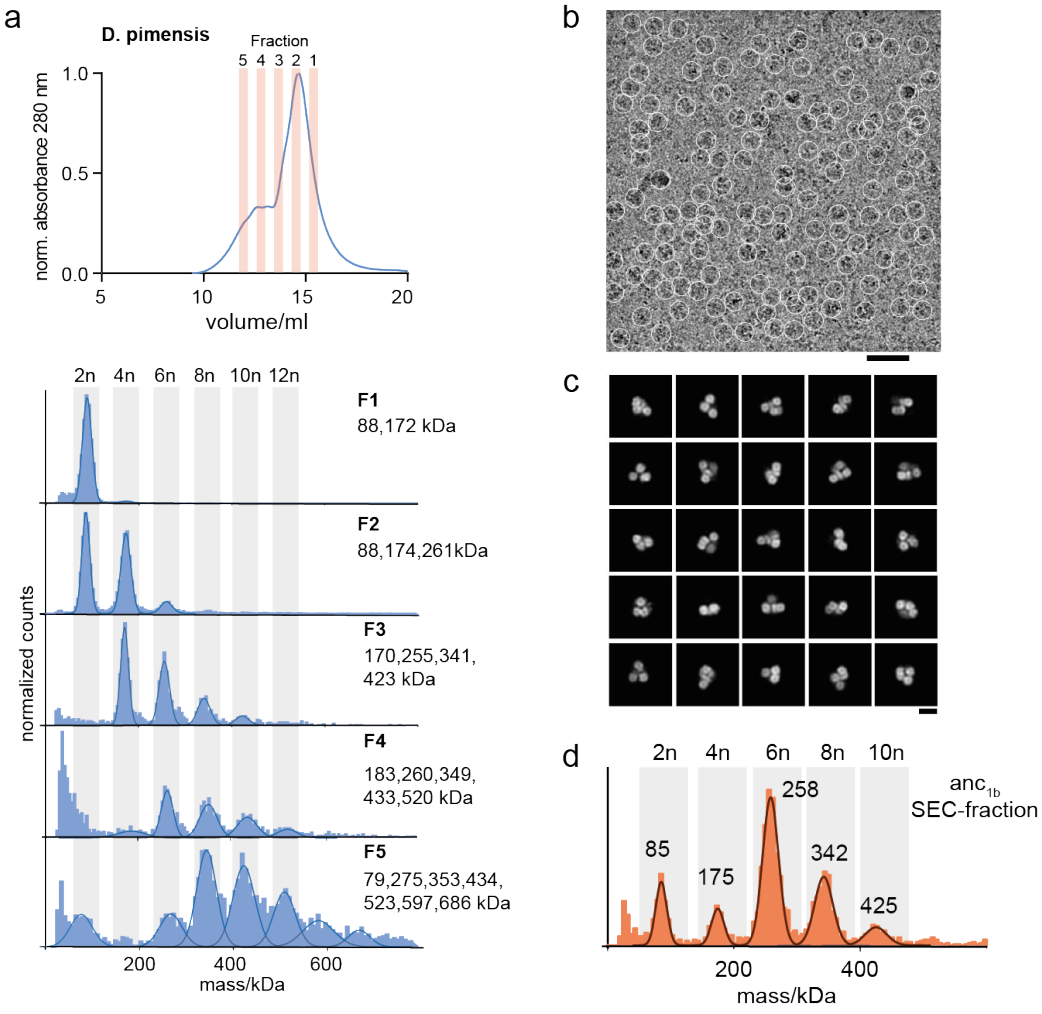
indicate the ancestral and derived amino acid, respectively). No combination of 4 was sufficient to induce hexameric assembly. The significance of the substitution p78K was shown by a reversal (K78p) in the hexameric anc_{3b}. (g) AlphaFold-multimer model of Anc3b color-coded according to the site-wise pLDDT score.



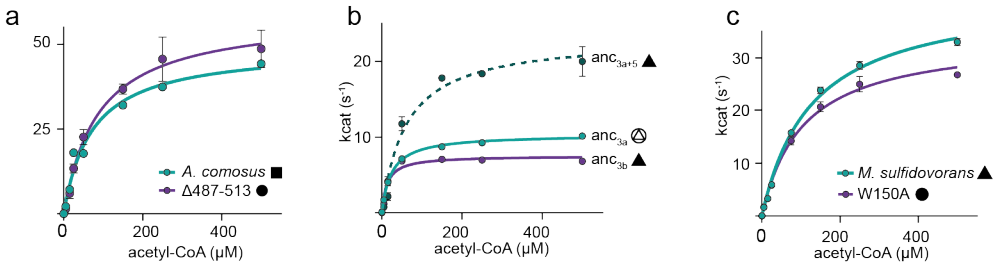
Extended Data Fig. 3 Inconclusive inference of the emergence of type II hexamers but conserved involvement of the JK-loop (a) Schematic representation of part of the CS phylogeny displaying the quaternary structures of characterized type II CS. Nodes corresponding to resurrected ancestral CS are indicated. (b) MP measurements of anc_{2a} and its descendants anc_{2c} and anc_{2d}. (c) Upper: AlphaFold-Multimer predictions of hexameric type II CS and nmCS with close ups on the interface that display the involvement of the JK-loop for all of them. Lower: pLDDT-colored structures of the same models. (d) Comparison of the extended C-terminus in the hexameric structure from *C. fasiculata* (AlphaFold-Multimer) and octameric

structure from *A. comosus* (cryo-EM). Only two dimers are displayed for both structures to highlight the cross-connection by the C-terminus between adjacent dimers.

Extended Data Fig. 4 Conservation of residues inducing oligomeric assembly in type II CS and nmCS. Alignment of the characterized type II CS and nmCS. Residues marked in green take part in the hexamer formation in the hexamer from *E. coli*. The extended C-terminus of nmCS is colored in yellow.

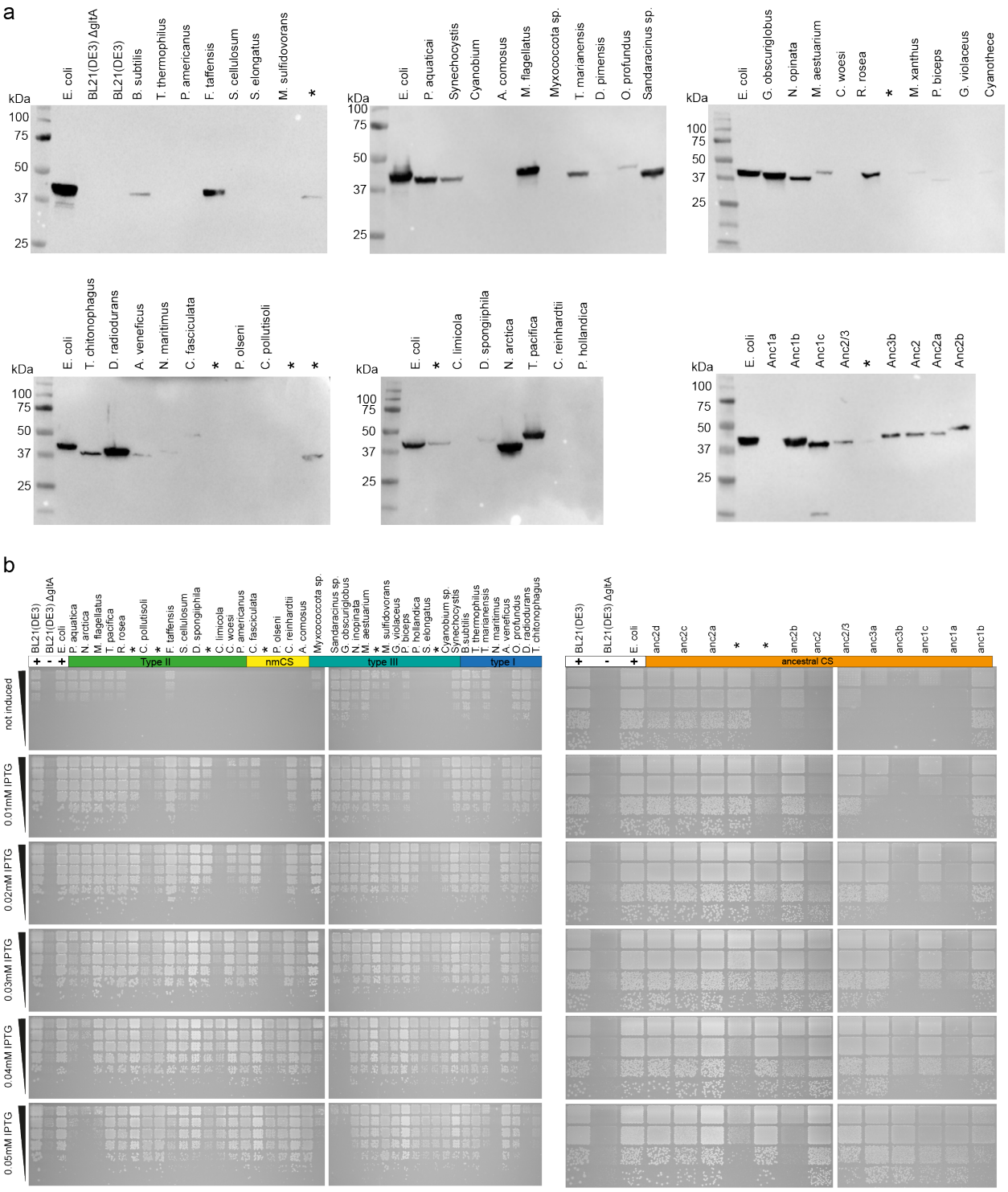


Extended Data Fig. 5 Polydisperse type I enzymes form multiples of dimers (a) SEC trace of polydisperse CS from *D. pimensis* and MP measurements of the isolated fractions. (b) Exemplary cryo-EM micrograph of oligomeric complexes of polydisperse anc_{1a}, low-pass filtered at 5 Å. Particle picks are depicted as white circles (diameter = 180 Å). Scale bar = 40 nm. (c) 2D class averages of polydisperse anc_{1a}. Scale bar = 10 nm. (d) MP measurement of the SEC-fraction used for cryo-EM.



Extended Data Fig. 6 Kinetic measurements of CS variants with saturated oxaloacetate

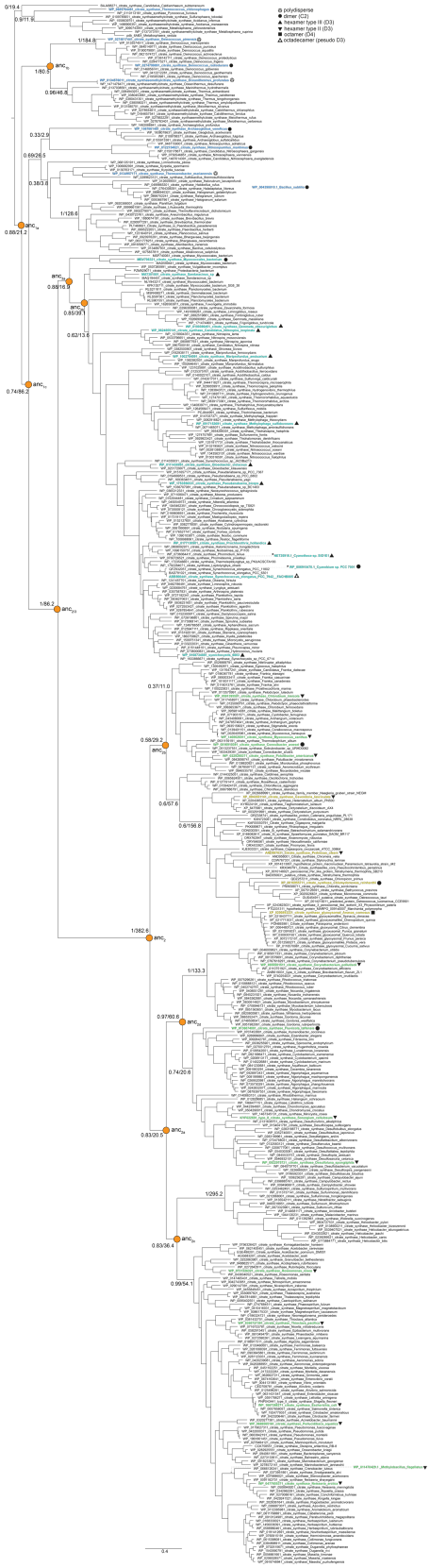
Michaelis-Menten kinetics of (a) extant CS from *A. comosus* and a variant that disrupts interfaces between dimeric subcomplexes and prevent assembly into larger oligomers (Δ487-513). Error bars = SD, n = 3 technical replicates; (b) the ancestral CS bracketing the emergence of hexamers within type III enzymes (anc_{3a}, anc_{3b}) and the minimal substitution construct to yield hexameric complexes (anc_{3a+5}). Error bars = SD, n = 3 technical replicates; (c) extant CS from *M. sulfidovorans* and a variant that disrupts interfaces between dimeric subcomplexes and prevent assembly into larger oligomers (W150A). Error bars = SD, n = 3 technical replicates W150A



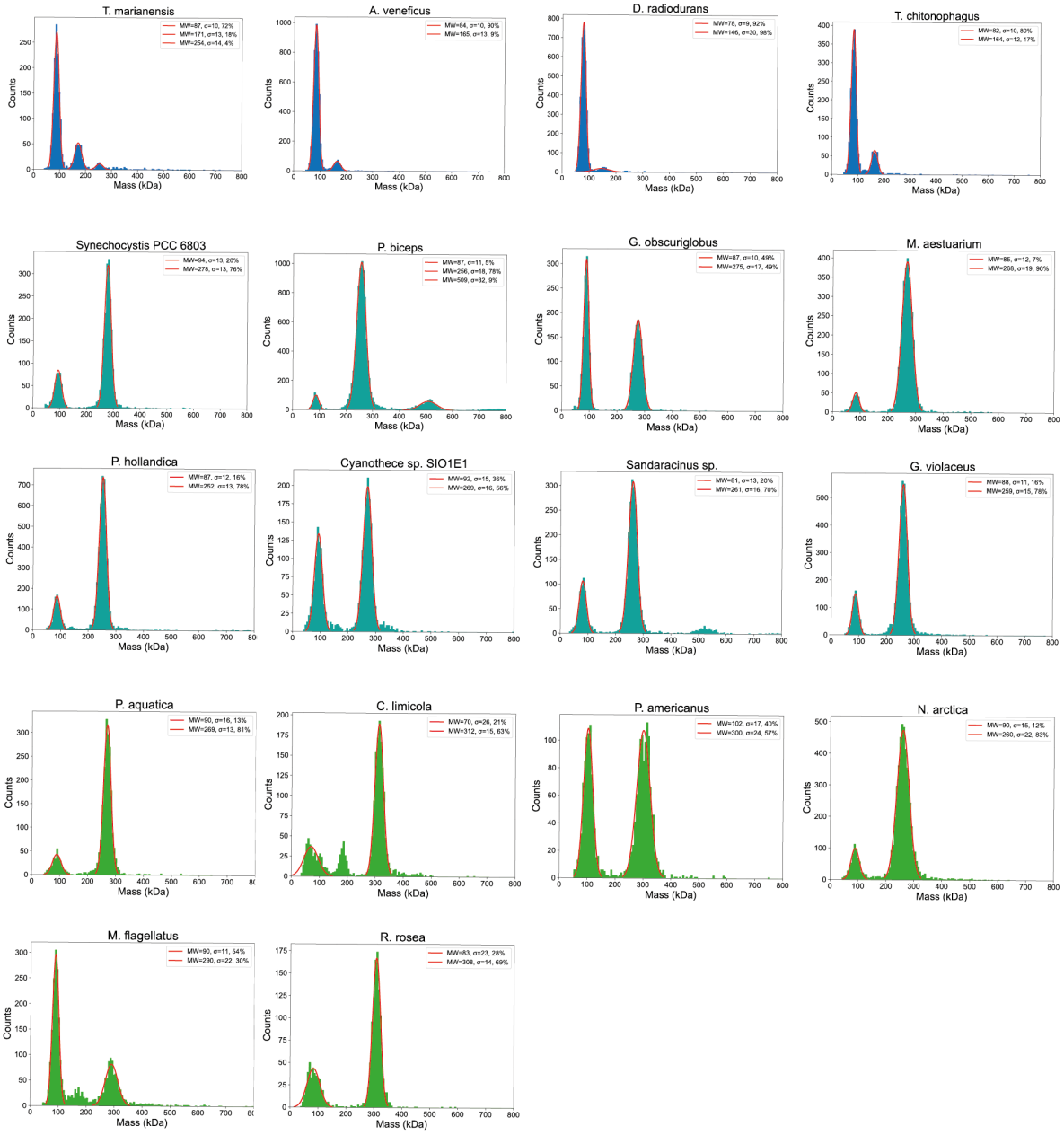
Extended Data Fig. 7 Protein production and complementation on solid media (a) Western Blots depicting the CS protein production of all tested extant and ancestral CS enzymes in *E. coli* using the same conditions as for the growth curves. (b) Spot-assays of complemented *E. coli*

BL21(DE3) ΔgltA strains. Cultures were spotted in a five-step serial dilution using a ratio of 1:5 for each step and incubated on M9-solid media supplemented with glucose and different IPTG concentrations. One representative plate is shown for each experiment, out of a total of three replicates for each plate.

Supplementary Information

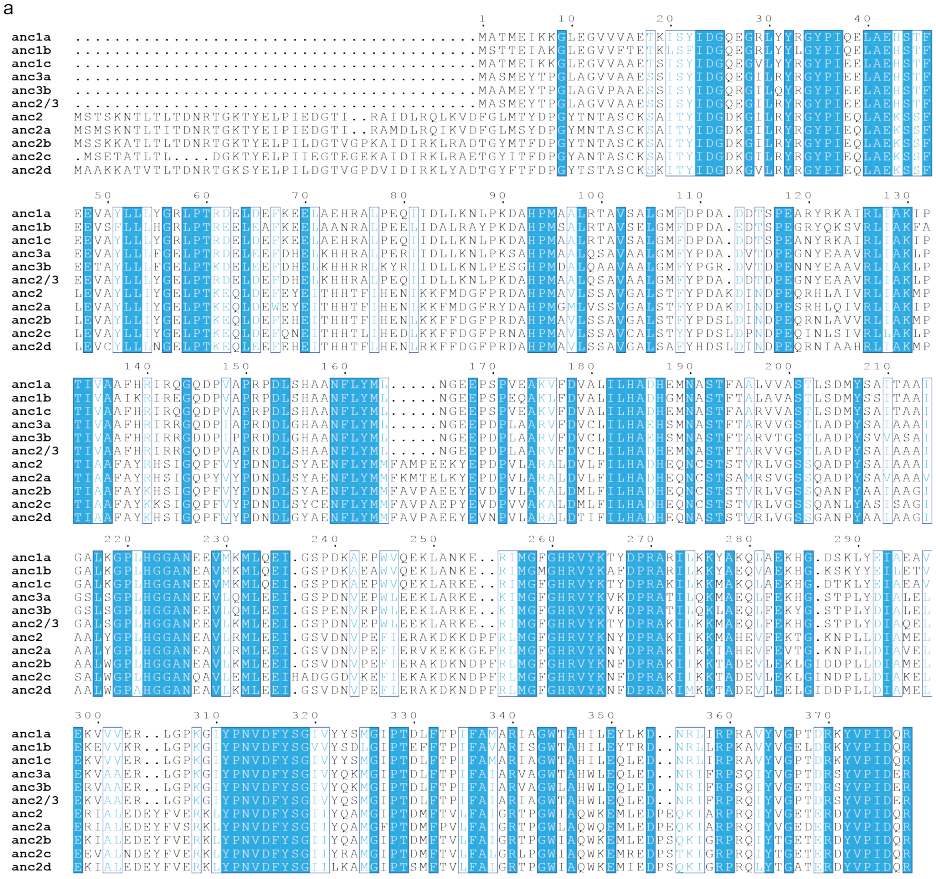


Supplementary Fig. 1 Phylogenetic tree of CS Full tree with all sequences (organism name and NCBI-identifier) to infer the evolutionary history of CS in Bacteria, Archaea and nmCS in Eukaryotes. Extant sequences that were purified are colored and their quaternary structure is indicated with a symbol. Internal nodes of the tree for which ancestral sequences were reconstructed via ASR are indicated. Branch supports values are shown for these nodes and additional important internal branches as Felsenstein's bootstrap values and approximate likelihood ratio test statistics (fbe/aLRT). Four sequences of characterized extant CS were not included in the inference of the main phylogeny used for ASR. Their position on the tree was inferred afterwards by recalculating the phylogeny and is indicated with dashed lines.

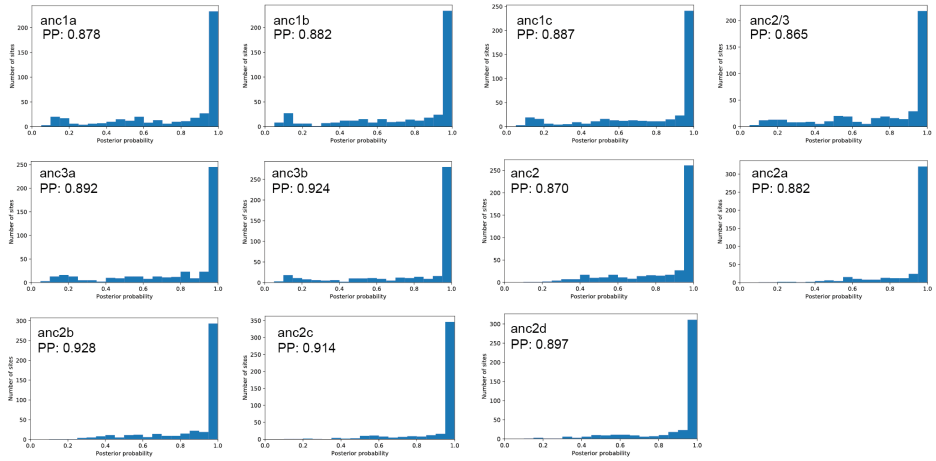


Supplementary Fig. 2 MP measurements of additional characterized extant CS

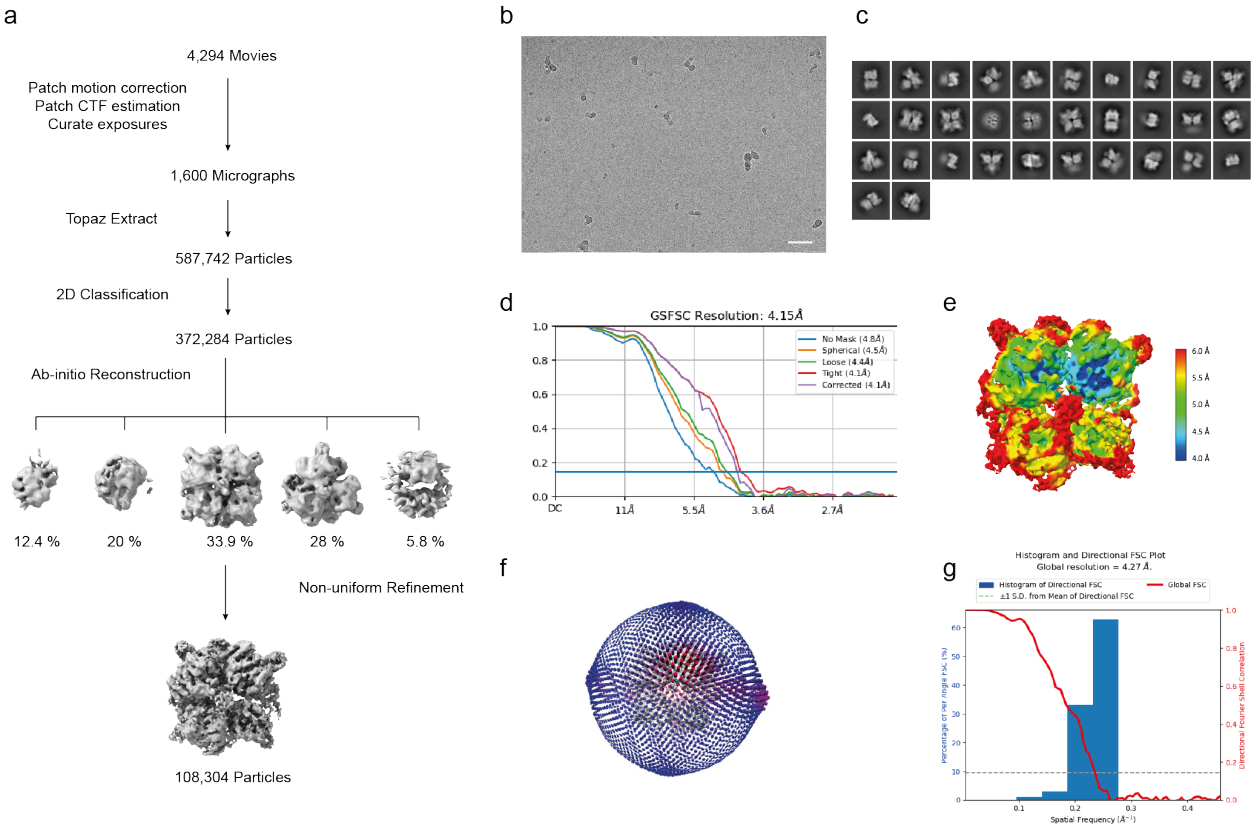
a



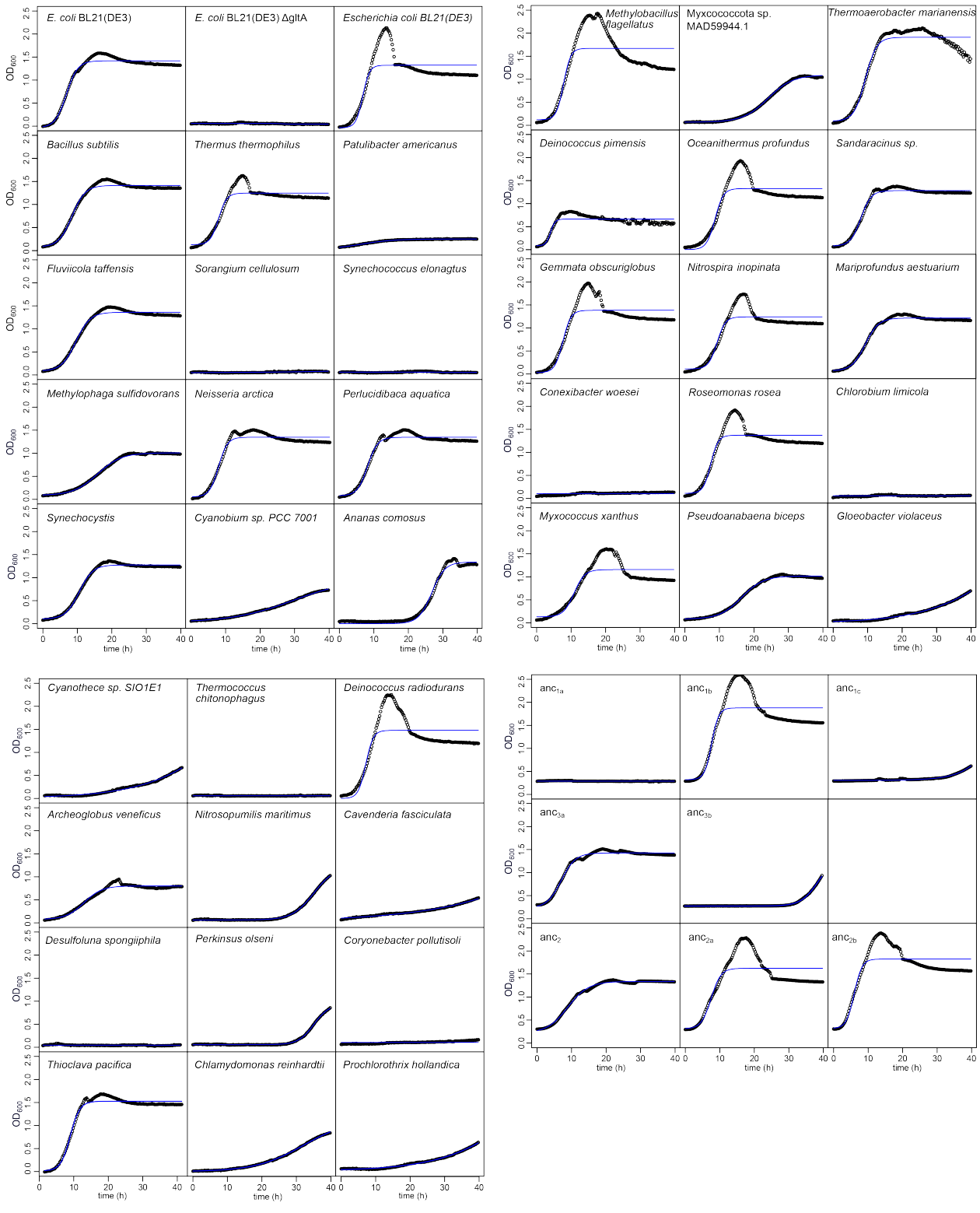
b



Supplementary Fig. 3 Sequences and confidence of reconstructed ancestral CS Sequences and robustness of the inferred ancestral proteins (a) Alignment of the amino acid sequences of the inferred ancestral proteins. (b) Histograms per-site of the posterior probabilities of the maximum aposteriori state across reconstructed sites for all eleven ancestral proteins.



Supplementary Fig. 4 Cryo-EM data processing. (a) An overview of the image processing procedure (see Methods). (b) A representative micrograph of *A. cosmosus* CS acquired from FEI Titan Kriso (scale bar, 50 nm). (c) 2D classes selected for ab-initio reconstruction. (d) Fourier shell correlation (FSC) curves for the reconstruction applying different masks. (e) The final density map colored by local resolution. (f) Euler angular distribution of particles used in the final 3D reconstruction. (g) The histogram of directional FSC.



Supplementary Fig. 5 Growth data of complemented strains with fitted growth curves

1 **Supplementary Table 1 List of DNA sequences of proteins used in this study**

Protein (Accession number)	DNA sequence
Anc 1a	ATG GCC ACG ATG GAG ATC AAA AAG GGA CTG GAG GGA GTG GTG GTT GCC GAA ACG AAA ATC TCT TAT ATT GAT GGC CAG GAA GGA CGT TTG TAT TAC CGT GGG TAT CCT ATC CAA GAA TTA GCT GAG CAT TCT ACC TTC GAG GAG GTC GCG TAT TTA CTT CTT TAT GGA CGC CTG CCA ACG CGT GAT GAA CTG GAC GAG TTC AAA GAA GAG CTT GCA GAA CAT CGT GCT CTG CCA GAA CAA ATT ATT GAT TTG TTA AAA AAC CTT CCA AAG GAT GCC CAC CCA ATG GCC GCT CTT CGC ACG GCA GTC TCC GCT CTG GGA ATG TTC GAT CCG GAT GCA GAC GAT ACC TCA CCT GAA GCT CGC TAC CGC AAG GCC ATC CGC TTG ATC GCA AAA ATT CCG ACG ATC GTA GCT GCC TTT CAC CGT ATT CGC CAA GGA CAG GAC CCA GTA GCA CCG CGT CCT GAC TTG AGT CAC GCT GCC AAC TTC CTG TAT ATG CTG AAT GGC GAA GAA CCT TCG CCG GTG GAG GCC AAA GTG TTT GAT GTG GCT CTT ATC CTT CAT GCG GAC CAT GAG ATG AAC GCT TCC ACC TTC GCC GCA TTA GTC GTC GCG AGT ACT TTA TCA GAC ATG TAC AGT GCC ATT ACG GCT GCC ATT GGG GCC TTG AAG GGG CCG CTT CAT GGG GGT GCC AAC GAA GAG GTG ATG AAG ATG TTA CAA GAA ATT GGC TCA CCT GAT AAA GCT GAA CCC TGG GTG CAG GAG AAA CTT GCG AAT AAG GAA CGC ATT ATG GGC TTT GGA CAC CGT GTT TAC AAG ACT TAT GAT CCC CGT GCC CGC ATT TTG AAA AAA TAC GCG AAG CAG CTT GCC GAG AAA CAC GGT GAT TCC AAG CTG TAC GAG ATT GCT GAG GCG GTT GAG AAA GTG GTC GTG GAG CGC TTG GGG CCG AAA GGT ATT TAT CCA AAT GTG GAC TTT TAT TCA GGG ATT GTC TAT TAC TCC ATG GGT ATC CCA ACA GAC TTG TTT ACA CCT ATC TTC GCA ATG GCC CGT ATT GCT GGA TGG ACA GCT CAT ATT CTT GAA TAT CTT AAA GAT AAT CGC CTG ATC CGC CCT CGT GCA GTG TAT GTC GGT CCC ACG GAT CGC AAG TAC GTA CCC ATT GAC CAA CGT
Anc 1b	ATG TCG ACT ACA GAA ATC GCT AAG GGC CTG GAG GGA GTG GTG TTC ACT GAA ACC AAA CTT TCC TTC ATC GAT GGG CAA GAG GGG CGT TTA TAC TAC CTT GGG TAT CCG ATC CAA GAG TTA GCG GAA CAT TCG ACT TTT GAG GAG GTA AGC TTC CTT TTA CTG CAC GGT CGT TTA CCG ACT CGC GAA GAG TTG GAG GCT TTC AAA GAA GAG CTT GCC GCC AAT CGT GCA TTA CCT GAG GAG CTT ATC GAC GCC TTG CGT GCC TAC CCC AAG GAC GCC CAC CCG ATG TCC GCC TTA CGT ACT GCC GTC AGT GAA TTA GGC ATG TTC GAT CCG GAT GCG GAG GAT ACA TCG CCG GAG GGG CGT TAC CAA AAA TCA GTG CGC CTT ATT GCA AAG TTT GCG ACA ATT GTG GCC GCC ATC AAA CGC ATT CGT GAG GGG CAA GAT CCA GTG GCG CCA CGT CCC GAT TTG TCC CAC GCG GCG AAC TTC TTA TAT ATG TTA AAT GGG GAA GAG CCG AGT CCC GAG CAG GCC AAG CTG TTT GAC GTA GCA TTG ATC CTG CAC GCT GAC CAC GGT ATG AAT GCG TCG ACG TTT ACC GCG TTA GCG GTA GCA TCG ACC CTT AGT GAC ATG TAT TCC TCC ATC ACC GCA GCC ATC GGC GCA CTT AAA GGT CCT TTA CAC GGT GGA GCA AAT GAA GCG GTT ATG AAG ATG TTG CAG GAA ATC GGG AGT CCG GAC AAA GCA GAA GCG TGG GTC CAA GAG AAA TTG GCG AAC AAG GAG CGC ATT ATG GGT ATG GGG CAC CGC GTC TAT AAA GCT TTC GAT CCT CGT GCT CGC ATC TTG AAA AAG TAC GCG GAA CAA GTG GCT GAA AAG CAT GGC AAG TCC AAG TAT TAC GAA ATT CTG GAG ACC GTA GAA AAG GAG GTC GTT AAA CGT CTT GGG CCT AAG GGA ATT TAC CCA AAC GTG GAC TTT TAC AGC GGG GTG GTG TAC TCA GAT CTT GGT ATC CCC ACA GAG TTT TTT ACG CCT ATT TTC GCA GTC GCA CGT ATC TCT GGT TGG ACG GCG CAT ATT CTG GAG TAT ACT CGC GAC AAT CGT CTG CTT CGT CCA AAG GCA GTG TAC GTG GGT GAA CTT GAC CGC AAG TAT GTC CCT ATT GAT CAA CGT
Anc 1c	ATG GCC ACT ATG GAG ATT AAG AAG GGG TTA GAA GGC GTG GTA GCG GCT GAG ACA TCG ATC TCT TAC ATC GAT GGT CAG GAG GGG GTC CTT TAC TAT CGT GGG TAT CCA ATC GAA GAG TTA GCA GAA CAC TCA ACA TTT GAG GAG GTG GCT TAT TTA TTA CTT TAC GGG CGT TTG CCC ACT CGT GAC GAA CTT GAC GAG TTT AAA GAG GAG CTG GCC GAA CAC CGT GCC TTA CCT GAA CAG ATC ATC GAC TTG CTT AAA AAT CTT CCC AAG GAC GCC CAT CCG ATG GCT GCG TTG CGT ACC GCT GTG TCA GCA CTT GGA ATG TTC GAC CCT GAC GCC GAT GAC ACC TCT CCG GAA GCG AAC TAC CGT AAA GCG ATC CGT CTT ATC GCA AAA ATT CCA ACG ATT GTG GCC GCC TTT CAC CGC ATT CGC CAA GGT CAA GAT CCC GTT GCG CCG CGT CCA GAC TTA TCA CAT GCT GCC AAT TTT CTT TAT ATG CTT AAC GGG GAG GAA CCC AGT CCT GTG GAG GCT AAA GTT TTC GAC GTG GCC CTT ATC CTT CAC GCA GAC CAT GAA ATG AAT GCA AGC ACG TTC GCC GCA CGC GTA GTG GCG TCG ACT TTA AGT GAC ATG TAT AGT GCG ATT ACG GCT GCC ATT GGT GCC TTG AAA GGT CCA CTT CAC GGA GGC GCG AAC GAG GAG GTC ATG AAA ATG TTG CAA GAA ATT GGC TCG CCA GAT AAA GCA GAG CCA TGG GTC CAA GAA AAA CTG GCG CGT AAG GAA CGC ATT ATG GGT TTT GGT CAC CGC GTA TAC AAG ACT TAT GAT CCG CGC GCG AAA ATC CTG AAA AAA ATG GCG AAG CAG TTG GCA GAA AAG CAT GGC GAT ACG AAG CTT TAC GAG ATC GCC GAA GCG GTC GAA AAG GTC GTT GTT GAG CGC CTT GGG CCG AAA GGC ATC TAT CCC AAC GTG GAC TTT TAC TCG GGT ATC GTT TAC TAC TCA ATG GGT ATT CCT ACA GAT TTG TTC ACT CCT ATT TTT GCA ATG GCT CGC ATC GCT GGC TGG ACG GCT CAT ATC TTG GAA CAG TTA GAA GAC AAC CGC CTT ATT CGT CCT CGT GCT GTT TAT GTG GGA CCT ACA GAC CGC AAA TAT GTC CCC ATC GAT CAA CGT
Anc 2/3	ATGGCATCTATGGAGTATACCCCTGGGTTGGCAGGAGTGGTGGCGGCTGAATCCTCAATTTCTTATATT GATGGACAGGAGGGTATTTTGCGTTATCGCGGGTATCCAATCGAAGAGCTTGCGGAGCACTCAACCTT CGAAGAGGTAGCATACCTGTTGCTGTACGGCGAATTACCCACGCGCGACGAGTTAGATGAATTTGATC ACGAACTGAAGCATCATCGCGCGTTACCAGAACAAATCATTGACCTTTTGAAAAATCTGCCGAAAGATG CCCATCCTATGGCCGCATTGCAGTCCGCCGTGGCGGCCTTAGGGATGTTTTATCCAGACGCAGACGAT ACGGATCCGGAGGGAAATTATGAAGCGGCAGTGCGTTTGATTGCAAAGTTACCAACGATCGTTGCTGC GTTTCACCGTATTCGTCGCGGACAGGACCCAGTTGCTCCTCGCGACGATCTTAGTCATGCGGCCAACT TCTTATACATGTTAAACGGCGAAGAACCAGACCCGCTTGCCGCACGCGTCTTCGACGTTTGCTTAATTC TGCATGCCGACCACGAAATGAACGCATCTACCTTTACTGCACGTGTAGTTGGCTCAACCTTGGCCGATC CTTATTCAGCAATCGCCGCAGCTATCGGCGCTCTTTCGGGGCCTTTACACGGTGGAGCTAACGAGGAA GTGTTAAAGATGCTGGAAGAGATTGGAAGTCCGGATAACGTGGAGCCGTGGCTGGAGGAAAAGTTGGC TCGTAAGGAGCGTATCATGGGCTTCGGACACCGTGTTTATAAAACCTATGACCCCCGCGCCAAAATTCT TAAAAAAATGGCCGAGCAATTGTTCGAAAAACACGGGTCAACGCCTTTATATGACATTGCCCAAGAACT TGAGAAGGTTGCAGCTGAACGTCTGGGGCCCAAGGGGATTTACCCTAACGTCGACTTCTACAGCGGAA TTGTATATCAGTCAATGGGGATCCCAACGGACTTATTTACGCCTATTTTCGCGATCGCGCGTATTGCCG

	GTTGGACAGCGCATTGGTTAGAGCAACTTGAAGACAATCGTATCTTCCGTCCTCGTCAGATCTATGTGG GCCCCACTGACCGCTCTTACGTACCCATCGATGAGCGCCTCGAGCACCACCACCACCACCACTGA
Anc 3a	ATG GCG AGT ATG GAA TAC ACA CCC GGT TTA GCC GGC GTC GTT GCC GCG GAG TCC AGC ATC TCG TAC ATC GAT GGC CAA GAA GGT ATC TTA CGC TAC CGT GGA TAC CCT ATC GAG GAG CTT GCG GAA CAC AGC ACA TTC GAG GAA GTC GCA TAC CTG CTT TTG TTT GGG GAA CTG CCT ACG CGT GAC GAG CTT GAA GAG TTT GAC CAC GAG CTG AAG CAC CAT CGT GCT TTG CCA GAG CGC ATT ATT GAC TTA TTG AAA AAT CTG CCT AAG TCG GCT CAC CCA ATG GCC GCC CTG CAA TCG GCA GTC GCA GCT TTA GGG ATG TTC TAT CCT GAT GCA GAT GTA ACA GAT CCC GAG GGC AAT TAC GAG GCA GCA GTT CGC CTG ATT GCT AAA TTG CCA ACT ATT GTA GCT GCT TTC CAC CGT ATT CGC CGC GGC CAA GAC CCT ATC GCA CCC CGT GAC GAC CTT GGT CAC GCT GCT AAC TTT CTT TAC ATG CTG AAT GGT GAG GAA CCG GAT CCT CTT GCT GCA CGT GTC TTC GAT GTC TGT CTT ATT TTG CAC GCA GAG CAC TCT ATG AAC GCT TCC ACT TTC ACA GCC CGT GTA GTT GGA TCT ACG TTG GCA GAC CCA TAC TCG GCG ATC GCT GCC GCG ATT GGC TCA CTG AGT GGT CCC TTG CAT GGC GGC GCG AAT GAA GAA GTC TTA CAG ATG CTG GAA GAA ATC GGA AGC CCT GAT AAT GTC GAG CCT TGG TTG GAG GAA AAA TTA GCG CGT AAA GAA AAA ATC ATG GGA TTT GGC CAT CGT GTC TAC AAG GTG AAA GAC CCT CGT GCT ACC ATT TTA CAG AAG ATG GCA GAA CAA CTT TTT GAG AAA CAT GGC TCA ACT CCG TTA TAT GAC ATC GCC CTT GAG CTT GAG AAG GTA GCG GCA GAG CGT TTG GGC CCC AAA GGC ATT TAT CCT AAT GTC GAT TTC TAT TCA GGA ATT GTA TAC CAA AAA ATG GGG ATC CCA ACT GAT CTG TTC ACA CCG ATT TTT GCA ATC GCA CGC GTT GCC GGA TGG ACC GCG CAT TGG TTA GAG CAA TTG GAA GAC AAC CGT ATC TTT CGT CCC TCA CAG ATT TAC GTG GGC CCA ACT GAT CGT TCA TAT GTA CCT ATT GAC GAA CGT
Anc 3b	ATG GCG GCC ATG GAA TAC ACT CCA GGC CTT GCG GGA GTT CCA GCT GCG GAG TCA AGC ATC TCC TAT ATC GAT GGC CAA CGT GGT ATT TTG CAA TAC CGT GGC TAC CCC ATT GAG GAA CTG GCT GAG CAT TCT ACA TTC GAG GAG ACA GCC TAT CTT CTT TTG TTC GGG GAA CTG CCA ACC AAA GAT GAA TTA GAG GAG TTC GAT CAC GAG CTG AAG CAC CAC CGT CGC CTT AAA TAT CGC ATT ATC GAC CTG TTA AAA AAT TTA CCT GAA AGT GGG CAT CCT ATG GAC GCC CTT CAA GCC GCG GTT GCT GCT CTG GGT ATG TTC TAC CCT GGG CGC GAC GTT ACA GAT CCT GAA AAT AAT TAC GAA GCT GCT GTT CGC CTT ATC GCT AAG CTT CCC ACC ATC GTG GCT GCG TTC CAC CGC ATT CGT CGT GGC GAT GAT CCT ATC CCT CCC CGC GAT GAC CTG GGA CAC GCC GCG AAT TTT TTA TAT ATG TTG AAT GGC GAA GAG CCT GAT CCG CTT GCC GCT CGT GTA TTC GAC GTC TGC CTT ATT TTA CAT GCA GAG CAT AGC ATG AAC GCC TCC ACC TTC ACT GCT CGT GTA ACT GGC TCG ACT CTG GCC GAC CCA TAC TCG GTT GTA GCT TCT GCC ATC GGT AGC CTG TCC GGC CCG CTT CAC GGT GGG GCA AAC GAG GAG GTG CTT CAG ATG TTA GAG GAG ATT GGC TCC CCT GAG AAT GTA CGT CCT TGG TTG GAG GAG AAA TTG GCA CGT AAG GAA AAA ATT ATG GGT TTC GGA CAC CGC GTG TAT AAG GTA AAA GAC CCC CGT GCG ACC ATC TTA CAG AAA TTA GCT GAA CAA TTA TTT GAA AAA TAT GGC TCA ACA CCA TTA TAT GAC ATT GCT CTT GAA TTG GAA CGT GTC GCC GCG GAG CGT TTG GGA CCG AAA GGC ATC TAT CCG AAT GTC GAT TTC TAT AGT GGC ATC GTT TAC CAG AAG ATG GGT ATC CCC ACG GAT TTA TTT ACG CCC ATC TTT GCC ATC GCT CGT GTA GCG GGA TGG CTG GCG CAC TGG CTT GAA CAG CTG GAA GAC AAC CGT ATC TTT CGT CCG TCT CAG ATC TAT GTC GGC GAG ACC GAT CGT TCA TAC GTC CCT ATC GAT GAG CGC
Anc 2	ATG AGT ACC TCG AAG AAC ACC TTA ACG CTT ACG GAT AAC CGC ACT GGT AAG ACC TAC GAG TTA CCC ATT GAG GAC GGT ACT ATC CGC GCA ATT GAC TTA CGC CAA CTG AAA GTG GAT TTT GGG TTG ATG ACG TAT GAT CCA GGC TAC ACC AAT ACT GCC TCG TGT AAA TCA GCA ATT ACC TAC ATT GAT GGA GAT AAA GGC ATT CTG CGT TAC CGC GGG TAC CCC ATC GAG CAA TTG GCA GAG AAA AGC TCG TTC CTT GAG GTG GCC TAC TTG CTT ATC TAT GGC GAG TTG CCT ACA AAG GAA CAG CTG GAT GAA TTT GAG TAC GAG ATT ACG CAT CAT ACT TTT ATC CAT GAG AAT ATT AAG AAA TTC ATG GAC GGG TTC CCA CGT GAC GCG CAT CCC ATG GCC GTT CTG TCT TCC GCA GTG GGG GCC TTA TCG ACC TTC TAT CCC GAC GCA AAG GAC ATT AAT GAT CCA GAG CAA CGC CAC CTG GCG ATC GTA CGC TTG ATT GCC AAG ATG CCT ACA ATT GCC GCG TTC GCC TAT CGT CAT TCA ATC GGG CAG CCT TTC GTT TAT CCG GAC AAC GAT TTG TCC TAT GCA GAG AAC TTT TTG TAC ATG ATG TTT GCC ATG CCT GAA GAG AAG TAT GAG CCC GAT CCT GTT TTA GCC CGT GCC CTG GAC GTT CTT TTC ATT TTA CAT GCG GAC CAC GAA CAG AAT TGC AGT ACT AGT ACA GTT CGC CTG GTC GGT TCT TCC CAG GCG GAC CCA TAT AGT GCT ATC GCC GCT GCC ATT GCC GCT TTA TAC GGC CCG CTG CAT GGT GGC GCT AAT GAG GCA GTT TTA CGT ATG CTG GAG GAA ATC GGC TCG GTA GAC AAC GTG CCT GAA TTT ATT GAA CGC GCC AAA GAC AAG AAA GAC CCT TTC CGT CTG ATG GGG TTC GGA CAT CGT GTA TAC AAG AAT TAC GAT CCC CGC GCG AAA ATC ATC AAG AAG ATG GCA CAT GAG GTT TTT GAA AAA ACA GGT AAG AAC CCT CTT CTT GAC ATC GCA ATG GAG TTG GAG CGT ATC GCC CTT GAA GAC GAG TAC TTC GTT GAG CGT AAG TTG TAT CCG AAC GTA GAT TTT TAC TCA GGG ATT ATT TAT CAG GCG ATG GGC ATT CCG ACC GAC ATG TTC ACC GTC TTA TTC GCA ATC GGA CGC ACT CCA GGT TGG ATT GCT CAG TGG AAG GAA ATG CTG GAG GAC CCA GAA CAG AAG ATT GCT CGC CCC CGC CAA ATC TAC GTG GGA GAG ACG GAG CGT GAC TAC GTA CCT ATC GAT CAA CGT
Anc 2a	ATG AGT ATG TCG AAG AAT ACC TTA ACG ATT ACT GAC AAC CGT ACT GGA AAG ACC TAC GAG ATC CCG ATC GAG GAC GGG ACG ATT CGT GCA ATG GAC TTG CGC CAA ATC AAA GTA GAC GAC GAT GAT TTC GGG TTG ATG AGC TAC GAC CCG GGT TAT ATG AAT ACT GCC TCT TGC AAG TCC GCA ATT ACT TAC ATC GAT GGC GAC AAG GGC ATT CTT CGC TAT CGT GGG TAT CCC ATC GAG CAA TTG GCC GAG AAA TCG TCG TTC CTT GAA GTA GCC TAT CTG CTT ATT TAT GGG GAA TTG CCC ACC AAG GAA CAA TTG GAT GAA TGG GAG TAT GAG ATT ACA CAT CAT ACA TTC ATC CAC GAG AAC ATT AAA AAA TTT ATG GAC GGC TTT CGC TAC GAC GCC CAC CCA ATG GGG ATG TTG GTG AGC TCT GTA GGT GCG CTG TCA ACC TTC TAT CCT GAC GCA AAG GAC ATT AAT GAT CCA GAA AGC CGC CAC TTA CAG ATC GTC CGC TTA ATC GCT AAG ATT CCT ACT ATC GCG GCT TTT GCT TAT CGT CAT TCC ATT GGA CAG CCT TAT GTC TAT CCC GAC AAT GAT CTT TCT TAC GCG GAA AAC TTT TTG TAT ATG ATG TTC AAA ATG ACC GAG TTG AAG TAC GAG CCG GAC CCC GTT TTG GCT CGC GCG TTG GAT GTA TTA TTC ATT CTT CAC GCA GAC CAT GAA CAA AAC TGC TCA ACG AGC GCT ATG CGC TCT GTC GGC TCC AGT CAA GCG GAC CCG TAT TCA GCG GTG GCT GCT GCA GTG GCG GCG TTA TAT GGC CCA CTT CAT GGC GGG GCT AAC GAG GCA GTA CTT CGC ATG CTT GAG GAA ATT GGT TCT GTA GAC AAC GTC CCT GAG TTT ATT GAG CGC GTG AAG GAA AAG AAG GGG GAA TTC CGT CTG ATG GGT TTT GGG CAT CGC GTG TAT AAG AAT TAC GAC CCA CGT GCC AAG ATT ATT AAG AAA ATC GCA CAT GAG GTC TTT GAG GTT ACG GGC AAG AAC CCA TTA TTG GAT ATT GCA GTG GAA TTA GAA CGC ATT GCC CTG GAA GAT GAA TAC TTC GTG AGC CGC AAA TTG TAC CCA AAC GTT GAT TTC TAC TCC GGC ATT ATT TAT CAG GCT ATG GGA TTC CCA

	ACC GAT ATG TTT CCG GTC CTG TTT GCT ATT GGT CGC ACG CCT GGG TGG TTA GCC CAG TGG CAA GAG ATG TTA GAA GAT CCA GAG CAG AAA ATC GCT CGC CCT CGC CAG ATT TAC GTC GGA GAA GAC GAG CGC GAT TAC GTG CCC ATC GAT CAA CGC
Anc 2b	ATG TCC TCG AAG AAG GCG ACA TTG ACG TTA ACT GAT AAC CGC ACG GGC AAA ACA TAT GAG TTA CCG ATC TTA GAC GGA ACA GTA GGT CCC AAG GCT ATT GAT ATT CGT AAA TTA CGT GCT GAT ACT GGA TAT ATG ACG TTC GAT CCA GGG TAC ACG AAT ACG GCA AGT TGC AAG TCG GCC ATT ACC TAT ATT GAT GGA GAC AAA GGC ATC TTG CGT TAT CGC GGT TAC CCA ATT GAG CAG CTG GCC GAA AAA TCG TCA TTT CTG GAG GTC GCC TAC CTG TTA ATC TAT GGT GAA TTG CCG ACT AAA GAA CAG TTA GAC GAA TTT GAG CAC GAG ATC ACC CAT CAT ACA TTC ATC CAT GAA AAC TTG AAG AAG TTT TTC GAC GGC TTC CCT CGC GAT GCA CAC CCT ATG GCT GTG CTG TCG TCG GCT GTT GGA GCA TTG AGT ACA TTT TAC CCT GAC AGC TTG GAC ATC AAT GAC CCC GAA CAG CGC AAT TTG GCA GTT GTG CGT TTG ATC GCG AAG ATG CCC ACA ATC GCC GCT TTC GCA TAT AAG CAC AGC ATC GGA CAG CCG TTC GTC TAT CCG GAT AAT GAT CTG AGC TAC GCG GAA AAC TTT CTG TAC ATG ATG TTC GCG GTC CCG GCC GAG GAA TAT GAA GTT GAT CCA GTC CTT GCG AAA GCA CTG GAT ATG TTG TTT ATC CTT CAT GCA GAC CAC GAA CAG AAT TGC AGT ACT TCC ACG GTT CGT CTG GTT GGC TCC TCA CAA GCT AAC CCG TAT GCG GCC ATT GCG GCG GGA ATC GCT GCG CTT TGG GGG CCC TTA CAT GGC GGG GCT AAC GAA GCT GTC TTG AAA ATG CTT GAG GAG ATC GGT TCC GTA GAC AAC GTC CCC GAG TTT ATT GAG CGT GCT AAA GAT AAG AAC GAC CCC TTC CGC TTG ATG GGT TTT GGA CAT CGT GTT TAC AAA AAT TTT GAT CCC CGC GCT AAA ATC ATC AAA AAG ACG GCA GAT GAA GTA TTG GAA AAA CTT GGT ATT GAT GAT CCA CTT TTG GAT ATC GCC ATG GAA TTG GAG AAA ATT GCA CTG GAA GAT GAA TAC TTT GTT GAG CGC AAG CTG TAC CCA AAT GTT GAT TTT TAC AGT GGC ATT ATT TAT CAG GCT ATG GGT ATC CCT ACA GAC ATG TTC ACG GTC CTG TTT GCA ATT GGC CGT ACT CCG GGC TGG ATC GCG CAA TGG AAG GAA ATG TTG GAA GAC CCA TCC CAA AAA ATC GGA CGT CCT CGC CAA ATT TAC ACG GGC GAA ACC GAG CGT GAC TAT GTC CCT ATT GAT CAG CGT
Anc 2c	ATGTCCGAAACAGCTACGCTTACTCTTGATGGCAAGACATACGAACTGCCCATTATTGAGGGAACGGAA GGTGAAAAAGCAATTGACATTCGTAAATTACGCGCTGAAACGGGTTATATCACCTTTGATCCTGGATAC GCCAATACGGCCTCGTGCAAGTCCGCCATCACATATATTGACGGGGATAAGGGAATCCTGCGTTACCG TGGGTACCCTATCGAGCAGTTAGCGGAGAAGTCCTCTTTCCTGGAAGTAGCCTATTTATTGATTTATGG TGAATTGCCCACAAAAGAACAATTAGACGAATTCCAAAATGAGATTACCCATCATACGCTGATTCACGAG GATTTAAAGAAGTTTTTCGATGGCTTCCCACGCAACGCGCACCCAATGGCAGTGTTATCCTCGGCGGTT TGCGCGCTTTCGACCTACTACCCGGACAGTCTTGACCCTAACGACCCTGAGCAAATTAACCTGTCGATC GTTCGCCTTCTTGCTAAACTGCCAACGATCGCTGCCTTCGCCTACAAAAAATCCATTGGTCAACCGTTT GTTTACCCCGACAATGATCTTTCCTATTGCGAAAATTTCCTGTACATGATGTTCGCGGTTCCCGCCGAAC CCTACGAAGTCGATCCTGTCGTAGCTAAAGCGTTGGATATGTTATTCATTTTGCACGCAGATCATGAGC AAAATTGCTCCACGAGTACTGTACGCCTTGTCGGGTCTTCGCAAGCTAATCTTTATGCAAGTATCAGTG CTGGGATTTCGGCACTTTGGGGGCCGTTGCATGGAGGAGCAAACCAGGCGGTTCTGGAAATGTTGGAA GAGATTCACGCAGATGGAGGGGACGTTAAGGAATTCATTGAGAAAGCGAAGGATAAGAATGATCCATT CCGTTTAATGGGATTTGGTCATCGTGTATATAAGAATTTTGACCCGCGTGCGAAAATCATCAAGAAAACA GCCGATGAGGTGTTAGAAAAGCTGGGGATTAATGATCCTCTTCTGGACATTGCGAAAGAATTAGAAGAA GTGGCGCTTAACGACGAATATTTCGTCGAACGCAAATTATATCCGAACGTCGATTTTTATTCAGGTATCA TCTATCGCGCCATGGGGATCCCGACGGATATGTTCACCGTTCTTTTTGCATTAGGACGCTTGCCGGGAT GGATCGCCCAGTGGAAAGAAATGCGTGAGGATCCGTCGACTAAAATCGGGCGTCCTCGTCAGATCTAT ACAGGAGAGACCGAGCGTGACTATGTGCCCATTGATCAACGCCTCGAGCACCACCACCACCACCACTG A
Anc 2d	ATGGCAGCTAAGAAGGCTACAGTTACGTTAACAGATAATCGTACTGGCAAATCATATGAGTTGCCGATT TTGGATGGGACGGTCGGACCTGATGTAATCGACATTCGTAAACTGTATGCTGACACTGGCTACTTTACA TTTGACCCGGGGTACACATCTACGGCGTCTTGCAAATCAAAGATTACCTACATTGATGGGGACAAGGGA GTGTTGCGTTATCGCGGGTATCCGATCGAACAGTTGGCCGAGAAGAGTTCATTTTTAGAGGTTTGCTAT CTTTTGTTAAATGGGGAATTGCCTACCAAGGAACAATTAGAGGAGTTTGAACACGAGATTACACATCATA CATTTCTTCATGAGAATTTGCGTAAATTCTTTGATGGATTCCCGCGCGACGCGCACCCGATGGCGGTAT TAAGTTCTGCTGTAGGGGCACTTTCAGCATTCTACCATGACAGTCTTGACATCAATGACCCGGAACAAC GCAATATTGCAGCTCATCGCCTGATTGCAAAGATGCCCACAATCGCAGCGTTCGCATACAAACACTCGA TTGGACAGCCGTTTGTTTATCCGGACAACGACTTGGGTTACGCCGAGAATTTTTTATACATGATGTTTGC TGTGCCCGCTGAGGAATACGAAGTCAACCCGGTTCTGGCACGCGCTTTGGACACTATCTTTATTCTGCA CGCGGACCATGAACAAAACGCCAGTACGTCGACGGTGCGCTTGGTCGGGAGCTCAGGCGCAAACCCT TATGCGGCCATTGCAGCTGGCATCGCCGCACTTTGGGGACCCGCGCATGGTGGTGCAAATGAGGCTG TGTTAAAAATGCTGGAGGAGATCGGGAGTGTTGACAATGTACCCGAATTTATCGAACGTGCTAAGGATA AAAATGACCCCTTCCGTTTGATGGGTTTTGGTCACCGCGTATACAAGAATTTCGACCCACGCGCAAAGA TTATGAAGAAAACAGCGGATGAAGTATTAGAAGAACTGGGCATCGATGACCCGCTTCTGGACATTGCTA TGGAACTTGAGAAGATCGCGCTGGAGGACGAATACTTCGTAGAGCGTAAGCTGTATCCTAATGTAGACT TTTACTCTGGAATCATTTTAAAGGCGATGGGTATCCCAACCTCGATGTTTACTGTGTTATTCGCTATTGG TCGCACCCCTGGCTGGATTGCTCAGTGGAAGGAAATGATCGAAGATCCATCACAAAAAATCGGCCGTC CGCGTCAGTTATACACTGGTGCAACAGAGCGCGATTATGTACCCATTGACCAACGCCTCGAGCACCAC CACCACCACCACTGA
***Archaeoglobus veneficus* WP_156786146.1**	ATG CAC GAC GGG CTT CAA GAC GTC CTT GCA TGC AAG TCA TCC ATT TGT CGT ATC GAG ATG ATT GAT GGA AAG GCG ATT TTA GAG TAC CGT GGG TAC GAT ATT CAT GAG TTG GCC CAG CAT AGC ACT TTC GAG GAA GTT GCC TAC TTA TTA TTA TTC GGT GAA TTG CCA ACC AAA TCC GAA CTG GAG GCT TTT TCC GAA GAG TTA AAG GAG TTA CGC GAA TTG CCG CCC CAG ATC ATC GGA CTG TTA ACG CAT TTA TCC CCT TTT TCA CAC CCT ATG GTC GTC CTG CGC ACT GCT ATC TCT TAC CTG GGT ACT ATG GAT AAA CAC ATC CAC CAT AAA TCA CAT GAA AAG TCT TTG CAA AAG GCT AAA TCC CTG ATT GCT AAG TTC CCA ACA ATT GTC GCT TAC TTC CAT CGT ATC CGT TCA GGT CAA AAC TTA GTC CAC CCG TCA GAT GAA TTG AGC CAC GCT GCT AAT TTT TTA TAT ATG CTG CAC GGC GTA GAG CCT ACT AAA ACG GAA GCT AAA GCG ATG GAC CTG GAT TTA GTT TTA CAC GCA GAC CAT GAA CTT AAT GCT TCG ACG TTT GCA GCC CGT ATT GCT GCA TCG ACT TTA GCC GAC ATT TAT GCG TGC GTG GTC GCC GCA ACG GGA ACA CTT ATG GGA CCC TTG CAC GGG GGA GCA GCG CAG AAA GTT ATG GAA ATG TTG CGT GAG ATC GCG GTG CCT TGG CGT GCC GAA GAG TAT GTC AAG ATG AAA TTG GAA CGT GGA GAA CGT ATC ATG GGA TTC GGT CAT CGT GTA TAC CGT GGT GTC GTA GAC CCC CGT ACA ATC GAA CTG CGC GCA CTG GCC GAG AAG CTG GCT AAG GAA AAG GAG CCA AAG TGG TTC GAA ATT TCC CGC GCT GTC GAA GAA GCT GTA TAC AAA TAT AAG GGA CTT TTT CCC AAT GTG GAC TTT TAC TCT GCT TCC GTG TAT GCA AAT CTT GGC

	ATC CCC GAT GAC TTG TTC ATT AAT ATT TTC GCC ATC TCT CGC ATT TCT GGG TGG ACC GCA CAC ATC ATT GAA CAG TAT GAG AAT AAC CGC TTA ATT CGC CCA CGC GCA TTC TAC GTG GGA GAG GCT GGC CGT AAG TAT GTG CCA ATC GAT CAG CGT GGA GAT
Bacillus subtilis **WP_004398810.1**	ATG ACA GCG ACA CGC GGT CTT GAA GGG GTT GTA GCA ACA ACA TCA TCT GTT AGT TCT ATT ATT GAT GAT ACC CTT ACA TAT GTG GGG TAT GAT ATC GAT GAT TTG ACA GAG AAT GCA AGT TTT GAA GAA ATC ATC TAC TTG CTT TGG CAT CTA AGG CTG CCA AAC AAA AAA GAG CTT GAA GAG TTA AAG CAG CAG CTT GCA AAG GAA GCT GCC GTT CCG CAA GAA ATT ATT GAG CAC TTC AAA TCT TAT TCC CTC GAA AAT GTT CAT CCC ATG GCT GCT CTT CGG ACA GCA ATT TCA TTA CTA GGC CTG CTT GAC AGC GAG GCT GAT ACC ATG AAC CCG GAA GCC AAC TAC AGA AAG GCG ATC CGC CTT CAG GCA AAG GTG CCT GGC CTG GTT GCA GCG TTT TCC AGA ATT CGC AAG GGT CTT GAG CCT GTT GAA CCG AGA GAA GAT TAC GGT ATT GCG GAG AAT TTT CTT TAC ACA TTA AAC GGT GAA GAG CCT TCA CCA ATT GAA GTT GAA GCC TTC AAC AAA GCA TTG ATT TTA CAT GCT GAC CAT GAG CTG AAC GCG TCA ACG TTC ACA GCG AGG GTC TGC GTC GCA ACA CTT TCT GAC ATC TAT TCA GGC ATT ACT GCA GCA ATC GGC GCG CTG AAA GGG CCT CTT CAC GGC GGA GCG AAT GAA GGC GTT ATG AAA ATG CTG ACA GAA ATC GGT GAA GTG GAA AAC GCT GAG CCG TAC ATC CGC GCT AAA CTT GAA AAG AAA GAA AAA ATC ATG GGA TTC GGC CAC CGT GTG TAT AAG CAT GGT GAT CCG CGC GCG AAG CAT CTG AAA GAA ATG AGC AAG CGC CTG ACC AAC CTG ACA GGC GAA AGC AAA TGG TAT GAG ATG TCG ATT CGT ATA GAA GAC ATC GTC ACA TCA GAG AAA AAG CTT CCG CCT AAC GTT GAT TTC TAT TCG GCT TCT GTA TAT CAC AGC CTC GGT ATT GAC CAT GAT TTG TTT ACA CCG ATT TTT GCT GTA AGC AGA ATG TCC GGC TGG CTC GCT CAT ATT CTT GAG CAG TAT GAC AAC AAC CGT CTG ATC CGC CCG CGG GCA GAT TAC ACA GGC CCT GAC AAA CAA AAA TTC GTT CCG ATT GAA GAA AGA GCC CTC GAG CAC CAC CAC CAC CAC CAC TGA
Deinococcus pimensis **WP_027481744.1**	ATG ACA ACC ACA ATC GCC AAA GGG CTT GAA GGA GTG TTA TTT ACA GAA ACG AAA CTG ACG TTT ATT AAC GGG ACC GAG GGC GTC CTT ACA CAT CTT GGC ATT CCC ATC CAA GAG TGG GCG GAA AAA TCT ACC TTC GAA GAG TTG TCA TTC GCG CTT CTG CAT GGG AAG CTT CCT ACA CGT GAA GAA CTG GCT GCA TTC GAT GCC GAC TTA AAG GCT AAC CGC GCT GTG CCT GCT GAA CTT ATC ACT GCG ATT CGT GCA TAC CCC CGC GAT GTC CAT CCC ATG CAA GCG CTT CGT ACT GCT GCG TCA CAT CTG GGG TTA TTG GAT GCG GAC GCG GAA GAT ACA TCC GCA GAA GGG CGT TAC CGT ATT GGC GTC CGT ATG ATT GCT CAG TTC TCG ACG ATT ATT GCA GCA ATT GCG CGC GCA CAA GAC GGA CAA GAC CCA GTG GAG CCT CGC GCT GAT CTG ACT CAC GCG GGC AAT TTC TTG TAT ATG TTG AAC GGT CAA GAG CCC TCA GCC GAA CAG GCC CGC CTG TTT GAT ATT GCA CTT ATT CTG CAC GCG GAC CAT GGT ATG AAC GCC TCG ACA TTT ACA GCC ATT GCG ACG GCC AGC ACT TTA TCA GAT ATG TAC AGT AGT GTC GTC TCT GCT ATC GGA GCC CTG AAA GGT CCT CTG CAT GGA GGT GCT AAC GAA GCA GTC ATG GAC ATG CTT GAC GAG ATT GGA ACT GTT GAT GCG GCA GTA CCA TTC ATC ACG CAG AAG TTA GAT AAC AAG GAA AAA ATC ATG GGG GTC GGC CAC CGC GTG TAC AAG TAC TTC GAC CCG CGC AGT CGC GTC TTA CGC GAC TAT GCT GCT CAT GTC TCC GAA AAA CAA GGA AAG TCC CAC TAC TAT TCT ATT TTA GAA ACA ATT GAA AAG GAA ATC GTC GAA CGC CTT GGG GGT AAA GGG ATC TAT CCG AAC GTT GAT TTC TAT TCC GGA GTT GTA TAC TCG GAT CTG GGG ATT AAA AAG TCG TTT TTT ACC CCC ATT TTC GCT CTT GCC CGC ATT TCA GGC TGG GTC GCT AGT GTA ACA GAG TAC ACT GCC AAC AAC CGT TTA CTT CGC CCC GAT GCG GTT TAC GTA GGC GCG CGC GAC CAA CAC TAT GTC GAG ATT GAC GAG CGC AAG
Deinococcus radiodurans **WP_027479989.1**	ATG TCA AAT ATC GCC AAA GGC TTA GAA GGC GTG CTG TTT ACC GAA AGT AAA CTT ACA TTT ATT AAT GGG TCA GAG GGT ATC TTG ACA CAC CTT GGA ATC CCT ATT CAG GAA TGG GCT GAG AAA AGC ACT TTC GAA GAG TTG AGC CTT GCG TTG TTG GAT GCC AAA CTT CCA ACC GCG GAA GAA TTA GCC AAA TTT GAC GCG GAA CTT AAA GCG AAC CGT GCG ATT CCG GAC CAG TTA GTC GGA ATC ATC CGT GAT ATG CCT AAG GGG GTT CAC CCT ATG CAA GCA TTG CGT ACC GCA GTT AGT TAC TTA GGA CTT CTG GAC CCG CAG GCG GAA GAC ATC ACT CCA GAA GCA CGT CGC GCT ATT AGT ACC CGC ATG ATC GCG CAG TTT AGT ACT ATC ATC GCT GCG ATC AAC CGC GCT CAG GAG GGC CAA GAC ATT GTA GCC CCC CGC GCG GAT CTG ACC CAC GCC GGC AAC TTC TTA TAT ATG TTA ACT GGT AAT GAA CCG ACA CCT GAG CAG GCA CGT CTT TTT GAC ATC GCA CTG GTT TTA CAT GCG GAT CAT GGA ATG AAT GCA AGC ACG TTC ACA GCT ATT GCG ACT TCG TCC ACA TTG AGT GAT ATG TAC AGT TGC ATG GTT TCA GCG ATC GGG GCT TTG AAG GGT CCG CTT CAT GGC GGT GCC AAC GAA GCA GTA ATG ACT ATG TTA GAC GAG ATT GGA ACA GTA GAC AAG GCT GAA GCG TAC ATT ACC GGA AAG CTT GAT AAC AAG GAA AAA ATC ATG GGT GTT GGC CAT CGC GTC TAT AAG TAT TTT GAT CCT CGT AGC CGT GTG CTT CGT GAC TAC GCG GAG CAT GTT GCA AAT AAG GAG GGT AAA TCC AAT TAC TAC CAG ATT TTG GAA GCG ATT GAG AAA ATT ATT GTA GAT CGC ATG GGC GCG AAG GGA ATT TAT CCA AAT GTT GAC TTC TAC TCT GGT ACC GTA TAT TCT GAC CTT GGG ATT AAA AAG GAG TAC TTC ACT CCG ATC TTC GCC TTG GCG CGT ATT TCA GGT TGG TGT GCT TCA GTG ATT GAG TAC TCA CAA GAT AAT CGC CTG CTT CGC CCC GAC GCA GAA TAT ACT GGT GCG CGT GAT CAG CAT TAC GTC GAC ATC AAA GAC CGC CAA
Nitrosopumilus maritimus **WP_012215482.1**	ATG GAG ACG AAG AAT ATC GGA TTG CGT GGT ATT GAA GTT GCG GAT ACT CGT ATC TCC AAT ATT GAC GGA GAA AAA GGG AAA CTG ATC TAT CGT GGG TTC GAT ATC CTT GAC TTA ACG AAG AAT AGC ACA TTC GAA GAA ACC GCT TAC CTG CTT CTT TAT GAT AAC CTT CCT ACC AAG GCT CAG TTG GAC GAA TTC AAC CAA AAA TTA GTG GAG GCG CGC TAT ATC CCG AAA CAA ATG CAG AAG AAC ATG GGG AAC TGG CGT AAA GAC GCG GAC CCT ATG GAT ATG CTG CAG GCA TTC GTA TCG GCT TTG GCA GGC TAC TAC GAT GAA GAG TTC TCA AAC AAG GAA GCG AGC TAC GAG AAA GCA ATT AAC CTG ATT GCA AAA GTA CCC ACA ATC ATT GCG AGC TGG CAA CGT ATC CGT AAT GGC TTA CCG ATT GTT GAT CCG GAT TCG TCG CTG AGT CAC GCA GCG AAT TTC TTA TAT ATG ATG AGT GGC GAA AAA CCT GAT CCG GAG GTC GAA AAA GTA TTT GAT GTG TGC TTA ATC CTT CAT GCT GAC CAC ACA TTT AAC GCC TCC ACA TTC ACC GCC CGC CAA GTC GCC TCC ACA CGT GCG CAT ATG TAC TCT GCA TCT TCT GCA GCT ATC GGT GCC TTA TCG GGG GAG TTA CAC GGG GGC GCT AAC ACA GAG GTG ATG AAG ATG TTA CTT GAA ATT AAA GAG ATT GAT AAA GTA GAA CCA TGG ATC AAG GAA AAA ATG AGC GCG GGC GAG CGT ATT ATG GGG ATG GGG CAC GCG GTC TAC CGC ACT TAT GAC CCA CGC GCC CAA GTG TTG AAG GAG TTA TCG CGC AAG TTA GCT GAA AAA ACA AAA GAA CCC TGG TTC GAT ATG ACG GAA AAA GTA GAA ACT ACG ACC ATC AGC GAA ATG AAA GCA CAG AAG GGG AAA GAC ATC TAC CCT AAT GTG GAC CTG TAC TCT GCC TCG ATC TAC TAC ATG CTG AAG ATT CCC GTT GAC CTT AAC ACG CCT ATT TTC GCG ATT TCA CGC GTA GTA GGC TGG GCA GCG CAC ATT ATC GAA GAG AAG TTC GCC GAA GCT GCC CCG AAA CCT GCC CTT TAC CGC CCC AAA GCG ACG TAC GTG GGC AAA TAT TGC GGT CCC GAG GGA TGT GAA TAC AAG ACA TTA GAT TTA CGC AAG

Oceanithermus profundus **WP_013457601.1**	ATG GCT GAG GCG GAA ATT GCC CGC GGG TTG GAG GGC GTT GTA TTC ACC GAG ACT GAA TTA TCA TTC ATT GAC GGT CAG AAT GGG CGT CTG TAT TAC CTT GGT TAT CCT ATC CAA GAA TTG GCT GAG CAT TCA AGT TTC GAA GAA GTT TCT TAC TTA CTG TTA CAT AAA CGC TTG CCC ACG GCC GAC GAG TTG GCC GTT TTT AAG GAA AAA TTA GTT AAA AAT CGT AGT ATC GCG GTT GAG AAG GTA GCA AGC TTT GCT GAC TAT CCG CGT ACA GCG CAC CCC ATG GCG TCG TTG CGC ACC GCC ATC AGT GAG TTA GGA CTG TTT GAC CCA ACT GAA GAA GAT ACT TCA TTT GAA TCA TTG TAC GAG AAG TCA GTT TCA TTG ATT TCG AAG TTT GCC ACA GTG ACC GCG GCG ATC AAG CGT TTG CGC GAG GGT CAT ATG CCA ATC GAA CCC GAC CCC GAA TTA TCG CAC GCA GCG AAC TTC TAT TAC ATG TTA AAC GGC AAA CGC CCG AGT CCT GAA CAA GAG AAG TTG CTG GAT GTG GCT CTT ATC CTG CAT GCA GAA CAT GGA ATG AAT GCC TCC ACG TTT ACT GCG TTA GCC GTA CAT TCG ACG GCA AGT GAC ATT TAT TCG AGT ATC GTT GCC GCG GTT GGC GCG TTG AAG GGA CCG CGC CAC GGT GGT GCG AAC GAA CAG GTT ATG AAA ATG GTT CAA GAG ATT GGT AAA CCC GAA GCT GCT CGT GGT TGG GTG CAA GGG AAA TTA GCC AAT AAG GAG CGC ATT ATG GGC ATG GGC CAT CGT GTA TAC AAA GCT CTT GAT CCC CGT GCT GTG ATC TTG AAA AAA TAC GCG GAA ATT GTC GCC AAA GCT CAT GGA AAG AGT ACA GAA TAC GAA ATT TTG ACA ATT GTA GAA GAA GAA GCG GGA AAG GTT TTG AAC CCC CGC GGA ATT TAC CCC AAT GTT GAC TTC TAT AGC GGT GTT GTA TAT TCA GAC CTG GGA ATC CCG ACA GAA TTT TTC ACA CCC GTA TTC GCC GTG GCC CGT ATC TCT GGC TGG ACT GCA CAT ATC CTT GAG TAC GCT CGT ATG GAT AAT CGC CTG TTA CGT CCT AAG GCC AAA TTC GTA GGG GAG TTA GAC CGT AAA TAC GTT CCA ATT GAT CAA CGC
Thermaerobacter marianensis **WP_013495717.1**	ATG ACA CAA CCA AAT CCA GCC GGG CTG GAA GAT GTG GTC GCT GGA ACG TCG GAA ATC TGC TTT ATC GAC GGG AAA GAA GGT CGC CTG TTA TAC CGC GGC TAT GAT GTC CGC GAC TTA GCG GAA CAT GCA TCT TTC GAG GAA GTG GTT TAC CTG TTA TGG CAC GGA CGT CTT CCC AAT CGT GCC GAA CTT GAC GCC TTC GTG CAA CAA CTT CGC AGT TTG CAG GCT CTT CCG GGC CAG GTT CTT GAT TTG ATC CAG CAT ATC CCT CCT GCC ACA CAT CCC ATG GCT GCT TTG CGT ACG GCA GTA AGT TAT TTA GGT ACG TTG GAT CCC GAC CAA GAG GAC CAG TCT CAG GAG GCC AAT CTG CGC CGT GCG ACG CGC CTT GTC GCT CAA TTT CCG ACG CTT GTC GCC GCG ATT CAG CGC GTA CGT CAG GGT AAG GAG CCG GTC GCA CCG CGT CGC GAT CTG TCA TTA GCG GCA AAT TTT TTA TAC ATG CTT CGT GGA GAA GAG CCC TCC CCT CTG CAT GCG GAC GTT ATG AAC GTT GCA CTT GTG CTG CAC GCA GAT CAC GAA TTA AAC GCC TCT ACG TTT GCA GCG CGT GTT GTC GCG GCA ACG CTT TCT GAC ATG TAC AGC GCC ATT ACG GCT GCG ATT GGT GCC TTG AAA GGC CCC TTG CAC GGT GGA GCA AAT GAG CAG GTT ATG CGT ACT TTG TTA GAG ATC GGT TCA CCG GAT AAA GCT GAG GCC TGG GTT AAG GAA GCA TTA GCC GCA AAA CGT CGC ATC ATG GGC TTT GGG CAC CGC GTC TAC AAG ACT GAA GAC CCG CGT GCT ACG ATT CTG CGC CGT CTG TCC CGT AAA GTA GCA GAG GCT GCC GGG GAT CTT CGC TGG TTC GAA ATT AGC CAG GCC GTG GAA AAG GTT GTG ACC GCG GAG AAA GGC TTG TAT CCT AAT GTC GAT TTC TAC TCC GCC TCA ACT TAT TAC GTT ATG GGA ATC CCG TTC GAA CTT TAT ACC CCA ATC TTT GCG GTT TCC CGC ATC TCA GGT TGG ACA GCA CAT GTC TTG GAG CAG TAT GCA AAT AAC CGC TTA ATC CGT CCA CGC GCT GAG TAT ACT GGT CCG ACA CGT CGC GAA TGG GTG CCC TTG GAA CAA CGC
Thermococcus chitonophagus **WP_068576448.1**	ATG GTA AAC AGT GGT AAA GGG TTG GAA AAC GTC TAC ATT GAT CAA AGC TCA ATT TGT TAT ATT GAC GGT TTT GGG GGA AAG CTG TAT TAT CGC GGC TAT TCC ATT GAG GAA TTG GCC GAA ATG TCC AGC TTT GAG GAA GTA ACT TTC TTG TTG TGG TAC GGG GAG CTG CCT ACC AAG AAG GAA TTG GAA GAG TTT TCA AAA GAA CTG GCT CGT TCT CGT AAG TTA CCT GAG GAA TTA TTA GAC TTA TTG TAC CGT ATT CCA AAA ACC GCG CAC CCC ATG GGA GCA CTT CGT ACC GCG GTC AGC TTC TTA GGG AAT TTG AAC GGC GAA CCA TCA GCC ACA CCG GAA GAC GTG TTC CGC AAA GGG ATC TCG GTG ACC GCT AAG ATT CCC GTT ATC GTC GCT ACT TTA TAC CGT ATC AAG AAT GGT TTA GAT CCT ATC GAA CCG TCC ACG AAG CTT TCC CAC GCG GCA AAC TTT TTA TAC ATG CTT CAC GGC GAA AAG CCA CCT AAA GAG TGG GAG AAA GCA ATG GAC GTT GCT CTT ATT CTT TAC GCT GAG CAC GAG ATC AAC GCT AGC ACC TTG ACT GTG ATG ACA GTC GGC TCC ACG CTG TCC GAT TAT TAT AGC GCC ATT GTG GCC GGT ATT GGT GCT CTT AAG GGC CCC TTA CAT GGC GGA GCC GTG GAG GGA GCG ATT CGT CAA TTT ATG GAG ATC GGT TCC CCC GAG AAG GTA GAA GAG TGG TTC TTC AAT GCG TTG AAA GAA AAG CGT CGT ATC ATG GGT GCA GGG CAC CGT GTA TAT AAG ACT TAT GAT CCT CGT GCA AAG GTG TTC AAA AAA TAT GCA CGT CTT TTA GGA GAT AAG ACC TTA TAC GAG ATT GCC GAA CGT TTA GAG CGT CTG GTG CTG GAA CAT TTA GGA AAG AAA GGG GTG TCG ATT AAC GTC GAT TAC TGG TCT GGA ATT GTG TTT CAC GCG ATG GGT ATC CCC ATT GAG TTG TAT ACT ACC ATT TTT GCA ATG GGT CGT ATC GCT GGA TGG ACC GCT CAT TTG GCG GAG TAT ATC TCG CAC AAT CGT CTG ATC CGT CCC CGT CTT CAG TAT GTC GGC GAA ATT GGT AAA AAA TAC GTT CCG GTT GAA ATG CGC
Thermus thermophilus **WP_014510561.1**	ATG GAA GTC GCT CGC GGG CTT GAG GGG GTA TTA TTT ACC GAA TCG CGT ATG TGC TAT ATT GAT GGA CAG CAA GGG AAA TTA TAC TAT TAC GGA ATT CCA ATT CAA GAA TTG GCG GAA AAG TCG TCG TTT GAA GAG ACT ACC TTC TTA TTA TTG CAT GGC CGC TTG CCC CGT CGT CAA GAG TTA GAA GAA TTT TCA GCG GCA CTT GCC CGC CGC CGT GCA TTA CCT GCT CAC CTT TTA GAA TCG TTC AAA CGC TAT CCT GTA TCG GCC CAC CCT ATG AGC TTT CTG CGC ACA GCT GTA TCA GAA CTG GGT ATG CTG GAC CCT ACA GAA GGG GAT ATC TCA CGT GAG GCT CTT TAT GAA AAG GGT TTA GAC TTA ATC GCC AAG TTT GCG ACC ATC GTA GCT GCT AAC AAA CGC CTG AAA GAA GGG AAA GAA CCA ATC CCT CCT CGC GAG GAT CTG TCC CAC GCT GCG AAT TTC CTG TAT ATG GCT AAC GGA GTA GAG CCT TCC CCA GAG CAA GCT CGT CTG ATG GAT GCG GCG CTG ATC TTA CAC GCA GAG CAC GGG TTT AAC GCG TCT ACG TTC ACA GCA ATT GCC GCT TTT TCG ACT CAG ACT GAC TTG TAT TCA GCC ATC ACT GCT GCT GTA GCC TCG CTG AAA GGT CCA CGT CAC GGA GGC GCT AAT GAA GCG GTA ATG CGT ATG ATT CAA GAG ATT GGA ACT CCG GAG TGT GCC CGT GAA TGG GTG CGT GAG AAA TTG GCC AAA AAG GAA CGC ATT ATG GGC ATG GGC CAT CGT GTT TAT AAG GCT TTC GAC CCG CGC GCA GGT GTG TTG GAA AAA TTA GCC CGC CTT GTA GCA GAA AAA CAC GGG CAC TCG AAA GAA TAT CAA ATC CTT AAA ATC GTG GAG GAA GAG GCC GGG AAG GTA CTT AAC CCG CGC GGG ATT TAT CCA AAT GTG GAT TTC TAC AGC GGT GTG GTT TAC AGT GAC CTT GGC TTT AGC CTG GAG TTT TTC ACA CCA ATT TTT GCG GTG GCC CGT ATC TCA GGT TGG GTG GGA CAC ATC CTT GAG TAC CAA GAA TTA GAT AAT CGC CTT CTG CGT CCA GGC GCG AAA TAC ATT GGA GAA TTG GAT GTC CCC TAC GTC CCT CTT GAA GCG CGT GAT
Ananas comosus **XP_020089322.1**	ATG GAG GGT GCC TTC GAT CAT AGT GCA CTT GCC CGT GCG CGT TTA GCC GTG TTA TCA GGA CAC CTT GCA GCG GTG TCA GCC GCA GGC GGT GGC GCG TCA CCC TTA GAG CGC AGC CCA GTG AGT GTT CAA GAG ATT CCA CCA CCC CCT CGT AAT CTG GGC GGG TCA CTG GCG GTT ATT GAT GGG CGT TCA GGC AAG AAA TTT GAA TTC AAG ATC AGC GAC GAA GGC ACG GTT CGT GCT ACT GAC TTT AAA AAG ATC ACA ACC GGA AAA AAC GAC AAA GGG TTA AAA CTT TAT GAT CCG

	GGG TAC TTA AAC ACT GCA CCT GTT CGT TCT AGC ATC TGC TAC ATC GAC GGA GAC GAG GGT ATC CTG CGT TAC CGC GGC TAC CCG ATT GAA GAA TTA GCT GAG AGC AGC TCG TAT CCA GAG GTT GCC TAT CTT TTA ATG TAT GGA AAC TTG CCA TCG AAA AGC CAG TTG GCT GAT TGG GAG TTC GCC ATT AGT CAA CAC AGC GCT GTT CCT CAG GGT GTA CTT GAT ATC ATC CAA GGA ATG CCC CAT GAC GCA CAC CCG ATG GGG GTC TTA GTG TGT GCC ATG AGC GCG TTA TCT GTG TTC CAT CCT GAT GCT AAC CCG GCC TTG CGC GGT CAA GAT CTT TAC AAG AGT AAA CAG CTT CGC GAC AAA CAG ATT GTT CGT ATT TTG GGG AAG GTT CCG ACT ATT GCA GCC GCG GCT TAC TTA CGT TTA GCT GGC CGT CCC CCC GTT TTG CCT AGT AAT AAT TTC TCT TAT TCG GAG AAT TTC CTG TAC ATG CTG GAT AGT TTG GGG AAC CGC TCA TAC AAG CCA AAT ACG CGC CTG GCA CGC GTA CTT GAT ATT TTA TTT ATT CTT CAT GCC GAA CAT GAA ATG AAC TGC AGT ACT GCT GCT GCC CGC CAC TTA GCG TCC AGC GGC GTA GAT GTC TTT ACG GCG TTG GCC GGT GCC GTC GGG GCC CTT TAC GGC CCC CTG CAT GGG GGC GCT AAT GAG GCC GTA CTG AAG ATG CTT AAC GAA ATT GGC ACA GTA GAG AAC ATT CCG GAT TTC ATC GAA GGC GTA AAG AAT CGT AAG CGT AAG ATG TCA GGG TTT GGG CAT CGT GTT TAT AAA AAC TAC GAT CCT CGT GCG AAG GTA ATT CGT AAA TTA GCA GAA GAA GTT TTC TCA ATT GTT GGC CGT GAT CCT TTG ATC GAA GTA GCT ATC GCA TTG GAG AAA GCA GCT CTG TCT GAT GAA TAC TTC ATC AAG CGT AAA CTT TAC CCC AAC GTT GAT TTT TAC AGC GGT TTA ATC TAC CGT GCT ATG GGG TTC CCG ACT GAA TTT TTC CCT GTT CTG TTT GCA ATC CCG CGC ATG GCA GGT TAC TTA GCT CAC TGG CGC GAA TCG CTG GAT GAT CCA GAT ACC AAG ATC ATG CGT CCC CAG CAG GTC TAC ACG GGG GTT TGG TTG CGT CAC TAT ACC CCC TTA CAG GAA CGC ACC GTT AGT AAC GAT GAA GAC AAA CTT GGG CAG GTG GCG GTG AGT AAT GCC ACG CGC CGT CGT TTG GCG GGA AGT CGC ATT
Cavenderia fasciculata **XP_004355114.1**	ATG TCG CAC ATC TCT CGC CTT AAC GTT ATT TCT GGT CAT CTG AAG GAC GAG CAG GAA GAA GAG TAC GTG GAT GAA CTG ACG TCC AAC AAT GTG AAC GCC GAT GCT AAG AAG GAG ACG CTT ACG GTC ATT GAT AAT CGT ACC GGG AAG CAG TAC AAT ATC AAG ATC AAG AAC CAG ACT ATC TCC GCG TTG GCA TTC CGT GAA ATC TCG GAA CAG AAA GGG GAC AAT GGT TTA GTG GTC TAC GAC CCT GCG TTT CAG AAC ACG GCA GTT GTT ACT TCG ACA ATC ACC TAT ATC GAC GGC GAC AAG GGT ATC TTG CGT TAT CGC GGC TTC CCT ATT GAG GAG CTG GCA GAG CGT TCA TCG TTC TTA GAA GTT GCT TAT CTG CTG ATT AAC GGT AAT TTG CCG AAT AAA AGT CAA TTA GAT GGG TGG TCG CAC AAA ATT ATG ACG CAC ACC TTC TTA CAC GAG AAC TTG GTG GGG CTT ATG AAG AGC TTT CGT TAC GAC GCC CAC CCC ATG GGC ATG TTG ATC TCA ACG GTA GCC GCC TTA GGA ACT TTT TAC CCC GAG GCG AAC CCA GCG CTT GCA GGA CAG GAC ATC TTT AAG TCG GAA AAT GTA CGC AAC AAA CAA ATG TTC CGC ATT ATC GGG AAG TTG CCC ACA ATC GCG GCT TGT GCT TGG CGC CAC CGT ATT GGC CGT CCG TAC AAC ACG CCG GTG AAT CAC CTG GGC TAC ACC GAG AAC TTC TTG TAT ATG CTG GAT AAA CTT TCC GAA CAA GAC TAT AAG CCA AAT CCG GTA TTA TGC CGT GCC TTG GAA ATC TTG TTT ATT CTT CAC GCC GAC CAC GAA TTA AAT TGC TCT ACG GCG GCC ATG CGT CAT ATC AGT TCT TCC AAT ACT GAT CCG TAC ACG GCA GTC GCT GGC GCC GCA GGC GCT CTG TAT GGG CCC TTG CAC GGC GGG GCC AAC GAG GCC GTG TTA GAC ATG TTG CAA CAA ATT GGT ACC AAG GAA AAC GTT GGA CAA TTC GTG GCC GAC GTA AAA GCG AAA AAA AAG AAA TTG ATG GGG TTT GGA CAT CGC ATT TAT AAG AAC TAT GAC CCA CGC GCA AAG ATC GTC CGT CGC GTG GCC TAT GAA GTA TTC GAA TCT CTG GGT AAA GAA CCC CTG ATT GAA GTG GCT ACA GAG CTG GAA CGC CAA GCA CTT AAC GAC GAG TAT TTC GTA ACA CGT AAG TTA TAC CCA AAT GTC GAC TTC TAT TCA GGC TTA ATC TAC AAG GCC ATG GGA TTC CCG ACA GAC ATG TTC CCA GTG TTG TTT ACA ATC CCA CGC GCC GTA GGA TGG TTG GCC CAT TGG ATC GAG CAT CAC GAA GAT CCG GAG ACC AAA ATC TAT CGC CCA CGC CAA GTG TAT AAG GGA GAA TGG TTC CGC AAT TAC GTG CCA ATC GAG GGG CGC CCT CCA GCC AAA GTT CGT ACC CAA GAG AGC TAT AGT TCG GCC ACC ACT AAG CGT TAC TCG AAG GTC ACC GGC TCC GGC GCA CAA
Chlamydomonas reinhardtii **XP_001695571.1**	ATG ATG TCC AGC CGA CTG GAT GTT CTG TCT AGG CAG ATG TGC GGT ATG CAG CTC TCC CCC ACG GCC GGT GAA GAG GAG CAG CGG CCC GCG CCG GCC GGT GGT CGC GGA ACG CTG ACA GTT GTC GAC AAC CGC ACT GGC AAG AAG TAC ACG CTC GAG ATC TCC GAT GGC GGC ACC ATC AAC GCC CAG GCG CTA AAG CAG ATT AAG GCC GGT GGA GAC GGC GTG GGC CTG CGC ACC TAC GAC CCT GGG TAC ACA AAC ACC ACC GCC GTC ATC TCG CGC ATC TCC TTC ATC GAT GGC GAC AAG GGT ATT CTG CGC TAC CGT GGC TAC CCC ATT GAG GAG CTG GCT GCG CGA TCA AAC ATG ATG GAG GTG GCG TAC CTG GTT CTC TAT GGC AGC CTG CCC ACG CAG TCG CAG CTG TCG GTG TTC CAC GAG GCT GTG ATG CGG CAC ACG GCG CTT CCG ACG GAG GTG ATT GAT GTG ATT CAC GCC ATG CCG CAC GAC TCG CAC CCA ATG GGC GTG CTC ATG ACC GGC ATC TGC GCG CTG TCC GCC ATG CAC CCA GAG GCG AAC CCG GCT CTG GCG GGC CAG GGT GTG TAC AAG ACG CGC GAG ATG CAG GAC AAG CAG ATC GTG CGG CTG CTG GGC AAG ATC CCC ACC ATC GCC GCC ATG GCC TAC CAC AAA AGC ACG GGT CGC AAG GCC GCG CCG CCC AAC CAG CGG CTG GAC TAC ACG GAG AAC TTC CTC TAC ATG CTG GAC GGC GGC TAC AAC CCG CAG TAC CGG CCC AAC CCG CGC CTG GCG CGC GCG CTG GAC ATT ATG TTC CTC CTG CAC GCG GAG CAC GAG ATG AAC TGC TCC ACG GCC GCC GTG CGC CAC CTG GCC TCG TCG GGC GTG GAC GTG TAC ACG GCG GTG GCG GGT GCG GTG GGC GCG CTG TAC GGC CCG CTG CAC GGC GGA GCC AAC GAG GCG GTG CTG CGC ATG CTG GCG CGC ATT GGC TCG GTT GAG AAC ATC CCG GCC TTC ATC GCG GGC GTC AAG AAC AAG AAG GAG AAG CTG TTT GGC TTT GGC CAC CGA GTG TAT CGC AAC TTC GAC CCG CGC GCC AAG ATC ATT AAG GAC GTG GCG GAG AAC GAG GTG TTC CCG CTG GTG GGC GTG GAC CCG CTC ATC GAG ATC GCC AAG GCG CTG CAG GAC GCG GCC CTG TCT GAC GAG TAC TTT GTG AGC CGC AAG CTG TAC CCC AAC GTG GAC TTC TAC AGC GGC CTG GTC TAC CGC GCG CTG GGC TTC CCG CCG CAG TTC TTC ACC GTG CTG TTC GCC ATC CCA CGC ATT GTG GGC TAC TGC GCG CAC TGG CGC GAG TCG CTG ACG GAC CCG GAC ACC AAG ATC ATC CGG CCG CAG CAG GAC TAC AAG GGC GTG TGG CTG CGG CAC TAC TCG GAC GTG GTC GTG CGC ACC GCC GAG GGT GCG GAC ACG CTC TCG AAG CTG CCG CCC TCG AAC GCC TAC AAC CGT CGG GTT GCC GGT GAG AAC TGG ATG
Chlorobium limicola **WP_059139550.1**	ATG ACG GTA ACT GAG ACG GGT AAC TCG CTT ACC ATC GTA GAT AAC CGC ACC GGT AAG AGT TAC GAA GTT CCT GTG GAG AAT GGA AGT ATC AAT ACC ATG GAA CTG CGC AAG ATC AAG GTT TCG GAA GAA GAT TTC GGT TTA CTT GGG TAT GAC CCT GGA TAT TTG AAC ACA GCA TCA TGT AAG TCT CGT ATT ACC TAC ATT GAT GGG GAT AAA GGC ATC TTA CGC TAC CGT GGT TAC CCC ATC GAG CAA TTA GCG GAA AAG AGT ACT TTC TTA GAA ACT GCG TAC TTG CTT ATC AAA GGG GAA CTT CCG GAC AAG GAG CGT CTG GCT GTG TGG ACG TAT AAT ATC CGT CAC CAT ACC ATG ACC CAC AAT AAC ATC GTG AAA TTC ATG GAC GGG TTC CGC TAC GAC GCA CAC CCG ATG GGA ATC TTG GTC GGA ACT GTA GGT GCT CTG TCT ACT TTC TAC CGC GAC GCT AAA GAT ATC GGA AGC GAG GAT TCG CGC AAG CTG CAG GTG CGC CGC CTT ATT GGA AAG ATC CCC ACC TTA GCT

	GCA ATG TCT TTC CGC CAT AGC ATG GGG TTC CCT TAC GTT ATG CCT GAT AAC GAT CTG TCC TAC GCT GGG AAT TTC CTT TCA ATG ATG TTT AAG ATG ACT GAA CTT CGT TAT AAA CCG AAC CCA GTT TTG GAA CGT GCA TTG GAC GTG TTG TTT ATC CTT CAT GCG GAC CAC GAG CAG AAC TGT TCT ACG TCT TCC TTA CGC GCG GTC GCT AGC TCT GGC GTA GAT CCT TTT TCA GCG ATC GCC GCA GGC TGT GCT GCC TTA TAT GGG CCG CTT CAT GGC GGG GCC AAT GAG GCT GTT ATT CGC ATG TTA ATG AAG ATC GGT TCC ATT GAC AAG ATT CCA GAA TTT ATT CAA TCG GTA AAG GAC GGC GAT GGT CGC TTG ATG GGT TTC GGA CAT CGC GTA TAC AAG AAT TAT GAC CCT CGC GCT AAA ATC ATT AAA GAT ATC GCT TTC GAA GTG TTC GAG GAG ACT GGT CGC AAT CCG ATG CTT GAC ATT GCC TTA GAA TTA GAA CGT ATT GCT CTT GAG GAC GAC TAC TTT GTA AGC CGT AAA TTG TAC CCT AAC GTC GAT TTT TAT TCG GGA TTG ATC TAT CAG GCG ATG GGA TTT CCG ATG GAC ATG TTC CCT GTT CTT TTT GCC ATC GGC CGC ATC CCA GGG TGG CTG GCG CAG TGG ATC GAA CAC GTC AAA GAT GGC GAA CAA AAA ATC GCA CGT CCC CGT CAG ATC TAT CTT GGG GAG GAT GAG CGC CAA TTT GTA GCG ATG GCT GAC CGC CCA AAG ACC CGT TTG GAT GAA CAG ATG GCC GGA ATC TGC CGT TTG
Conexibacter woesei **WP_081691039.1**	ATG AGT GAG ACT CAG ACT ACG GGG GAC GGG GCG GTA GCA GCT GCA AAT GCG GAC ACC CTG ACA GTG ACG GAT AAT CGT ACC GGC CAG ACT TAC GAG GTA CCC ATC ACT GAC GGA ACT GTA CGC GCT ATG GAT TTC CGC CAG ATG AAA ACT TCA GAC GAC GAC TTT GGT CTG ATG ACC TAC GAT CCT GCA TTT ACG AAT ACA GCA TCC TGT CGT TCG GCG ATT ACC TAT CTT GAT GGA GAA AAC GGG GTA TTG GAA TAT CGT GGC TAC CCT ATC GAG CAA TTG GCC GAG CGC TCC ACA TAT CTG GAG GTG GCC TAC CTG TTA GTC CAC GGC GAG CTT CCA ACT ACT CCA CAG CTG GAT CAG TGG AAG CAC GAC ATT ACG ACT CAC ACA TTT GTG CAC GAG AAC GTT AAG GAG TTT GTA GGG GGG TTT CGT CAC GAC GCT CAC CCG ATG GGG ATG CTT GTC GGA AGT GTT GGA GCA CTG TCA ACG TTC TAC AAA GAC GCT AAC GAA ATT AGC GAC CCG GAA AAC CGC GCA CTG CAG ACA ATC CGT TTG ATC GCT AAG ATG CCG ACA TTG GCT GCT TTT GCC TAT CGT CAC ATT ATG GGA CAA CCA TAT TAC TAC CCT GAT AAT GAC TTA GAC TAT CCT GGA AAC TTC CTT TCG ATG ATG TAC AAG ATG ACA GAA CTG AAA TAC AAG CCA GAT CCG CGT TTG GAA CGC GCG CTT GAC GTT CTT TTT ATC TTA CAC GCA GAT CAC GAG CAG AAT TGT TCA ACC AAT GCT GTG CGT GCC GTA GGG TCG TCT CAA GTG GAC CCA TAC AGT GCC GTA GCG GCA GGA GTT GCC GCC CTG TAT GGG CCA TTG CAT GGA GGC GCA AAT GAG GCA GTT CTT CGT ATG TTG AAG CGC ATC GGA AAC AAG GAA AAC ATT CCT GAC TTT ATG CAA GGA GTC AAA GAT GGA AAT GAA CGC TTA ATG GGA TTC GGC CAC CGC GTG TAC AAA AAT TAC GAT CCC CGT GCA ACA ATT ATC AAA AAG GCA TGT GAC GAC GTA TTT GAG GTA ACA GGG GTA AAT CCT CTG TTA GAT ATC GCT CTG GAA GTA GAG AAA ATC GCG CTT GAG GAT GAG TAT TTT GTC TCC CGT AAG TTA TAC CCG AAC GTA GAC TTT TAC TCG GGC TTA ATC TAC GAA GCA CTG GGT TTG CCT ATG GAC ATG TTT CCT GTA ATG TTC GCG ATC CCT CGC ACA AGT GGA TGG ATT GCA CAG TGG TTA GAG ATG GTG CAA GAC AAA GAG CAA AAA ATT GCG CGC CCA CGC CAA ATC TAC ACT GGT GAA CGT ACC CGC GAC TAC GTC GGC ATT GAT GCC CGT AAG
Corynebacterium pollutisoli **WP_085550151.1**	ATG GCC ACC GAT AAC ACC GAA AAG GCG GTG TTG AAT TAT CCC GGC GGC GAA TTT GAA ATG GAC ATT ATC AAG GCG ACA GAA GGT AAC GAC GGT GTA GTG CTT GGG AAG ATG CTG GCG GAT ACC GGG TTG GTT ACA TAC GAC CCT GGA TAC GTT AGT ACA GGC TCG TGT GAA TCG GCG ATT ACA TAC ATC GAT GGG GAC AAT GGT ATC CTG CGT CAT CGT GGC TAT GAC ATC GCA GAC TTG GCC GAG AAT GCG ACG TTC AAC GAA GTC TCG TAT CTG CTT ATC AAC GGC GAA CTG CCT ACG GTA GAA GAG TTG CAT AAA TTC TCT GAC GAG ATC CGT CAT CAC ACC CTT CTG GAC GAA GAT TTT AAG TCT CAA TTC AAT GTG TTT CCA CGC GAC GCT CAT CCT ATG AGC GTG CTG GCG AGC TCG GTT AAC ATT TTG AGT ACT TAT TAC CAG GAT CAG CTT GAT CCG CTG AAC GAA GAG CAT CAG CAT AAA GCA ACA GTA CGT CTT ATG GCT AAG GTG CCA ATG CTT GCT GCC TAC GCG TAT CGC GCA TCA AAA GGG GCC CCC TAC ATG TAC CCA GAT AAC TCT CTT AAT GCC CGT GAA AAT TTC CTT CGC ATG ATG TTT GGG TAC CCA ACG GAG CCA TAT GAG ATT GAC GAG GTC ATG GTC AAG GCT TTG GAT AAA TTA CTT ATC CTG CAC GCC GAC CAC GAA CAG AAC TGC AGT ACT TCT ACT GTA CGC ATG ATT GGG AGC GCA CAA GCG AAC ATG TTC GTC TCT GTA GCC GGC GGA ATT AAC GCT TTG TCT GGG CCA CTG CAC GGT GGG GCG AAC CAG GCG GTG CTT GAG ATG CTG GAA GAA ATT GAT GCG AAT GGG GGC GAC GCA ACG GAT TTT ATG AAC CGT GTC AAA AAT AAA GAA AAG GGT GTT CGT CTT ATG GGA TTC GGA CAC CGC GTA TAC CGC AAT TAC GAC CCA CGT GCT GCC ATT GTG AAG GAG ACA GCG CAC GAA ATT TTG GAA CAT CTT GGT GGC GAT CAT CTG CTG GAT TTG GCC ATG AAA CTG GAA GAG ATC GCA CTT AAT GAC GAA TAC TTT ATT TCA CGT AAA TTG TAC CCG AAC GTG GAC TTC TAC ACA GGG CTT ATT TAC CGT GCA ATG GGG TTT CCA ACT GAC TTC TTT ACA GTC CTG TTC GCA ATT GGG CGT TTG CCT GGC TGG ATC GCC CAT TAC CGC GAG CAG TTG GCT ACT ACG ACC AAG ATC AAT CGT CCA CGT CAG ATT TAT ACG GGT AAC ACC TTA CGT ACA GTA ACA CCT CGT GAG GAA CGC
Desulfoluna spongiiphila **WP_092207573.1**	ATG TCA GAA GTA GTT AAA ATG ATT ATC GAA GGC AAG ACA TAC GAG TTT CCC GTA GTT GTG GGA AGC GAA GGA GAG AAA GCT TTT GAC ATC ACC AAG CTG CGT CAA CAA ACG GGG TAC ATT ACA ATG GAT CCC GGC TTT GGC AAT ACG GGC AGT TGT ACT TCG GCA ATC ACC TTC ATG GAC GGT GAG AAG GGG ATC CTG CGT TAT CGC GGA ATT CCG GTG GAG CAG CTT GCG GAG CAT TCG TCA TTC GTC GAG ACG GCG TAT CTT TTA ATC AAT GGA CGC CTT CCG AAT CGC CAG GAA CTT ACT CGC ACA AGT GTG CAG CTT AAT GAC GAC TCC TTG ATT CAT GAG GAT ATG CAA ATC TTC TAC CAA AAT TTT CCC CGC CGT GCG CAC CCT ATG GGG ATT TTG AGT GCC ATG CTG AAT GCT TTA CGT TCC TTC TAT CCT GAA TTA GAA GAC ACT GAC CCT GAA GGG ATC AAC CGT ACA TTC CTT CGT TTG TTA TCG AAG ATT CGT ACT ATG GCC GCG ATG TCT TAT CGC ATC AGC CGC GGT CAT AAA GTC GTT TAC CCC CGC GCG GAC TAC TCC TAT TGC GCT AAC TTC CTT AAT ATG ATG TTT GAT TCT CCG GTG CGT CCT TAC GAG TTG GAT CCT GAT ATT GTT GAG GCG TTA AAT GTG TTT TGG ATT CTT CAT GGG GAC CAT GAG CAA AAC TGC AGT ACC GCG GCC GTT CGT GTG GTG GGA TCT GGT CGC GTC AAT CTT TAC GCA GCC ATT TCA GCT GGC ATC GCC GCA CTT TGG GGC CCC CTG CAT GGG GGC GCA AAT CAG GCC GTA ATT GAG ATG TTG ACT AAT ATC CAT CAG TCC GGG GCG GAC ATT GAA GAT GTC GTT AAA CGT GCA AAG GAC AAG GAT GAT CCA TTT CGT CTT ATG GGA TTT GGA CAT CGT GTG TAT AAG ACA TAC GAT CCT CGT GCC AAG ATC ATG AAA AAA ATG TGC GAC ACT GTT TTG CCA AAA TTA AAA GTG GAA GAT CCT CTG TTA GAC ATC GCT CGT CGT TTG GAG GAG GTA GCG CTG ACC GAC CCA TAC TTC GTA GAC AAG AAT TTG TAC CCT AAT GTT GAC TTC TAT TCT GGA ATC GTG TTG CGT GCT ATG GGG ATC CCT ACT GAA ATG TTC ACC GTT ATG TTC GCT ATT GGT CGC TTG CCT GGT TGG ATC AGT CAG TGG AAG GAG GGA GCT GAT GAT CCG AAC TGG AAG ATT TCT CGT CCA CGC CAG GTG TAC ACC GGT AAT ACC GTA ACC GAT TAT TTA CCA ATG AAC CAG CGC

Escherichia coli **WP_166726827.1**	ATG GCT GAT ACA AAA GCA AAA CTC ACC CTC AAC GGG GAT ACA GCT GTT GAA CTG GAT GTG CTG AAA GGC ACG CTG GGT CAA GAT GTT ATT GAT ATC CGT ACT CTC GGT TCA AAA GGT GTG TTC ACC TTT GAC CCA GGC TTC ACT TCA ACC GCA TCC TGC GAA TCT AAA ATT ACT TTT ATT GAT GGT GAT GAA GGT ATT TTG CTG CAC CGC GGT TTC CCG ATC GAT CAG CTG GCG ACC GAT TCT AAC TAC CTG GAA GTT TGT TAC ATC CTG CTG AAT GGT GAA AAA CCG ACT CAG GAA CAG TAT GAC GAA TTT AAA ACT ACG GTG ACC CGT CAT ACC ATG ATC CAC GAG CAG ATT ACC CGT CTG TTC CAT GCT TTC CGT CGC GAC TCG CAT CCA ATG GCA GTC ATG TGT GGT ATT ACC GGC GCG CTG GCG GCG TTC TAT CAC GAC TCG CTG GAT GTT AAC AAT CCT CGT CAC CGT GAA ATT GCC GCG TTC CGC CTG CTG TCG AAA ATG CCG ACC ATG GCC GCG ATG TGT TAC AAG TAT TCC ATT GGT CAG CCA TTT GTT TAC CCG CGC AAC GAT CTC TCC TAC GCC GGT AAC TTC CTG AAT ATG ATG TTC TCC ACG CCG TGC GAA CCG TAT GAA GTT AAT CCG ATT CTG GAA CGT GCT ATG GAC CGT ATT CTG ATC CTG CAC GCT GAC CAT GAA CAG AAC GCC TCT ACC TCC ACC GTG CGT ACC GCT GGC TCT TCG GGT GCG AAC CCG TTT GCC TGT ATC GCA GCA GGT ATT GCT TCA CTG TGG GGA CCT GCG CAC GGC GGT GCT AAC GAA GCG GCG CTG AAA ATG CTG GAA GAA ATC AGC TCC GTT AAA CAC ATT CCG GAA TTT GTT CGT CGT GCG AAA GAC AAA AAT GAT TCT TTC CGC CTG ATG GGC TTC GGT CAC CGC GTG TAC AAA AAT TAC GAC CCG CGC GCC ACC GTA ATG CGT GAA ACC TGC CAT GAA GTG CTG AAA GAG CTG GGC ACG AAG GAT GAC CTG CTG GAA GTG GCT ATG GAG CTG GAA AAC ATC GCG CTG AAC GAC CCG TAC TTT ATC GAG AAG AAA CTG TAC CCG AAC GTC GAT TTC TAC TCT GGT ATC ATC CTG AAA GCG ATG GGT ATT CCG TCT TCC ATG TTC ACC GTC ATT TTC GCA ATG GCA CGT ACC GTT GGC TGG ATC GCC CAC TGG AGC GAA ATG CAC AGT GAC GGT ATG AAG ATT GCC CGT CCG CGT CAG CTG TAT ACA GGA TAT GAA AAA CGC GAC TTT AAA AGC GAT ATC AAG CGT TAA
Fluviicola taffensis **WP_0136874881.1**	ATG TCG AAG ATC GCA AAA ATC GAA TTG GAC GGC AAA GTT TAC GAA TTT CCA GTG GTT CAG GGA ACG GAA AAC GAG CTT GCA ATT GAC ATT AGT AAG TTG CGC CAA GAG ACA GGC TAC GTC ACT CTT GAC ACC GGC TAC AAG AAC ACC GGC GCC ACT ACA AGT GCA ATC ACC TTC CTT GAC GGG GAA GAG GGT ATT TTA CGT TAT CGT GGG TAT CCA ATT GAG CAA CTT GCA GAG AGC GCC ACC TTT ATT GAA GTG GCT TAC CTG TTA ATC TAT GGC GAA CTT CCA ACG CAG GTT CAA CTT GAC GAT TTT ATT TCT TCT ATT ACC AAG CAC ACC TTA GTC CAT GAG GAC ATG AAA CAG TTT TTC GAG GCT TAC CCA GCG CAG GCA CAT CCT ATG GGT GTA TTA AGT TCC ATG GTA TGT TCA TTG AGT ACG TTC TAC CCC GAG TCT TTA GAT CCT AAT CGC TCT GCG GAA AAA AAA AAT CGT ACG ATC CTG CGC CTT CTT GCC AAA CTT CCA ACT CTG GCA GCT TGG TCG TAT AAA AAC AGT ATG CGT CAT CCC TTC ATG TAT CCG AAA AAT GAA TAT GAT TAT GTG AAG AAC TTT CTT TAC ATG ATG TTC GCA ATG CCT ACG GAA GAT TAC AAG GTT GAT CCG GTA GTA GTC GAT GCC CTT AAT AAA CTG TTA ATT CTG CAC GCA GAC CAC GAG CAG AAC TGC AGC ACG AGT ACT GTT CGC ATC GTA GGC TCA TCT CAG GCG AAC CTG TAT GCG TCA ATC TCT GCG GGT ATT AGT GCC TTA TGG GGA CCA CTG CAT GGT GGG GCG AAT CAA GCG GTA ATC GAG ATG CTT GAG AAG ATC CAT AAC GAC GGT GGT GAC GTT GAC AAG TGG GTC CTG AAA GCC AAA GAC AAG GAT GAT CCG TTT CGC CTG ATG GGT TTT GGT CAC CGT GTT TAT AAA AAT TTT GAT CCG CGT GCC ACC ATT ATC AAG AAA GCT GCC GAC GAT GTC TTA GAG AAA TTG AAG ATC AAC GAC CCC TTG CTG GAT ATT GCC AAA AAA CTT GAA AAG TAC GCG CTT GAA GAC GAA TAC TTT AAG AGC CGC TCT CTT TAT CCA AAC GTC GAT TTC TAC TCC GGG ATC ATT TAC AAA GCT CTT GGT ATC CCG TCC GAA ATG TTT ACT GTC ATG TTC GCA CTG GGT CGC TTG CCA GGG TGG ATT GCA CAG TGG AAA GAA ATG AGT GAG GGT GGA GAA CCT ATT GGA CGT CCT CGC CAG ATC TAC ACA GGC GAG ACC ACC CGC GAA TAC GTC CCG ATC GCG AAA CGT GGT
Methylobacillus flagellates **WP_011478429.1**	ATG AAG CCC AGT AAG GTC ACG CTT ACG TTC GAT GAT GGT AGC GCC CCT ATC GAA CTG CCA ATT TTG CCT GGC AAA TTA GGT CCT AGT GTA ATT GAT ATT CGT TCA CTT TCA AAA CAC GGC TAC TTT ACT TAT GAT CCA GGT TTT CTT AGT ACA GCG TCG TGC GAC TCT AAC ATC ACC TTC ATT GAT GGC GAA GAG GGT CTT TTG TTC TAC CGT GGT TAC CCC ATC GAG CAA TTA GCT GAG CAC TCC GAT TTT ATG GAG GTC AGC TAC TTG TTG ATC TAT GGC GAG CTT CCG AAC GCG GAG CAA AAG GAG AAA TTC ACG AAT ACG ATC ACA CGT CAT TCT ATG GTC CAT GAC CAG CTG ACA CAC ATC TTC CGT GGG TTT CGC CGC GAT GCA CAC CCT ATG GCG GTC ATG GTG GGC GTC GTA GGT TCA ATG TCC GCT TTC TAC CAC GAA GCA ATT GAC GTT TCG GAT CCG CGC AAC CGT GAA TTT GCG GCC CAC CGC TTG CTT GCA AAG GTC CCT ACT ATC GCC GCA TGG TCG TAC AAA TAT AAC GTC GGC CAG CCT TTC ATG TAC CCC AAA AAC CGC TTT AAC TAC GCG GAA AAT TTC ATG CAC ATG ATG TTC GCC ACA CCT TGC GAG GAT TAT GAA CCG AAC CCC GTT CTT GCT AAA GCG TTC GAA CGT ATT TTG ATT CTG CAT GCA GAT CAT GAG CAG AAT GCA TCG ACA TCA ACC GTT CGT CTG GCG GGG TCA TCT GGC GCC AAC CCT TTT GCC TGC GTC GCA GCA GGA ATT GCG AGC CTT TGG GGT CCG GCT CAT GGC GGC GCA AAT GAA GCC ACT CTG AAG ATG CTT GAG GAG ATC GGC GAC ATT TCG CGC ATC GGT GAG TAT ATT AAG CGT GCC AAA GAC AAG ACT GAT AGC TTC CGT CTT ATG GGA TTC GGA CAC CGC GTA TAT CGT AAT ATG GAT CCC CGC GCT GCC ATC ATG CGC CAA ACC TGC CAC CAA GTG TTG GAC GAA CTT GGG TTG AAC GAC GAT CCG ATG TTT AAG CTG GCG TTG GAG TTA GAG AAA ATC GCG TTA GAG GAT GAA TAC TTC GTG TCC CGC CGT CTG TAT CCG AAT GTA GAT TTC TAT TCA GGT ATC GTG ATG CGC GCG ATG GGC ATC CCG AAC TCG ATG TTT ACG GCG ATC TTC GCA TTG GCG CGT ACG GTC GGA TGG GTG GCG CAG TGG AAT GAA ATG ATG TCG GAC CCT GGG TCG AAA ATT GGA CGC CCC CGC CAG TTA TAT ACT GGA CCT GAG CGC CGT GAT TTC CTG CCT CTG GAC CAA CGT
Myxococcus xanthus **WP_140862691.1**	ATG GCC AAG GAC ACG CTG ACG ATC ACC GAC AAT CGG ACC GGG AAG ACC TAC GAG GTC CCG ATC GAG AAC GGC TGT ATT CGC ACC AAT GCT CTC CGC CAG ATC AAA ACG GGC GAC GAT GAC TTT GGG TTG ATG GGC TAC GAT CCG GCG TTC CTC AAT ACC GCC AAC TGC AAG AGC GCC ATC ACC TTC ATC GAC GGA GAC AAG GGC ATC CTG GAG TAC CGG GGC TAC CCC ATC GAG CAG CTC GCG GAG AAG TCG AGC TTC CTC GAG GTC GCC TAC CTC CTG CTG AAC GGT GAG CTT CCC TCC GAG AAG GAG CTG CAG CAG TTC AAC CAC CTC GTC ACG CAC CAC ACC TAC GTG CAC GAG AAC GTG AAG AGC TTC ATG GAC GGG TTC CGC TAC GAC GCG CAC CCC ATG TCC ATG GTG AGC TCC ACG GTG GCG GCG CTG TCC AGC TTC TAC CCC GAC GCG AAG AAC ATC AAG GAC CCG CTG AGC CGC CGC ATC CAG ATC ACG CGG CTC ATC GCG AAG ATG CCC ACC ATC GCC GCG TTC TCC TAC CGG CAC ACG ATG GGC CTG CCG TAC ATC TAC CCG GAC AAC GAC CTG TCC TAC GTC GCC AAC TTC CTG GCG ATG ATC AAG CGC ATC GGC ACG TCG ACC TAC AAG GTC CAC CCG GTG CTC GAG CGC GCG CTC GAC ATC CTC TTC ATC CTG CAC GCG GAC CAC GAG CAG AAC TGC TCC ACC ACG TCC GTG CGC ACG GTG GGC TCG TCC GAG GTG GAT CCG TAC TCG GCG GTC AGC GCG GGC ATC GGC GCC CTC TAC GGC CCG CTG CAC GGC GGC GCC AAC GAG GCG GTG CTC CGC ATG CTC CGC GAC ATC GGC CAC GTC TCC AAG ATC CCG GAG TTC ATC AAG TCG GTG AAG AGC GGC

	GAG GGT GAG AAG AAG CTG ATG GGC TTC GGT CAC CGC GTC TAC AAG TCC TAC GAT CCG CGC GCC AAG GTC ATC AAG CGC GTG GCC GAC GAG GTC TTC GAG GTG ACG GGC AAG AAC CCG CTG CTG GAG ATC GCC GTC GAG CTG GAG CGC ATC GCG CTC GAG GAC GAG TAC TTC GTG AAG CGG AAG CTG TAC CCG AAC GTC GAC TTC TAC TCG GGC CTC ATC TAC GAG GCG ATG GGC TTC CAG GTC GAG ATG TTC CCG GTG CTC TTC GCC ATC CCC CGC ACG GTG GGC TGG TGC GCG CAG TGG GAG GAG ATG GTG ACG GAC AAC GAG CAG AAG ATT GCC CGT CCG CGC CAG GTG TTC ACG GGC GCC CCG CGC CGC GAC TAC GTC CCC ATG GAC AAG CGC GCC GTC AAG
Neisseria arctica **WP_047760537.1**	ATG AGT AAG ACT ATT AAA TTA CAA GTC GAA GGA CAA GAG GCG TTC GAA CTG CCC GTT TTG GAG GGT ACG TTA GGT AAC GAT GTA GTG GAC ATT CGT ACT TTT ACT AAA GAG ACC GGT ATG TTT ACT TTC GAT CCG GGG TTT GTT TCG ACC GGG TCC TGT GAG TCG AAG ATT ACC TTT ATC GAT GGT GAG AAG GGC CAA TTA TAC TAC CGT GGC TAC CCT ATC GAA CAG TTG GCT GAA AAT AGC GAC TAC CTT GAA ACC TGC TAC CTG CTT ATC TTC GGG GAG CTT CCC ACC GCT GAA CAG AAA GCA CAA TTT GAA CAC ACT GTG TCG CGC CAC ACG ATG GTC CAC GAA CAA CTT ACT TGG TTC TTC CGC GGA TTC CGT CGC GAC GCT CAC CCG ATG GCA ATG ATG GTC GGT GTG GTA GGC GCA CTT GCA GCT TTT TAT CAA GAT AGC TTG GAC ATT AAC AAC GCT GAC CAT CGT CGC ATT GCT GAG TTT CGC CTT ATT TCT AAA ATT CCG ACA ATC GCG GCC ATG TGT TAC CGC TAC TCA AAC GGA CTT CCG TTC AAC TAT CCT CGC AAC GAT TTA TCT TAT ACG GCA AAC TTT ATG CAC ATG ATG TTC GCT ACA CCG TGC GAG GAG TAT CAG CCC AAC CCC GTT CTG GTG CGC GCA CTT GAC CGC ATT TTT ATC TTA CAC GCA GAT CAC GAA CAA AAC GCA TCT ACT TCC ACA GTG CGT CTG GCA GGA AGC TCG GGA GCG AAT CCG TTT GCA TGC ATC GCG GCG GGT ATT GCA AGC TTG TGG GGT CCC GCT CAC GGT GGT GCA AAC GAG GCG GTG CTG AAA ATG CTG GAC GAA ATC GGC GAT GTG AGT CGT GTG CCG GAA TTT ATG GAA GGT GTG AAG GAG CGC AAG TAC CGT TTA ATG GGC TTC GGT CAC CGC GTC TAC CGC AAC ATG GAC CCG CGT GCT AAC ATT ATG CGC GAG ACT TGC TAT GAA GTA CTG AAG GAA TTG GGG CTG GAA AAT GAC CCC AAA TTC AAG CTT GCC ATG GAA TTG GAG CAG ATC GCG TTA AAC GAT CCG TAC TTC GTG GAG CGT AAG TTA TAT CCT AAC GTT GAT TTC TAT AGC GGG ATT GTA CTT TCT GCC CTT GGG ATT CCT GTC TCA ATG TTT ACA GTC ATT TTT GCT TTA GCA CGT ACC GTT GGC TGG ATT AGT CAC TGG CAC GAA ATG ATT GCT GAC CCT GGT CAT AAG ATT TCT CGT CCT CGT CAA TTG TAT ACG GGG GCT GTA CGT CGT GAT TAT GTT CCA GTC GAC AAA CGC
Patulibacter americanus **WP_022926837.1**	ATG AGC GAG ACA CAA GAG CGC GGT GGC GAC GCG GCG GTC GCA ACA GAT CAG CAA ACG CTG ACT GTC GTC GAC AAT CGT ACC GGC AAG CAG TAC GAA GTC CCC ATC AGT GAT GGA ACT GTC CGT GCA ATG GAC TTT CGC CAG ATT AAA ACT GGA GAC GAC GAC TTT GGA TTG ATG ACG TAC GAC CCG GCA TAT ACA AAT ACG GCA AGC TGC CGT TCA TCG ATC ACC TTT ATC GAC GGC GAT AAA GGC ATC TTA GAA TAT CGT GGC TAC CCG ATT GAG GAA TTG GCT GAG AAG GCA ACT TTC CTG GAG GTA GCT TAC CTG TTG ATT TTT GGT GAG TTG CCT ACG ACT GAT GAA TTA GCA GAG TGG ACG CAC GCT ATT ACG ACC CAC ACA TTT GTG CAT GAA AAC GTA AAA AGC TTC GTA GAA GGA TTT CGT TAT GAT GCA CAT CCG ATG GGA ATG TTG CTT GCA TCG GTA GGG GCG CTT TCG ACC TTC TAT CCG GAA GCA AAA GAG ATT AAA GAT GAA CAA TTA CGC TAC GAA CAA ATC GTC CGT CTT ATC GCC AAG GTT CCT ACT CTG GCG GCC TTT GCG TAC CGC CAC GCC AAG GGC ATG CCT TAC GTA TAC CCG GAC AAC GAC TTA AAT TAT GCC GAA AAC TTT CTG GGT ATG TTG TTC AAA ATT GCA GAA CTT AAA TAT CAG GCT GAC CCT CGT ATT GCT AAA GCC CTG GAC GTG CTG CTG ATC TTA CAC GCA GAC CAC GAA CAG AAC TGC AGT ACC TCG TCA GTA CGT GCA GTA GGG TCA AGC CAA GTC GAT CCA TAT AGT GCA GTA ACC GCC GGA ATC GCA GCC CTG TAT GGT CCC CTT CAC GGG GGA GCG AAT CAA GCG GTT TTA GAA ATG CTT CAG GAG ATT GGC ACA GTC GAA AAT GTA CCT GCC TAT ATC GAG AGT GTC AAA TCC GGG GAT AAA AAG TTG ATG GGC TTT GGC CAC CGT GTG TAT AAA AAT TTC GAT CCG CGT GCC AAA ATT ATT AAG AAG GCC GTT GAT GAT GTA TTT GAG GTT ACT GGA GTG TCT CCT CTT TTG GAG GTT GCC CAG GAA TTG GAA CGT ATT GCC TTG GAG GAC GAC TAC TTT GTA AAA CGC AAA CTT TAC CCG AAC GTC GAT TTC TAC TCT GGA CTT ATC TAT GAG GCC CTT GGC CTT CCC GTG GAA ATG TTC ACA GTT TTG TTC GCA ATC GGC CGC ACG CCT GGG TGG ATT GCG CAG TGG CTT GAG ATG GTG CAA GAT AAG GAG CAA AAA ATC GCC CGC CCT CGT CAG ATT TAT ACC GGG GAA CGT ACC CGT AGT TAC GTT CCT ATT GAT GAT CGT TCT
Perkinsus olseni **AND95703.1**	ATG GAC CGC TTG AAT ACC GTA GCT GCC CAT GTC GCA CCA ACT CTT ACA ATT ACG GAT AAT CGC AGC GGA AAG AAG ATC GAC GTG AAA GTT AAA GAT GGG ACG ATC CGC GCA ACG GAT CTT AAA CAG CTT GGT ATT CAA ACA TAC GAC CCG GGC TAT ATG AAC ACT ACC TGT TGC GTA TCA CGC ATC TCT TTT ATT GAT GGG AAC CGT GGA GTG CTT CGT TAC CGC GGC TAT CCT ATT GAA CAA TTA GCC CAG GAT AGC AGT TTC ACT GAG GTT GCC TTT CTG TTG CAG TAT GGA GAA CTG CCC ACG CGT GGC CAA TTA ACG CAG TGG GAG CAA GAC CTG ATG CGT CAC AGT ATG CTG CAT CAA GAT GTA GCC ACG CTT ACG TCC GCG TTC CGT TAC GAC GCG CAT CCG ATG GGC ATG TTC GTA TCA GCC ATG GCT GCT CTG GGA ACG TTG CAC CCC GAG ACC AAC CCA GCA TTA ACT AGC CAA AAC ATT TAT AAG GAT AAG GCG GTT CGT AAC AAA GAA ATT GCC CGC ATC TTG GGT TCA GCT ACA ACT TTA GCT GCT ATG GCA TAC CGT CAT CGC ATG GGC CGC CCT TTT AAC ACC CCA AAC TCC GAT CTT GGT TAT ACT GAA AAT TTC CTT TAT ATG TTA GAT GTG TCC CAC GAA GGC ACG TCT TAT CGT CCG CAC CCC AAA CTG GTA AAA GCT TTG GAC GTG CTT TTC GTC CTG CAC GCC GAG CAC GAA CTT AAC TGT TCG ACA AGC GCC ATG CGT CAT TTA ACT AGT TCA AAT GTC GAT GTT TAT ACA GCA TGT GCC GCT GCG GCC GGT GCG TTG TAC GGA CCC CGT CAT GGG GGA GCG AAT GAA GCG GTC TTA CGC ATG TTG GAG CGC ATT GGG TCA ATT GAG AAC GTA CCT ACC TTC GTA GAT AAG GTC AAA CAG CGT AAA GAG TTG TTA ATG GGC TTC GGA CAT CGC GTT TAT AAG CAT TAT GAT CCT CGC GCG ACC ATC GTA CGC AAA ATC GCA GAG GAT GTC TTC GAG ATT TGC GGT CGC GAA CCG CTT ATT GAG GTA GCC ATG GAT TTA GAA CGC ATC GCA TTG TCG GAT CCG TAC TTT AAA GAT CGC AAG CTT TAC CCT AAC GTA GAT TTT TAC TCA GGC CTG ATT TAT AAG GCC ATG GGT TTT CCG ACT GAT ATG TTC CCA ATC TTA TTT ACC GTT CCT CGT CTG GCA GGC TGG CTG GCC CAT TGG TCC GAA TGG ATC GAC GAT CCT GAG AAC CGT ATC TAC CGC CCC TTC CAG GTT TAC AAG GGC TAT GAT AAG CGC GAT TAC GTG CAA GTT GAT GAA CGT AGC GAG AAA GTT CAC TCT AAG GAA TTG ACT GTA AAG CGT AGT GCC TTT AAC CGT CGC CGC GAC GCG TCG CTT CAG TCA GCA GCC GCG GAC CGC TCG GAA TAC GAG TGG ATC AAA GAT
Perlucidibaca aquatica **WP_068858516.1**	ATG ACA GCG AAA ACC GCC ACA CTT ACC ATT GGC GAC GAA ACT TTA ACA CTT CCC ATC ACG AGC GGT ACA CTG GGA CCA GAC GTC ATT GAT GTT AAA GAT GTC CTG GCT AAA GGT TAT TTT ACG TAC GAT CCG GGG TTT ATG GCA ACG AGT GCG TGT GAA TCA AAG ATT ACA TTC ATC GAC GGA GAT AAA GGG ATC TTG CTG CAC CGT GGT TAC CCT ATC GAT CAG TTA GCG GAG AAA GCG GAA TAT CTG GAG GTA TGT TAT CTT CTT TTG AAC GGT GAG CTT CCA AAC GCG CAG CAA AAA

	GCC GAG TTC ACA AAA CTT GTT ACA GAT CAT ACG ATG GTC AAC GAT CAA CTG CGT AAG TTT TTC GAA GGG TTT CGC CGC GAC GCT CAC CCG ATG GCT GTG ATG TGT GGC GTG GTG GGG GCG CTG AGC GCA TTT TAT CAC GAC GGT TTA GAC GTA AAT GAT CCT AAG CAT CGT GAG ATC ACA GCG ATC CGT CTG ATC GCA AAA ATC CCT ACC ATT GCT GCA ATG TCC TAC AAG TAT ACA GCA GGA CAG CCT TTT ACC TAC CCT CGT AAC GAC CTG AAC TAC GCG GAG AAT TTT TTG CAC ATG ATG TTC TCC GTG CCA GCG GAC GTC AAC TAT AAG GTC AAC CCG ATT CTT GCG AAG GCT ATG GAT CGC ATT TTC ACC CTG CAT GCA GAC CAT GAA CAA AAT GCG TCT ACA TCA ACG GTT CGC CTT GCT GGC TCA ACG GGA GCG AAT CCG TAT GCA TGC ATC GCC GCC GGT ATT GCC GCG TTA TGG GGC CCC GCC CAC GGA GGA GCG AAC GAA GCC GTA TTG AAG ATG TTA GAT GAA ATT GGT AGC GTA GAG AAT GTC GCG GAT TTC ATG GAA AAA GTG AAA ACA AAA GAG GTG AAG TTG ATG GGC TTT GGC CAC CGT GTA TAC AAG AAT TTT GAT CCT CGT GCT AAA GTA ATG AAA CAA ACG TGT GAT GAA GTG TTG GCC GCA TTG AAC ATC AAC GAT CCT CAA TTA GCG TTG GCT ATG GAG TTA GAA CGT ATC GCT CTG AAC GAT CCA TAT TTT GTA GAA CGC AAA CTG TAC CCG AAC GTT GAT TTT TAC TCA GGG ATT ATC TTA AAA GCG ATT GGA ATT CCA ACA TCA ATG TTC ACC GTC ATT TTT GCT CTT GCC CGC ACA GTG GGA TGG ATT TCA CAC TGG CTT GAA ATG CAT TCC GGG CCG TAC AAG ATT GGG CGC CCC CGT CAG TTG TAC ACG GGG TAC GTG CAA CGC GAC CTT CCA CCG GTA GAG AAA CGT
Roseomonas rosea **WP_073132632.1**	ATG TCA GAT GCT TCG ACG CCG GGA TCG GTC ACG ATT ACG CTT GAC GGC ACC AAT AAA TCC TCG CGT GCC CCC TTG GTA CAG GCA TCC GTT GGC CCC GCA GTA GCA GAT ATT CGC AAG TTA TAC GCT GAC CTT GGC GTA TTT ACT TTC GAC CCG GGT TTC GGT ATG ACC GCT GCC TGC GAA TCA AAG ATC ACA TAT ATT GAC GGA GAT AAG GGA GTA CTG CTG TAT CGC GGG TAC CCC ATC GAA CAG CTT GCG GAA AAT AGC GAT TTT ACT GAA GTG TGT TAC CTG TTA CTG CAC GGT GAG CTT CCA AAC GAT GCG CAG TTA AAG GAA TTT AGC CAT AAT GTT ACA ATG CAC ACT ATG GTC CAT GAA CAA ATT CGC AAT TTC TTC AAT GGG TTT CGC CGC GAT GCT CAC CCG ATG GCG ATT TTG TGC GGC GTC GTC GGA GCC TTG TCT GCG TTC TAC CAT GAC TCA TTG GAT ATC TCT GAC CCG CGT CAA CGC GAA ATT GCG GCC TTT CGT CTT ATT GCC AAA GTA CCG ACG ATC GCT GCT ATG GCG TAC AAG TAT TCA ATC GGC CAA CCT TTC GTG TAC CCA CGC AAT GAC CTT TCA TAC GCC GAA AAC TTC TTA TAT ATG CTG AAC GCC GTT CCG GCG GAG GAA TAC AAG GTG AAC CCC ATT CTG GCA CGT GCG ATG GAT CGT ATT CTG GTC TTA CAT GCC GAT CAT GAG CAG AAC GCT AGC ACT AGC ACG GTA CGT CTG GCT GGC AGC ACA GGT GCA AAT CCG TAT GCC TGC ATC GCA GCG GGT ATC GCG GCT CTT TGG GGA CCA GCG CAT GGG GGA GCC AAT GAG GCG GTT CTT AAG ATG TTA GGG GAA ATT AAA ACT CCC GAA AAC ATC CCT GAC TTC ATC GAG AAG GTG AAG GAT AAG AAC TCA TCC GTG AAA CTT ATG GGG TTT GGT CAT CGT GTA TAC AAG AAT TTC GAC CCA CGC GCT AAA ATC ATG AAG GAA ACA TGT CAC GAA GTA TTG GCA GAG CTG GGC ATC AAG GAT GAA CCT CTG CTT GAC ATG GCT ATG GAG ATG GAA CGT ATT GCT CTG TCC GAT GAC TAC TTT GTA AGC CGT AAG CTG TAT CCT AAC GTG GAC TTC TAC TCC GGG ATT ATC CTT AAA GCT ATG GGA ATC CCC ACA TAC ATG TTC ACT GTT CTT TTT GCC GTT GCC CGC ACG GTA GGC TGG GTT TCG CAA TGG AAA GAA ATG ATC GAA GAG CCG GGG CAA CGT ATC GGA CGC CCG CGT CAG GTG TAC ACA GGA GAA ACT AAC CGC GAT TAC GTG CCG ATG GGA AGT CGC GGG
Sorangium cellulosum **KYF62269.1**	ATG AAA AAG AAT GGA GCG GAG CAG ACT ATT GAG CTT CCT GTT ATT ACA GGG ACA GAG GGA GAG AAA GCT ATT GAC GTA GCT TCC CTG CGT TCG AAA ACC GGC TAC GTG TGC ATC GAT CCC GCC TTC GTC AAT ACT GCG AGC ACG ACA TCA TCG ATT ACA TTT TTA GAT GGT GAG AAA GGG ATT TTG CGC TAT CGC GGT ATC CCG ATC GAA CAA CTG GGG GAA AAA TCT ACG TTT GTA GAA ACT TCA TAC TTG TTA ATC TAT GGC AAG TTG CCA AAT AAG AGC GAA CTT TCC CGT TTC TCG ACG CTG CTT ACT CGT CAT AGT CTT ATC CAC GAG GAT ATG AAG CGC TTC TTC GAT GGG TAT CCG TCC ACG GCT CAC CCT ATG GCA ATC TTA TCC GCT ATG GTC CTG AGT TTG TCG AGC TAT TAT CCA GAA GCG ATT GAT GTC AAG AAT ACG GAT TTT TTG GAT ATT ACG ATC GCT CGC TTA CTT TCC AAA GTT CGT ACT ATC GCC GCT TTC TCG TAT AAA AAG TCG ATT GGA CAG CCC TTT GTC TAC CCA CAT AAC TCA CTG TCA TAT TGT GCA AAT TTC TTG AAC ATG ATG TTC TCA GTT CCC GCC GAG CCT TAT GAG ATT GAT GAA GAC ATC GTC AAG GTG CTT AAC CTG TTG CTG ATT TTA CAT GCT GAT CAC GAA CAG AAT TGC TCC ACA TCC ACT GTT CGC CTG GTA GGT AGC GCA AAA ACC AAC TTA TTC GCG TCC ATC GCT GCC GGC ATC TGT GCG CTT TGG GGA CCC TTA CAC GGC GGT GCA AAC CAG GAG GTA GTC ACA ATG TTA GAG GCA ATT CGC GCC GAT GGT GGT GAC GTT AGT AAG TTT GTA AAG CTG GCA AAA GAC AAG ACA TCC GGT TTT CGT TTG ATG GGT TTC GGC CAT CGT GTA TAT AAA AAC TAT GAC CCA CGC GCG GTT TTG ATT AAG GCA GCA GCG GAT AAG ATT CTG AGC AAA TTA GGG ATT AAA GAT CCC CTT CTT GAT ATC GCG CAA CGT CTG GAG GAG ACT GCT CTT AAG GAT CCT TAT TTC GTA GAA CGC AAG TTA TAC CCC AAC GTA GAC TTC TAT TCC GGT ATC ATC TAT CGC GCG CTT GGC TTC CCA ACG AAT ATG TTT ACG GTT ATG TTC GCT TTA GGT CGC CTG CCC GGC TGG ATT GCG CAC TGG AAA GAA ATG AAT GAT GAT CCC AGC GGT AAA ATC GGG CGC CCA CGT CAG ATT TAC ACG GGG GAA AAT GCA ACA GAT TAT ACT CCG TTA GAC GAC CGT GCG
Thioclava pacifica **WP_038072739.1**	ATG GCA GAG AAC CAG AAA ACC GCA ACG CTG TCC TTG GAT GGT ACC ACT TAC GAG TTG CCC GTA TTG AGT CCC TCA GCG GGC CCA GAT GTT CTT GAT ATC CGC AAG CTG TAC GGG CAA GCT GAC GTT TTC ACA TAT GAT CCG GGA TTC ACG TCT ACA GCC TCT TGC GAA AGC AAG ATC ACC TAT ATC GAT GGC GAT AAG GGG GAA TTG CTG TAC CGC GGG TAC CCG ATT GAC CAA TTG GCT GAG AAA TCC CAT TAC CTG GAG GTA TGT TAC CTT CTT CTT TAC GGG GAA TTA CCA ACT CGC GCC CAA ATG GAG GAC TTC GAG TAC CGC ATC ACG CGC CAT ACA ATG ATC CAC GAG CAG ATT CAT AAT TTC TTC CGC GGG TTT CGC CGT GAT TCG CAT CCC ATG GCA ACC ATG GTT GGA GTA GTG GGC GCA ATG TCG GCA TTC TAT CAC GAT AGT TTA GAT ATC AAC GAT CCG TGG CAG CGT GAA GTA GCT TCA ATG CGC TTA ATC GCC AAA CTG CCG ACG ATC GCG GCT ATG GCG TAC AAA TAT TCG ATC GGA CAA CCC TTT GTT TAT CCG CGC AAC GAC CTT AGT TAT GCC GAA AAT TTT CTG AAT ATG TGT TTC TCG GTA CCA GCA GAA CCC TAT AAG GTG GAA CCG GCG CTG GCT AAG GCT ATG GAC CGC ATC TTT ACG TTA CAT GCT GAC CAT GAA CAG AAC GCA TCA ACA AGT ACC GTG CGT CTT GCT GGT TCC AGT GGG GCA AAC CCG TTT GCC TGT ATT GCG GCG GGA ATC GCT TGT TTA TGG GGG CCA GCA CAC GGT GGA GCA AAT CAG GCA TGT CTT GAA ATG CTG CGT GAA ATC GGA TCT GTC GAC AAG ATT CCT GAG TAT ATC AAG CGC GCA AAA GAC AAA GAC GAT CCA TTT CGC CTG ATG GGA TTC GGT CAC CGC GTG TAC AAA AAT TTC GAT CCT CGT GCT AAA GTA ATG AAA GAA TCA GCG GAT GAG GTA TTA GAT TTA TTA GGG ATT CAC GAT AAC CCA ACC CTT CAG GTT GCC AAG GAG TTA GAG AAG ATT GCT CTT GAA GAC GAA TAT TTC GTA TCT AAG AAG CTT TAC CCT AAT GTA GAT TTT TAC TCC GGC ATC ATT CTT GAA GCA ATG GGC TTC CCC ACT TCG ATG TTC ACG CCT ATT TTT GCA CTT TCC CGC ACT GTA GGT TGG ATC TCG CAG TGG AAA

	GAA ATG ATC GAA GAC CCC ACG AAT AAG ATT GGG CGC CCC CGC CAG CTT TAT ACA GGT GCT ACA TTC CGC GAC TAC GTG GAT GTT GAA AAG CGT
Nitrospira inopinata **WP_062485514.1**	ATG ATG CCC CAC GAT TTT ATG CCA GGT CTT GCA GGG GTG CCA GCG GCA ACT AGT TCA ATC TCC GAC GTT GAC GGC CAG CGT GGA GTT TTG GAA TAC CGT GGG ATT CGT GTA GAG GAG CTG TGC GCC AAG TCG TCA TAT TTG GAA ACA GCA TAT TTA TTA CTG TTC GGG CGT CTT CCT ACT CGC GCA GAA CTT CAC CGT TGG ACC GCG GAC GTA ACC CAC CAT CGT CGC GTA AAA TTC CGT ATC GTA GAC TTG TTG AAG TGC TTA CCC GAG CAG GGC CAT CCC ATG GAC GCT CTG CAA GCC GCG GTC GCG GCC CTT GGG ATG TTT TAC CCT GGG CGC AAC GTC AAA GAT GCT GAC AAC AAT TAC TGG TCG GCC GTC CGT CTT GTG GCG AAG TTA CCC ACT ATT GTC GCG GCA TGG GCT CGT ATT CGC CGT GGT GAC GAC CCC ATT CCA CCA CGT GAT GAC TTG GGA TTT TCG GAG AAT TTT CTG TAT ATG CTT ACA GAA ACT GCG CCC CCG CCA TTA TGG AGT GAA GTC TTT GAC GAC TGT CTT ATT TTA CAC GCT GAA CAT ACA ATG AAT GCT AGC ACT TTC GCA GGG CTG GTA ACT GCA AGC ACA TTG GCA GAT CCA TAT ACG GTT GTG GCG TCA TCT ATT GGC GCA CTG AAA GGC CCA CTG CAC GGA GGT GCC AAT GAG GAA GTA GTT CAG ATG TTA ATT GAG ATT GGG TCT CCA GAT CGT GCT CGT ACG TAT GTT GAA GAA AAG CTG CGT GCT AAG CAA AAG TTA ATG GGC TTT GGT CAC CGC GTT TAC AAG GTA AAA GAC CCT CGC GCA ACG GTA CTT CAA GAA CTG TGC CGT CGC TTA TTT ACG GAG AGT GGA CCG TCA CCC TTT TAC AAT ATC GCG TTA GAG GTT GAG CGT GCA GCG GGA GAA CGT TTG GGA GAT AAA GGG GTT TAT CCC AAC GTG GAT TTT TAC AGT GGT ATT ATC TAC GAC AAG ATG GGG ATC GAG ATG GAT CTT TTT ACT CCT ATT TTC GCG ATG GCC CGC GTC GCT GGA TGG CTG GCT CAT TGG CTT GAA CAA CTT GGT GAA AAT AAA CTT TTT CGC CCA GAC CAG ATC TAC TCA GGA GAA CAT AAC CGC CCG TAC GTT CCG ATC GAC CAG CGC
Cyanobium sp. **PCC7001** **WP_006910478.1**	ATG GCC GGG AGC TTG AGT GAT AGT GTA CCT GGA TCG ACT GGA GGG GCA ACG GCT GCC CCA CCA TTT CGT CCC GGG TTG GAA GGT GTA CCA GCA ACT CAG AGC GCG ATC TGT GAC ATT GAT GGA CAA AAG GGT CGT TTA ACA TAC CGC GGG TAC GAC GCG GGG GAG CTG GCC GCT CAT AGC ACG TTC TTG GAA ACA ACC TAC TTA TTA ATT TGG GGA GAG CTT CCC ACC GCC GAA GGT TTG CGT CAG TTT GAA CAC GAG GTT CAG ATG CAT CGC CGT GTT TCC TTC CGT ATT CGT GAC ATG ATG AAA TGC TTC CCC GCT ACA GGC CAT CCC ATG GAT GCA TTG CAG TCG TCC GCC GCC AGC TTG GGT CTG TTT TAT TCC CGC CGT GCG CTG GAC AAT CCC GAG TAT ATT GCA GAA GCG GTG GTA CGC TTA ATT GCC AAA ATT CCC ACT ATG GTA GCA GCC TTC CAA CTG ATT CGC AAG GGC CAG GAT CCG ATC CAA CCG CGC GAC GAT CTT CCA TTC GCG TCG AAC TTT CTT TAC ATG CTG ACT GAG CAG GAA CCT GAT CCT CTG GCG GCT CGC ATC TTC GAT GCA TGT CTG ATC TTA CAT GCT GAG CAT TCT CTG AAC GCA AGC ACG TTC TCG GCG CGC GTC ACT GCT TCT ACT TTG ACT GAC CCC TAC GCC GTC GTG GCG AGC GCA GTT GGC ACG TTG GCT GGC CCA CTG CAC GGT GGC GCA AAC GAG GAT GTC CTG GCC ATG TTG GAG GCT ATC GGG AGT GCA GAC CAG GTG GAA CCG TGG TTG GAC CGT GCA ATC GCC CAG AAA CAA AAG ATT ATG GGT TTT GGT CAT CGC GAG TAC AAG GTG AAA GAC CCT CGC GCC GTG ATT TTA CAG GGC TTA GCG GAA CAG CTT TTC CAC CGT TTC GGG CAC GAC CCC TTG TAT GAC TTA GCG CGT AAA CTG GAA GAA GCT GCG GCA GAA CGC TTA GGC CCG AAA GGT ATC TAC CCG AAC GTA GAT TTT TAT TCC GGC TTG GTC TAC CGT AAG TTA GGG ATC CCG CGC GAT CTT TTT ACA CCT ATT TTT GCC ATC GCC CGC ACA GCA GGT TGG CTT GCG CAC TGG AAG GAG CAA TTA GGT GCA AAC CGC ATC TTC CGT CCA TCA CAG ATC TAC ACA GGA CCT GTA CCC CGC GAC TGG GTC CCC CTT GAA GCC CGT
Cyanothece sp. **SIO1E1** **NET35916.1**	ATG ACA TTC TGT GAA TAT AAA CCC GGG CTG AAG GAA TTC CGG CGA CCC AAA GTT CAG TAA GTT ATG TTG ATG GTC AGC GTG GAA TCC TGG AGT ATC GCG GCA TCC AAA TCG AAG CAC TTG CCG CTA AGT CAA ATT TCC TTG AGA CAG CCT ACC TTT TGA TCT GGG GTG GTT TAC CTA CTC ATG AGG AGT TGG CGT CAT TTG AAT CGG AAA TTC GCT ACC ACC GTC GTT TGA AGT ACC GTA TCC GCG ACA TGA TGA AGT GCT TTC CGG AAA GCG GTC ACC CGA TGG ATA GCC TGC AGG CGT GTG CAG CCG CCC TGG GTC TGT TCT ATT CGC GTC GCG CAC TTG ATA ACC CTG TTT ATA TTC GCG CTG CGG TAG TTC GTC TGC TTG CCA AAA TCC CTA CTA TGG TAG CTG CTT TTC AGA TGA TGC GCA AAG GAA ATG ACC CCA TCC AAC CGC GCG ACG ATC TGA GCT ACT CTG CCA ACT TTC TGT ATA TGC TTA ATG AAC GTG AAC CGG ATC CGT TAG CTG CTC ACA TCT TCG ATG TGT GTC TTA CGC TTC ATG CCG AGC ATA CTA TCA ACG CGT CAA CCT TTT CCG CAA TGG TGA CCG CCA GTA CAC TTA CAG ATC CCT ACG CGG TTG TCG CGT CAG CCG TGG GTA CAT TAG CAG GAC CCC TTC ATG GAG GCG CGA ATG AAG AGG TGC TTT CTA TGC TGG AGG AAA TTG GAA GCG TGG GTA ATG TGC GCC CGT ACT TGG AAG ATT GTC TGC AGC GCA AAG CGC GCA TTA TGG GAT TCG GTC ACC GCG TGT ACA AGG TCA AAG ATC CAC GCG CCA TTA TTC TGC AAG ACC TGG CAG AAC AGT TGT TTG AAA AGC TTG GTG GCG ATC GTT ACT ACG ACA TTG CAG TAG AGC TTG AGC GCC AGG TCT CTG AAA AAT TGG GGC ACA AAG GCA TCT ATC CAA ACG TCG ACT TTT ATT CGG GAT TGG TCT ATC GTA AGT TGG GTA TTC CCC GCG ATA TGT TCA CGC CGG TCT TCG CTA TTG CCC GTG TTG CTG GGT GGT TGG CGC ATT GGA AGG AGC AAC TTG CTG AGA ACC GCA TCT TTC GTC CGA CTC AGA TCT ACA CTG GAC CTC GCC ACA TTC CTT ATG TAG CTA TTG CTG ATC GTC ATC CAC CCC AGG AGG AGG TCA TTC TTA CTA ATT TAC TTG CAG ATT
Gemmata obscuriglobus **WP_010038645.1**	ATG AGT ACT GAA ATT GAG TAT AAG CCT GGC TTG GAG GAT GTT CCT GCA GCA AAA AGT GCC GTA TCC TTC CTG GAC GGG AAG AAA GCC GTT TTA GAG TAT CGC GGT ATC CCA GTT GAA GTG CTG GCT AAG GAG TCG TCG TTC GAG GAG GTT TCA TGG TTA TTG GTC AAA GGG GAT TTA CCA ACA CAA AAG CAG CTT GCT GAA TTC GAC CAC GAT CTG CGT CAG CGC CGT GCC ATC CAC TTC CGC CTG AAA GAT CTT ATT AAA TGT ATG CCT GCT GAT GGC CAC CCG ATG GAT GCC CTG CAT GCT GGG GTC GCT GCG CTG GGC ATG TTT TAC CCA TGT CCT ACA GTC TCC AAT CCA GCC AAG AAT TGG GAT GCA ACC TGC CGC CTT ATT GCT GCC CTG CCA ACT CTT GTG GCC GCC TTC GCT CGC GTT CGC CGT GGA GAG GAG ATT TTA GAT CCA CGT AGC GAC CTT GAC CAC GCT GGG AAC TTT TAT TAC ATG TTA TTC GGT AAG GAG CCT TCA CCT GCA ACT CGC AAA GTC TTA GAT GCC TGC TTG ATC CTG CAT GCA GAG CAT CAA ATG AAC GCA TCA ACT TTC ACA GCT CGT GTT ACA GGT TCC ACG CTT GCT ACG CCG TAC CAT ACT ATT GCC TCT GCA ATC GGT AGT TTA TCA GGG CCC TTA CAT GGC GGG GCT AAT GAA GAG GCA TTA CGC CAG TTT GAG GAA ATT GGA GGT CCC GAA CAG GTA AAA GGA TGG CTT GAC GCA AAA CGC GCC GTG GAT AGC AAA TAC AAA GTG ATG GGT ATG GGA CAC CGC GTT TAT AAA GTA AAG GAT GCG CGC GCA ACG GTT TTA CAG GAG ATC GCG GAA CAT ATG TTT GCT GAG ACT GCC CGT CCT AAG ACT TAT GAA ACG GCA CTG GAG TTA GAG CGC CTT TGC GCT GGT ATT TAC GGC CCA AAG GGA ATC TAT CCC AAC GTC GAT TTT TAT TCA GGA GTG GTT TAT CAG TCA TTG GGT ATT CCC ACC GAT GTG TTC ACT CCT ATT TTC GCT ATC GCC CGC GTA TCT GGG TGG CTG GCA CAT TGG ACT GAG CAG CTT GTT GGC AAT CGT ATT TTC CGT CCG GAA CAA ATC TCG ATC GGA AAA ACA GAT GTG AAA TAC GTG CCG TTA GAG CAG CGC GCC

Gloeobacter violaceus **WP_011143005.1**	ATG TCA GGG GAG TAT GTG CCT GGA TTA GAA GGC GTA CCA GCA ACA CGT TCA AAT ATC TCG TTT GTG GAC GGA AAA GCA GGT GTG TTA GAG TAT CGC GGT ATT CCG ATC GAT CAG TTG GCC GAG TCA TCT ACA TTT TTA GAA ACA GCT TTT TTG CTG ATT TTC GAT CAT TTG CCT ACG AAG GAC GAA TTA TTG AGC TTC GAA GTT GAG ATT TTA GGC CAT CGT CGT GTA AAA TAT CGC ATT CGC GAC ATG ATC AAA AGC TTC CCC GAG AGT GGC CGT CCC ATG GTC GCC CTT CAG TCT TGC ATC GCG GCG CTT GGC TTA TTT TAC CCC TTG CAG AAC GAA TCG AAA GAA AAA TAC GCC TAC GAT AGT ACT ATC CGT TTA TTA GCT AAA ATG CCC ACC ATG GTA GCC ACC TTT CAT CAA ATG CGC CTG GGG AAC GAT CCC ATC CCA CCA CGC GAT GAT TTG GGG CAT GCC GCA AAC TTC TTG TAT ATG TTG ACG GGA AAA GAA CCG GAC CCG CGT GCA GCT CGT ATC TTT GAC GTT TGC CTG ATG TTG CAT GCA GAA CAC ACG GTC AAT GCT TCT ACC TTC AGC GCC CTT GTT ACG GCA TCA ACT TTG GCA GAC CCA TAC ACT GTA ATC ACC AGT GCG GCA GGA ACA TTG TCC GGC CCT CTG CAC GGA GGT GCG AAC GAG GAA GTC ATT CGC ATG TTA AAG GAA ATC GGC ACC ATC GAA CGT GTA CGC CCA TAC TTA GAA AAT CGT TTA GCT CGT AAA GAA AAG ATC ATG GGC GTC GGC CAC CGC GTA TAT AAG GTG AAA GAC CCA CGC GCA ACG ATT CTT CAA AAC TTA GCT CAA GAG CTT TTT GAT CGC TTC GGC CAC AGC CGT TTA TAC GAC ATT GCC GTG GAA GTG GAG CGC GTT TGC GAC GAG CTT TTG GGT CAA AAA GGA ATC TAT CCG AAT GTA GAT TTC TAT TCA GGG TTG GTC TAT GAG AAA ATG GGG ATT CCG GCT GAT ATG TTT ACA CCG GTC TTT GCG ATT TCC CGC GTC GCC GGA TGG CTG GCA CAT TGG CAT GAA CAG TTG GCG GAC AAT CGT ATT TTC CGC CCA ACG CAA GTT TAT ACT GGC AGC CAT AAC GTG GAG TTC ACG CCC CTT TCG TTG CGC AGC TAC GCA
Mariprofundus aestuarium **WP_100278409.1**	ATG GTG GAT AGC CAG AAC GCA GAA GAT TCG ATC GCG TGT TTT TCC CCT GGT TTA GCA GGT ATC CCC GTA GCA GAG TCA GCT GTT TCA TAC GTT GAC GGA AAG AAA GGG TAT CTG GAA TAC CGT GGA ATT GAT ATC GAG CAG CTG GCC GAG AAA AGC TCT TTT GAA GAG TCG AGT TAT CTT TTG CTG TAC GGA GAG CTG CCA ACG GCT AAA GAG TTA AAA GCG TTC GAT GAG AAG TTG AAG CAT CAC CGT CGT ATC AAA TTC CGT ATT CGC GAC ATG ATG AAA TGC TTT CCA GAG TCA GCA CAT CCC ATG GAT AGT CTT CAA GCA ACG GTG GCG GCT CTT GGT ATG TTT TAT CCG TTC CCA ACC AAT CTT GAA GGA GAA TTG GAC AAA GAA ACG ATT GAT AGC GTC TGT GTT AAT TTA ATT TCA AAA ATG CCT ACG CTT GTA GCC GCT CAC GCG CGT ATG CGT CAT GGT GAC GAC CCC ATT GCC CCG CGT GAC GAC TTG TCT CAC GCT GCT AAT TTC TAC TAC ATG CTT ACG GGA GAT GAG CCG TAT CCC CTG ACC GAG CGC ATT TTG GAT GTC GCC TTC ATC GTA CAT GCA GAA CAC GAG ATG AAC GCA TCC ACC TTT AGT GCG TTG GTC ACC GCC TCT ACC CAT GCG GAT CCA TAC ACA TCG ATT TCA GCA GCC ATT GGC GCG CTT TCC GGT CCG CTT CAT GGA GGC GCC AAT GAG GAC GTG TTA CAC ATG CTG GAT GAG ATC GGA AGT GTG GAC AAT GTA GAA GAC TAC TTC AAT CGT AAA TTG GCG AAG CGC GAA AAA TTT GCC GGC CTT GGT CAC CGC GTC TAT AAG ACG AAG GAC CCA CGC GCC ACG CTT TTA CAG AAA CTG TAC GTT GAA CTG ACA GAA AAG GTG GGG GTA GAC ACA ACG TAT GAG ATC GCA CGC CAG ATT GAA AAA TTA TCG CGT TCG ACA TTA GGG GCT AAG GGA GTT TGC CCA AAT GTT GAT TTC TAC AGC GGG ATC GTT TAC CGC AAA ATG GGA ATC CCG ACC GAT CTT TTC ACA CCG ATT TTT GCT ATC GCC CGC GTG TCG GGT TGG CTG GCA CAT TGG AAA GAA CAA CTG GGT CAG GGG CGC ATC TAC CGT CCA AGC CAA ATT TAT ACG GGT AAG TAC GGC CAA TCC TAT GTA CCA GTC GAA GAA CGC
Methylophaga sulfidovorans **WP_091715369.1**	ATG AGC GAC TTT CTT CCT GGG CTT GAA GGG GTT CCA GCT ACC AAA AGC GCA ATT TCT TTC ATT GAC GGG GAG AAA GGC ATT CTT AGC TAT CGC GGT TAT CCT CTG GAG ACG CTT GCT GAA AAT AGC ACT TTC GAG GAA ACG ACG CTT CTG CTT TTA GAT GGC GAG TTG CCG ACG AAG AAG GCG CTT AAT GAC TTT AGC CAG CAA CTG AAG GAC AAT TAT CGC ATC AAG TAT CAC ATT CGC CAA ATG ATG CGT CAT TTT CCT CAT ACA GGA CAT CCG ATG GAT ATG CTT CAG ACA GCC GTT TCC TCT TTA GGC ATG TTT TAT CCC GGC ACT GAA TGC CTG ACT GAC GCC AAC TCC TGC GAA GAC CTG GAC TAT GTT CGC AAT ATG ACA GTG AAT ATT ATT GCA CAG ATG GCC CCA TTA GTG GCG ATG TGG GAA CAT ATC CGT AAT GGA TGG GAC CCT GTT AAC CCA AAG CAT GAC CTT AGC GTC GCC GAG AAC CTG CTT TAC ATG TTC AAT GGG GAA GAG CCG GAT CCT CTG ATG GCG AAA ATC ATG GAT GTG TGC CTG ATC CTG CAT GCC GAG CAT ACA CTG AAT GCT TCT ACC TTT GCC GCG TTA GTA GCT GGG TCA ACA CTG GCG ACC CCA TAC TCC GTT ATC AGT GCG GCA ATC GGG ACA TTG TCG GGT CCA TTG CAT GGT GGA GCG AAT CAG CGT GTC GTT GGC ATG CTG CAG GAG ATT GGG AGC CCG AAG AAC GTC GAA AGT TGG GTA GAC GAA AAA TTA AAA AAC AAG GAA GTT ATT TGG GGA ATG GGT CAC CGC GAA TAT AAG GTC AAG GAC CCG CGC GCG ACT ATT TTG CAT AAA CTG GTA GAG CAG TTG GTA GCC GAA CGT GGG GGT CAC CTT GAC GAT ATG TTT GAC ACA GCC TTA AAA CTT GAG GAG GTT TGC GCT GAT CGT CTT GGG CAC AAA GGC GTT TAT CCT AAT GTA GAC TTT TAT TCT GGC ATC TTG TAT TCG GAA ATG GGT ATC CCC GAG GAC GAG TTT ACG GCC TTG TTC GCT GTG GCT CGC AGC GCA GGA TGG CTG GCT CAT TGG CGT GAA CAG ATT TCA GAT AAT CGT ATC TAT CGT CCT ACC CAA ATC TAT GTC GGC TCC GAT ATG CGC GAT TAC ACA CCA ATC GAA GAG CGC
Myxococcales bacterium **MAD59944.1**	ATG TCA TTT ACT CCT GGA TTA GCT GGC GTG ATT GCC GCG GAG TCT TCA ATT TCA ACA ATC GAC GGA CAA AAC GGG ATC CTG CGT TAT CGC GGT ATT CGC ATC GAG GAT TTG GCC GAG AAT TCC TCC TTC GAG GAA GTA GCC TAT TTG TTA TTG TTT AAC GCA CTG CCT ACA ACA GAG CAA TTA TCC GCC TTC GAC GCG CGT TTA AAG GAA CAC CGT TCA ATC CCG GAA GGC ATC GTG ACA TTA CTG GGG ACA TTA CCG GAT ACA ACG CAT CCA ATG GTT GCC TTG CAG TCA GCA GTG AGC GCG CTT GCC GCG TAC TAC CCA CAT TTT GAA GTG GAA GAC ACA GTG AAA AAT GAA GAG AGT ATC ATC CGT ATG ATC GCC TGT TTT CCT ACC ATC GTT GCC AGC TTC GAT CGT TTG CGC AAG GGA AAA GAA GTT TTA GCG CCA AAC AGC GAT CTG GAC CAT GCA GCC AAC TTT CTT TAT ATG CTG AAC GGA GAA GTT CCA GAC GAG AAA CAA CGT CGC ATT ATC GAC GTC GCA TTG GTG TTG CAC GCG GAG CAT GGG TTC AAC GCT TCT ACT TTT ACA GGT CGC GTA GTG GGT TCC ACA ACT GCC AAT CCG TAT GGT AGC CTT TCA GCA GCA GTT GGT TCT TTG TCA GGG CCT TTA CAT GGG GGT GCT AAC GAG CGT GTA TTG CAC ATG TTG GCG GAG ATC GAC GGT GGG GTA GAC GGC GTA GAG AGC TGG TTT ACC CAA GCG CAA GCT GAG AAG CGC AAA ATT ATG GGT CTT GGA CAC CGC GTT TAT AAA GTA AAG GAC CCC CGC GCC ACT GTG CTT CAG GGA ATG GCC CGT GAT TTA TTC GCT TCT CAC GGC AGC ACA CCT TTA TAT GAC ATG GCG GCG CGT TTG GAG GAG ATT GCA AAA GAA CGT TTG GGC GAG AAG GGG ATC GCG CCC AAT GTC GAT TTT TAC AGT GGC TTA GTA TAC CAG AAA CTT GGT ATG GAG ATC GAC CTG TTT ACT CCG TTC TTT GGA ATT GCG CGT ATT GCG GGC TGG GGC TCA CAC TGG ATT GAG CAA CTG GGT AAC AAT CGT ATC TAC CGT CCG ACT CAA CTG TAT GTA GGA GCG ACA GAA GCG CAG TAT GTT CCT ATG TCT GAA CGC
Prochlorothrix hollandica	ATG TCG ATT GAC GAA TAC AAG CCC GGA TTA GAG GGT GTA CCG GCG ACG TTG TCA TCA ATT TCC TAT GTT GAT GGG CAA AAG GGC GTA CTG GAG TAT CGT GGA ATT CCC ATC GAA CAA CTT GCT CAA AAC TCA AAT TTC CTG GAA ACA GCG TAT TTA CTG ATT TGG GGT ACA CTG CCG AGT

WP_017713050.1	ACC GAG GAA TTG ACG AGT TTT GAA CAA GAA ATC TAC CAA CAC CGC CGC CTT AAA TAT CGC ATC CGC GAT ATG ATG AAG TGC TTC CCG GAA AGC GGG CAC CCG ATG GAT GCT CTT CAA TCG TGC GCA GCC GCA TTG GGG CTT TTC TAT GCC CGC CGT GCA CTG GAC GAC CCA GCC TAC ATC CAT CGT GCC GTC GTG CGT TTG TTA GCG AAA ATT CCG ACC ATG GTA GCG GCT TTT CAA CTT ATG CGT AAG GGA AAT GAT CCG GTA CAA CCC CGC GAT GAT CTT AAT TAC GCC GCA AAT TTT CTG TAT ATG CTG AAC GAA AAG CAA CCG GAT CCA TTA GCA GCT CGC ATC TTC GAC GTG TGT TTA ATC CTT CAT GCA GAG CAC ACC ATT AAC GCT TCG ACA TTC TCT GCC ATG GTC ACG GCC AGT ACT TTG ACC GAT CCG TAC GCG GTA ATT GCT AGT GCA GTC GGA ACG TTG GCC GGT CCT TTA CAC GGC GGC GCA AAC GAA GAG GTG CTT GAC ATG TTA GAG AAC ATC GGC ACT GTG GAC CGC GTA GCC AAC TAT GTT GAC GAA TGC ATC GCA ACT AAA AGC CGT ATC ATG GGA TTC GGA CAC CGC GTA TAC AAA GTA AAA GAT CCG CGT GCG ACC ATT TTG CAG GGA TTA GTA GAG CAG CTT TTT GAT GAG TTT GGC TCT GAT TTG TAC TAC GGA ATC GCC CTT GAG CTT GAA CGT GTT GTA AGT GAT CGT TTA GGA CAT AAG GGG ATT TAC CCT AAT GTT GAC TTC TAC TCA GGA TTA GTA TAT CGC AAG CTG GGT ATT CCC ACA GAT TTG TTC ACC CCA GTT TTT GCT ATC GCG CGT GTC GCG GGT TGG CTT GCG CAT TGG AAG GAA CAG TTG GCT GCA AAC CGT ATT TTT CGC CCC ACG CAA ATC TAT ACA GGA AAC CAC GGC CAG TCT TAC GTG AAG TTG GAC GAC CGC TCT ACC CCT TTG CCC CCA GGT CTT CTG GCA CTG GGA CAG GAG TTG
Pseudanabaena biceps **WP_175355653.1**	ATG GCC ATC GGC GAG TAT AAA CCA GGA TTA GAA GGC GTG CCC GCG ACC CAG AGC AAC ATT TCA TAC GTC GAC GGA AAA GCG GGG CTT CTG GAG TAT CGT GGG ATC CGT ATC GAA GAA TTG TGC GTA CAC AGT AGT TTC TTG GAG ACA TCA TAT CTG TTA ATC TTT GGT GAG TTA CCT ACG TCG GCG AAG TTA AAG GAG TTT GAA GTT GAT ATT ACC CAC CGT CGT CGC ATC AAA TAT CGT ATC CGT GAT ATG ATT AAA TCG TTT CCT GAT AAC GCA CAC CCC ATG GAT GCC TTG CAA ACC AGT GTG GCG GCA CTG GGC ATG TTT TAT CCG TTG GGA GAT TTC CAC GAC GCG GAT TAT ATC TAT CAG GCG ACC GTG CGT CTG CTT GCC AAA GTG CCC ACT ATG GTT GCA GCC TTC CAT ATG ATG CGT CAA GGC AAT GAT CCA GTT ATG CCA CGC GAC GAT CTG GAC TAT GCG TCG AAC TTC CTT TAC ATG CTG AAT GAG AAA GTG CCG GAC CCT TTA GCC GCG CGT ATC TTT GAC GTT TGT CTG ACT TTA CAC GCG GAA CAC ACT GTT AAT GCA AGC ACG TTT GCG GCG TTG GTC ACG GCT TCA ACG TTA ACG GAT CCC TAC GCG GTT ATT ACA TCG GCT ATC GGA ACG TTA GCC GGC CCA CTG CAC GGG GGC GCC AAC GAA CAA GTG ATG ATG ATG TTG GAG GAA ATC GGT TCC GTT GAT AAT GTT ACC GCG TAT TTA GAA CGT AAG ATC GAG CGC AAG GAG AAG CTT ATG GGC TTT GGC CAC CGT ATT TAC AAA GTA AAG GAT CCG CGT GCT ATT GTA TTG CAA GAG TTA GTC CAT AAA ATG TTT GAC CAG TTC GGA CAT GAC CAT TAC TAC GAT ATC GCC TTG GAG TTA GAG AAA CAG GCA TTT GAG AAA CTT TCG TCA AAG GGA ATC CAC CCG AAC GTA GAT TTT TAC TCT GGG TTA GTC TAC AAG AAG CTT GGA ATC CCT AGC AAC TTA TTC ACG ACA ATC TTC GCT ATC GCC CGT GTG CCA GGT TGG CTT GCT CAT TGG AAA GAA CAA TTG AGT GAC AAT CGT TTA TTC CGC CCG ACT CAG GTC TAC ACG GGG CTT CAT GAT GTT ACG TAT TTA CCT ATT GAA CAT CGT
Sandaracinus sp. **MAQ19445.1**	ATG CTG ATG GCA GAT AAG CCT GAC TAT GTC CCT GGG CTG GCA GGT GTA CCT GCA GCT CGT TCC TCG GTG TGT CTT ATT GAC GGG CAA GTC GGG AAA CTG CAG TAC CGT GGA TAT CCT ATT GAA CAG TTA GCA GAA TCG TGT AGT TAT GAG GAG GTC GTC TAC TTG CTG CTG TTC GGA GAG TTA CCA ACC AAG GCA CAG CTT GAG AGC TTC GAT GCG GAA TTA AAG GAG GAG CGC GGA TTG AAG TTC CGT ATC GTC GAT GTC TTA AAG AAC TTG CCC GAG CGC GGC CAC CCG ATG GAC GCC TTA ACA GCA GCC GTC GCA TGC ATG GGC ATG TTT TAC CCG GGG GAT CAT GTC GAA GAT TTA GAA TTT CGC CGT CTG TGC GCA ATC CGC CTT GTT GCG AAA CTG CCT ACA GTT GTT GCT GCT TGG CAT CGC ATC CGC CGT GGG GAC AAC CCA ATC GAG CCA CGT ACA GAC CTT GGG CAT GCC GCG AAC TTC CTG TAC ATG CTG GAG GGA CAT GAG CCT GAT GAA TTA GAG GCC AAA GTC ATG GAT GTC GCC TTG ATT TTG CAT GCT GAA CAT TCC ATG AAC GCA TCT ACT TTC ACC GCG CGT GTA ACC GGT AGC ACA TTG TCA GAC CCG TAT GCA GCA GTT GCT AGC GCA ATC GGT AGC CTT GCT GGC CCA CTG CAT GGT GGG GCT AAC GAG CGC GTT CTG GAG ATG TTA CGC ACG ATC GAA GAT TCC ACA TCG GAG GCC GTG CGT CGC TGG GCA GAA GAC AAA CTG GCA CGC AAA GAG AAA ATC ATG GGC TTT GGA CAT CGC GTC TAT AAG GTT AAA GAC CCA CGT GCT AAC ATT TTG CAG AAA TTA GCC ACG CAA CTG TTT GAG AAG CGT GGG CGC ACA CCT ATC TAC GAC ATT GCA GTG GAG TTG GAA AAG CAA ATG GGA GAA TTA GTG GGT CAT AAG GGA ATT GCA CCG AAT GTT GAC TTT TAT AGC GGA ATC GTT TAC GAA AAA ATG GGC ATC CCA GTT GAT TTG TTT ACT CCT GTC TTT GCG ATT GCC CGC GTC GGA GGT TGG CTG GCC CAC CTT CTT GAA CAG TTG CAA GAC AAT CGT ATC TTT CGC CCG ACG CAA ATT TGG ATT GGT GAG GCC GAC CGC ACC GTA CCA CCT CTG GCT GAT CGT GGG
Synechococcus elongatus **BAD79102.1**	ATG ACT GCC GTC AGC GAG TTT CGG CCT GGC CTA GAA GGC GTG CCC GCC ACA CTC TCG AGC ATT AGC TTT GTC GAT GGC CAG CGC GGC GTC CTA GAG TAT CGC GGC ATC AGC ATC GAG CAA CTG GCG CAA CAG AGC AGT TTT CTG GAA ACC GCC TAC CTG TTG ATT TGG GGC CAT CTA CCA ACT CAG CAG GAA TTG ACC GAG TTC GAG CAC GAA ATT CGC TAC CAC CGC CGC ATC AAG TTC CGC ATC CGG GAC ATG ATG AAA TGC TTC CCC GAT AGC GGC CAT CCT ATG GAT GCC CTG CAG GCG AGC GCC GCA GCC CTC GGG TTG TTC TAT TCG CGG CGC GCC TTG GAT GAT CCC GAA TAC ATT CGG GCG GCC GTT GTG CGT TTG CTA GCC AAA ATT CCG ACG ATG GTG GCT GCC TTC CAG CTG ATC CGC AAG GGT AAC GAC CCA ATT CAG CCC CGC GAT GAA CTG GAC TAC GCC GCC AAC TTT CTC TAC ATG CTG ACG GAG CGC GAG CCC GAT CCA GTC GCA GCT CGG ATT TTT GAT ATT TGC CTC ACC CTG CAC GCC GAA CAT ACG ATC AAC GCC TCG ACC TTC TCG GCG ATG GTC ACA GCT TCG ACC CTG ACC GAT CCC TAC GCT GTC GTT GCT TCT GCC GTT GGC ACC TTG GCT GGC CCC CTC CAT GGC GGC GCC AAT GAA GAA GTG CTG GAC ATG CTG GAG GCG ATC GGT TCC GTC GAG AAT GTT GAG CCC TAC CTC GAC CAC TGC ATT GCC ACC AAA ACG CGC ATT ATG GGC TTT GGG CAC CGT GTC TAC AAA GTC AAG GAT CCG CGG GCA GTC ATT CTG CAA AAT CTG GCC GAG CAA CTG TTC GAT ATC TTC GGC CAT GAT CCC TAC TAC GAA ATC GCG GTC GCA GTC GAA AAG GCA GCA GCC GAG CGA CTC AGC CAC AAG GGC ATT TAC CCC AAC GTC GAT TTC TAC TCC GGC TTG GTC TAT CGC AAG CTC GGT ATT CCT AGC GAT CTA TTC ACA CCG GTG TTT GCG ATC GCG CGG GTT GCG GGC TGG CTC GCC CAC TGG AAA GAG CAG CTG AAC GAA AAT CGG ATC TTC CGC CCC ACT CAG ATC TAC ACG GGC AGC CAC AAC CTC GAC TAC ACC CCG ATC GCC GAT CGG GAT TTG GCG ATC GAA TCT GAT
Synechocystis sp. **WP_193386667.1**	ATG AAT TAT ATG ATG ACT GAT AAC GAA GTG TTT AAA GAA GGC CTA GCC GGA GTC CCC GCC GCT AAA TCG AGG GTG AGC CAT GTG GAT GGC ACC GAC GGT ATT TTG GAG TAC CGG GGC ATT CGC ATC GAA GAA TTA GCC AAA TCC AGT AGT TTT ATC GAA GTA GCC TAT CTG CTC ATC TGG GGT AAA TTG CCC ACC CAG GCA GAG ATC GAA GAG TTT GAG TAC GAA ATT CGC ACC CAT CGA CGC ATT AAG TAC CAC ATC CGG GAC ATG ATG AAA TGT TTT CCC GAA ACA GGG CAC CCC ATG

	GAT GCC TTG CAA ACT TCA GCG GCG GCC TTG GGA TTG TTC TAT GCT CGA CGG GCC TTG GAT GAC CCC AAA TAT ATC CGG GCG GCG GTG GTG CGT CTG TTA GCC AAA ATC CCC ACC ATG GTG GCA GCT TTC CAC ATG ATC CGG GAG GGT AAC GAT CCC ATT CAG CCC AAT GAT AAA TTG GAT TAC GCT TCC AAC TTC CTT TAC ATG CTG ACG GAG AAG GAG CCA GAC CCC TTT GCC GCC AAA GTG TTT GAT GTG TGT TTG ACC CTC CAT GCT GAG CAC ACC ATG AAT GCG TCC ACC TTT TCG GCC CGG GTA ACG GCT TCT ACT CTC ACG GAT CCC TAT GCA GTG GTT GCC TCG GCG GTG GGG ACT TTG GCG GGG CCG CTC CAC GGG GGA GCC AAC GAA GAA GTG CTA AAT ATG CTT GAA GAA ATT GGC TCA GTG GAA AAT GTC CGC CCC TAC GTG GAA AAA TGC CTG GCC AAC AAA CAG CGC ATC ATG GGC TTT GGC CAC CGA GTT TAT AAA GTC AAA GAC CCC CGG GCA ATT ATT TTG CAG GAT TTG GCT GAA CAG TTA TTT GCC AAA ATG GGC CAC GAC GAA TAT TAC GAA ATC GCA GTG GAG TTG GAA AAA GTA GTG GAA GAA TAC GTG GGT CAA AAG GGC ATT TAC CCC AAT GTG GAC TTC TAT TCC GGT TTG GTT TAC CGC AAG CTA GAC ATC CCC GCC GAT CTG TTT ACG CCC CTA TTT GCG ATC GCC AGG GTG GCG GGT TGG TTG GCC CAC TGG AAG GAA CAA TTA TCA GTC AAT AAA ATT TAC CGT CCT ACC CAA ATT TAC ATC GGT GAC CAT AAT TTA TCT TAT GTT CCC ATG ACA GAA CGG GTA GTT TCC GTG GCC CGC AAT GAA GAC CCC AAT GCG ATT ATT

1