Electronic Supplementary Material (ESI) for Metallomics.
This journal is © The Royal Society of Chemistry 2016

**Supporting information**

**Copper-responsive gene expression in the methanotroph *Methylosinus trichosporium* OB3b**

Grace E. Kenney[a], Monica Sadek[b], and Amy C. Rosenzweig[a,b]

[a]Departments of Molecular Biosciences and [b]Chemistry, Northwestern University, Evanston, IL 60208

**Supplemental Materials and Methods**

**RNA isolation and cDNA synthesis.** All experiments involving RNA were carried out using RNase-free reagents and supplies in a dedicated work area. All RNA experiments included three distinct biological replicates. qRT-PCR experiments, data analysis, and data reporting were designed in an effort to comply with the MIQE guidelines.[1] At each timepoint, culture volumes containing approximately 3E9 cells were added to a stop solution of ice cold 5% phenol/95% EtOH at a v/v ratio of 10:1.25 and centrifuged at 6000 x g for 20 min at 4 ºC. The resulting pellets were resuspended in 0.6 mL TE containing 20 mg lysozyme (Sigma) and 100 µL proteinase K (Qiagen) and incubated for 30 m at room temperature with shaking. Then 5 mL 65 ºC Qiazol was added, and RNA purification continued using an RNeasy Midi Kit (Qiagen), including an on-column DNase treatment, following the protocol described in the RNEasy Lipid Handbook (Qiagen). The RNA was eluted in RNase-free water and quantified using a Nanodrop (Thermo). Residual gDNA was removed using a Turbo DNAfree kit (Ambion) as described by the manufacturer, and the remaining RNA was cleaned up using an isopropanol/ethanol precipitation and resuspended in 20 µL RNase-free $H_2O$ (Life Technologies.) Final RNA concentrations were determined using a Nanodrop 1000 (Thermo) and confirmed along with RNA integrity with a BioAnalyzer (Agilent) (Table S4). cDNA was synthesized using 1 µg starting RNA and a SuperScript VILO cDNA synthesis kit (Invitrogen) using random hexamer primers, following the manufacturer's instructions.

**Reference gene selection for qRT-PCR**. Historically, qRT-PCR studies have often relied upon several assumptions: that RNA quality does not significantly affect results, that primers are

equally efficient, and that so-called housekeeping genes used as reference genes in one study are stable across other conditions. None of these assumptions are now accepted.[2-5] The RNA quality of all samples was analyzed and almost all samples used for analysis had an RNA Integrity Number (RIN) >7.5, indicating high quality RNA (Table S4). Primer efficiency values were calculated for each primer set (Table S3), and the presence of multiple products or primer-dimers was ruled out using both melting curves and analysis on a 4% NuSieve GTC agarose gel (Fig. S5).

The handful of previous RT-PCR studies performed on *M. trichosporium* OB3b used 16s rRNA as a reference gene.[6,7] However, 16s rRNA has very high transcription levels[8] and the dilution levels required to achieve $C_q$ values above 10 may push uncommon transcripts of other genes below the limit of detection. Additionally, 16s rRNA is exceptionally stable, and may not reflect integrity or quantity of total RNA particularly well.[9] Finally, use of multiple reference genes is now the preferred option for qRT-PCR studies. For this analysis, 14 potential reference genes were chosen from 10 major biological systems (Table S2.)[10,11] stability and optimal number of reference genes were chosen using GeNorm$^{PLUS}$, and the most stable set of genes was used for data analysis (Fig. S6).[5]

**Extended TEM sample preparation**. A Pelco BioWave Microwave with a cold spot and vacuum chamber was used for processing steps. Glutaraldehyde-fixed bacteria were enrobed in 2% low gelling temperature Agarose II (aMReSCO) and further processed in a secondary fixative of 2% $OsO_4$ in buffer, followed by two DI water washes, and an acetone dehydration series (30, 50, 70, 90 and 100% 2x). The EMBed 812 resin was infiltrated with acetone. Acetone to resin ratios 2:1 and 1:1 were processed in the microwave for 3 min and left on the bench for

30 min. to 2 hours. The 1:2 mix was processed overnight at 4ºC. Two 100% resin steps were infiltrated at RT for 1 – 4 hours each. A change of fresh resin was used to polymerize samples in 00 beem capsules in a 60ºC oven for 24 hours. The ultra thin sections (~90nm) were created using a Leica Ultracut S ultramicrotome and collected on copper mesh grids. A post stain was applied by floating the grid on a drop of 3% uranyl acetate for 7-10 min. The grid was rinsed in $CO_2$ free DI water and then floated on a drop of lead citrate for 7 min. After rinsing in $CO_2$ free DI water, the stained sections were allowed to dry before acquiring images using a Gatan Orius camera on a JEOL 1230 TEM (80kV accelerating voltage).

**Characterization of PP358 Mbn**. After cellular harvest, spent medium was optionally incubated with 1mM (final) $CuSO_4$ and filtered over a .2μ PTFE filter (Millipore) protected by a GF/B glass microfiber filter (Whatman) in order to remove particulate matter. The clarified spent medium was loaded onto a Diaion HP-20 column, which was washed in 10mM NaOAc or $NH_4OAc$. Mbn was eluted with 60% MeOH/40% 10mM NaOAc or $NH_4OAc$ as previously described.[12] Eluate was pooled and lyophilized; the resulting Mbn was resuspended in MilliQ water. Mbn samples were analyzed via mass spectrometry on an AmaZonX ion trap mass spectrometer (Bruker) using electrospray ionization. Mass spectra were acquired in both positive and negative ion modes over a mass range from 100-2000 Da.

**Cloning, isolation, purification, and preliminary analysis of MbnI and MmoD**. A codon-optimized *Ms. trichosporium* OB3b MmoD sequence was cloned into pPR-IBA1, providing a construct with a C-terminal Strep-tag (Genscript, IBA) (Supplemental Textfile 1). A similar construct was also assembled using the native *Ms. trichosporium* OB3b MbnI sequence, cloned

into the vector using BsaI (NEB) (sequence in Supplemental Textfile 2). Both constructs were expressed in C41(DE3) cells grown in LB medium with or without copper supplementation to 0.5 mM, induced in early logarithmic phase growth with 0.5 mM IPTG, and grown overnight at 18 ºC. Cells were lysed via sonication (1s on/3s off at 4 ºC for 10 m total) in lysis buffer (25 mM MOPS pH 7.2, 250mM NaCl, 10% glycerol, Roche cOmplete protease inhibitors), and after clarification at 21000 x g and 4 ºC for 1h, the lysate was loaded onto a 15 mL Strep-Tactin column (IBA) equilibrated in 25mM MOPS pH 7.2, 250mM NaCl, and 10% glycerol. The column was washed with loading buffer, the protein was eluted in loading buffer with 2.5 mM desthiobiotin, and the column was regenerated in 1 mM HABA. Proteins were concentrated and desalted in Amicon spin concentrators with a 3 kDa cutoff and quantified using the Bradford assay (Sigma). Desalted protein was loaded on a heparin column and eluted with a gradient of 0-100% 2 M NaCl. Metal content was analyzed in 5% nitric acid using a iCap Q inductively coupled mass spectrometer (Thermo) with a standard curve from 0-100 ppb Cu using a multi-element standard and 5 ppb indium, lithium, scandium, and yttrium internal standards (both Inorganic Ventures).

**Assembly of the PP358 genome**. *De novo* assembly of the *Ms. trichosporium* OB3b PP358 genome was performed within the software package CLC Genomics Workbench v 8.0 (CLCbio, Cambridge, MA). Raw reads were quality trimmed (Q20) and mapped against a reference genome of *Ms. trichosporium* OB3b (i.e. ADVE02000001-ADVE02000003) using low stringency (0.5 length fraction and 0.8 similarity fraction). All three genomic elements were covered deeply, at an average depth of 163-186X. Reads mapping to the reference genome were recovered and used for *de novo* assembly using default settings within CLC Genomics. Short

contigs (<200 bp) and contigs with low coverage (<50X) were removed from the assembly, and the mapping reads recovered again. A subsequent *de novo* assembly was performed on the remaining contigs.

**CopD bioinformatics**. All ORFs containing the CopD PFAM (PF05425) were obtained from the JGI-IMG database. The predicted number of helices was calculated using TMHMM.[13] For genes of interest, including CopD from WT and mutant *Ms. trichosporium* OB3b, transmembrane homology was confirmed via TopCons,[14] and the JGI-IMG annotation of additional domains was supplemented by analysis via HHPred[15] (Supplemental textfile 1). Subgroups of sequences were aligned via MAFFT[16] in order to further analyze gene architecture and sequence conservation. An extended profile hidden Markov model covering the full 8-helix region was constructed via HMMER 3.1[17] using a MAFFT (FFT-NS-2)[18] alignment of all *copD* sequences predicted to have 8 helices, with additional N- or C-terminal domains manually trimmed. All sequences were aligned against the new profile HMM via HMMalign[17] in order to identify very highly conserved residues.

**Supplemental Results and Discussion**

**Analysis of PP358 variants.** Many variants observed in the PP358 genome are unlikely to be responsible for the phenotype of that strain. Silent mutants, mutations in transposases and other gene mobility elements are poor candidates (File S1). Many other genes have clear roles that are unlikely to be directly related to copper homeostasis or methane metabolism, such as components of polyketide synthetase systems or a nicotinic acid phosphoribosyltransferase

(Table S1). Some potentially relevant proteins remain (Table S1), including several hypothetical proteins. These cannot necessarily be ruled out; MettrDRAFT_0214 and MettrDRAFT_2026 are hypothetical proteins with no useful genomic context. Genomic neighborhood or the presence of certain domains does suggest that other hypothetical proteins are less likely to be involved. MettrDRAFT_2787 appears to be related to PopZ (pole-organizing protein Z), which is involved in cell division, MettrDRAFT_2979 is located near RNA-related proteins (including pseudouridine synthase and RNA methyltransferase), and MettrDRAFT_3481 is likely to be an alpha/beta hydrolase.

Two histidine kinases from two-component signaling systems are also present (Table S1), but in these systems, substrate sensing depends on the periplasmic component of the system. Disruption of periplasmic copper sensing seems unlikely to result in one of the key PP358 phenotypes, namely low cellular copper content: while regulation may be disrupted, transport should not be. A mutation in a gene involved in methanol dehydrogenase (Table S1) assembly may broadly be associated with methane metabolism, but *methanol* metabolism is not copper-dependent; methanol dehydrogenase metabolizes methanol produced by both MMOs. The mutation in a *Surf1* homologue (Table S1) is intriguing since *Surf1* is involved in copper loading of periplasmic enzymes but the gene is located in a cytochrome *c* operon, suggesting that it is specifically involved in cytochrome *c* copper loading. A mutation in *mbnB*, a gene predicted to be involved in Mbn biosynthesis, initially drew our attention, but it merely eliminates an internal stop codon present in the reference genome but found in no other MbnB homologues and not observed in wildtype *Ms. trichosporium* OB3b as cultured in our laboratory. CopD is by far the strongest candidate because it is directly involved in copper homeostasis, directly adjacent to pMMO, and appears to be co-regulated with pMMO by copper addition.

**CopD bioinformatics.** Identification of the various predicted protein architectures present in the CopD family suggests that the existing profile hidden Markov model (HMM), which covers only four helices of the standard 8-helix CopD sequence, is inadequate to describe the larger family. Using sequences predicted to have 8 transmembrane helices and no additional domains, a new profile HMM was constructed via HMMbuild[17] (Supplemental File S5). All sequences were then aligned to this model using HMMalign.[17] Using the resulting alignment, conserved residues were identified. Two highly conserved histidine residues that are present in predicted periplasmic loops in full-length CopD sequences may represent copper-binding ligands (Fig. S4).

**Supplemental Files**

File S1. Variants table for *Ms. trichosporium* OB3b PP358 when compared to the wildtype genome.

File S2. Raw CNRQ data for qRT-PCR reactions.

File S3. CopD traits and calculations (20150916_jgi_copD.xlsx). Data from the JGI/IMG database augmented by predicted numbers of transmembrane helices and other calculations.

File S4. CopD sequences. (20140917_jgi_allcopD.fa). All CopD sequences in FASTA format.

File S5. 8-helix CopD HMM. (20150916_8-helix-copD.hmm).

## Supplemental Figures

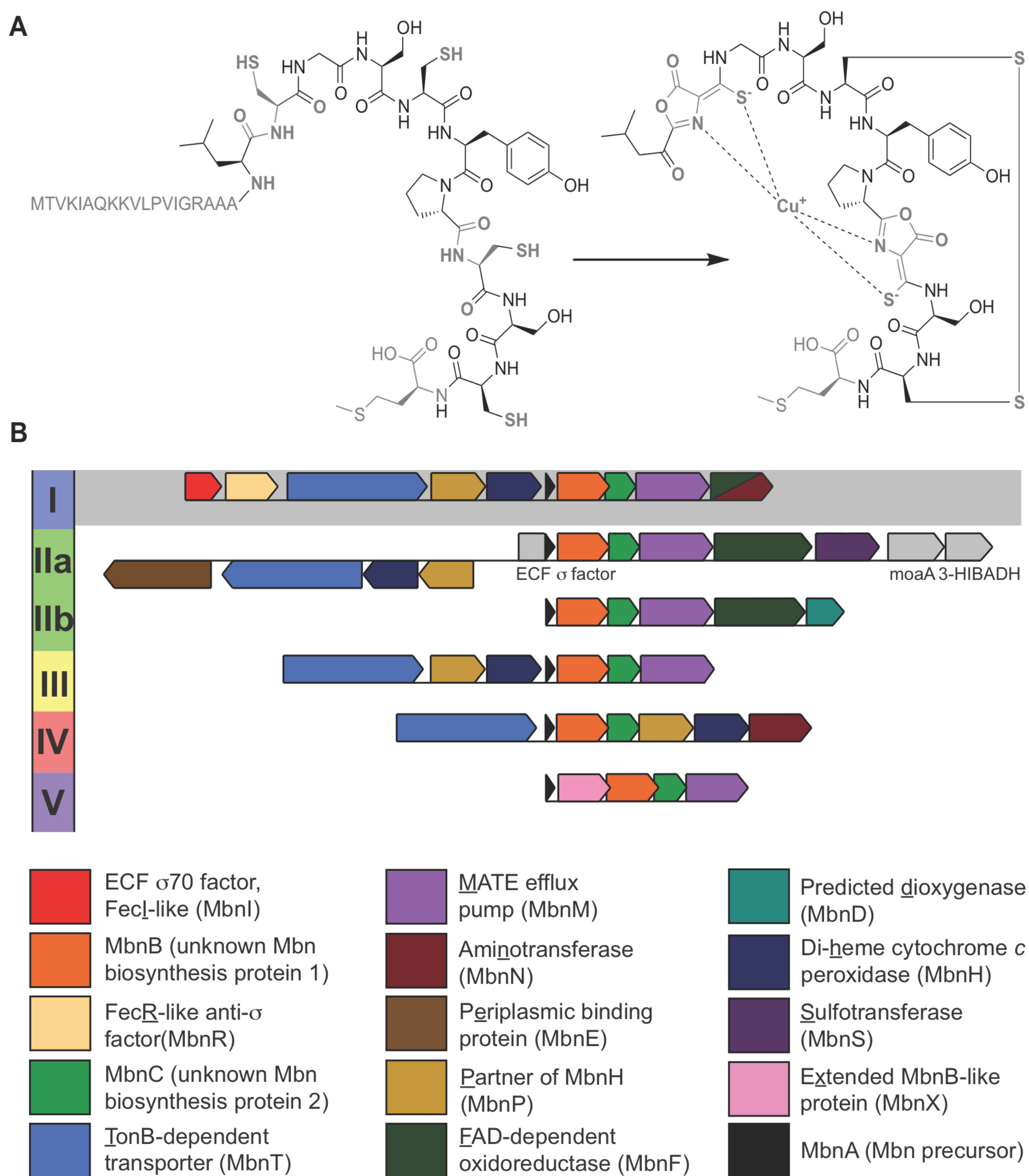



**Fig. S1.** Methanobactin structure and genetic context. (*A*) The ribosomally produced precursor peptide MbnA in *Ms. trichosporium* OB3b is proposed to be post-translationally modified to form Mbn (shown here as CuMbn). (*B*) The 35 full or partial Mbn operons identified in bacterial genomes can be divided into several groups using sequence-based phylogenetics and operon content. The *Ms. trichosporium* OB3b Mbn operon belongs to Group I.

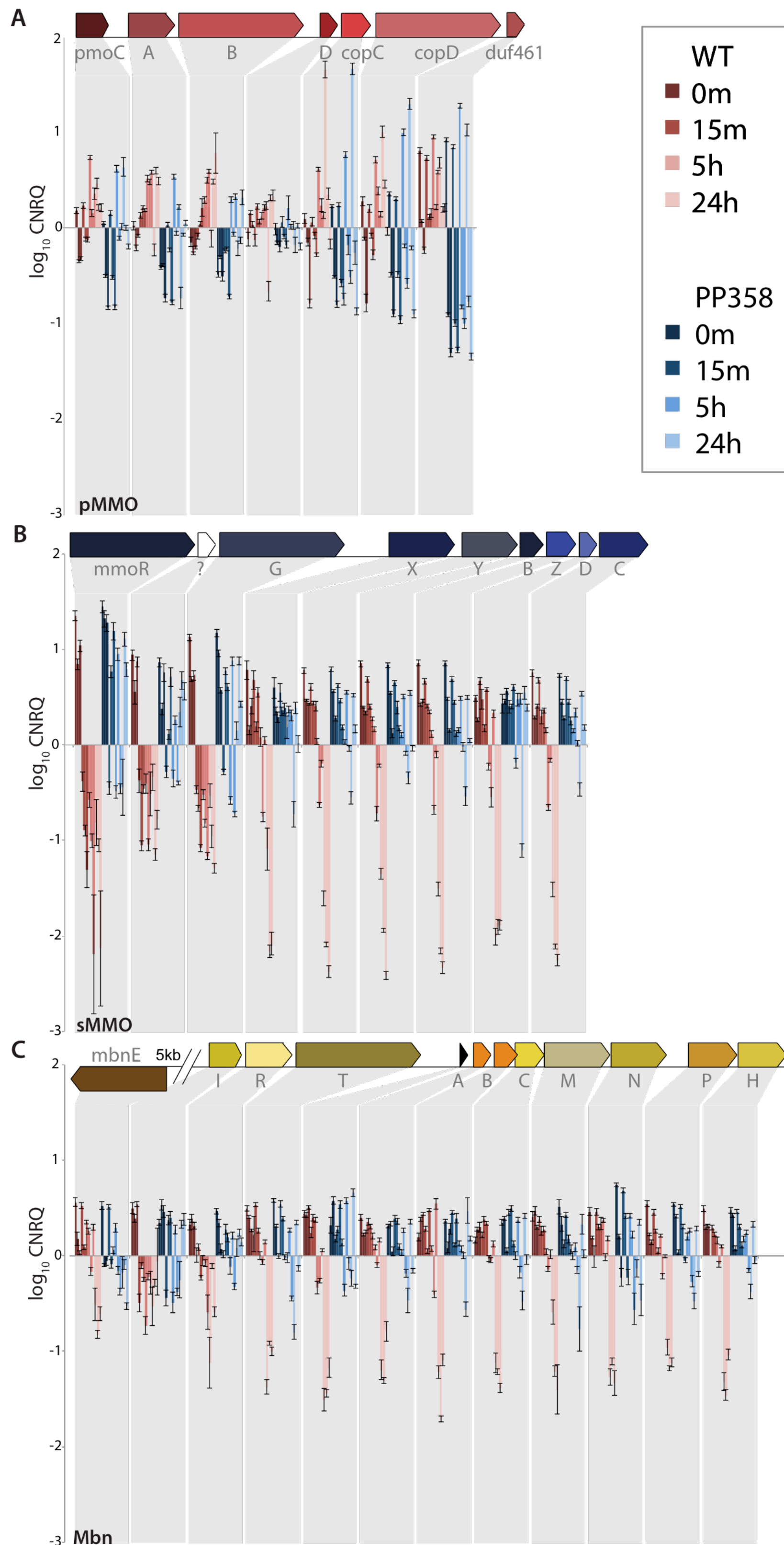



**Fig. S2.** Individual $\log_{10}$-transformed CNRQ values for each replicate run of each gene in the (*A*) pMMO, (*B*) sMMO, and (*C*) Mbn operons. Error bars represent standard error.

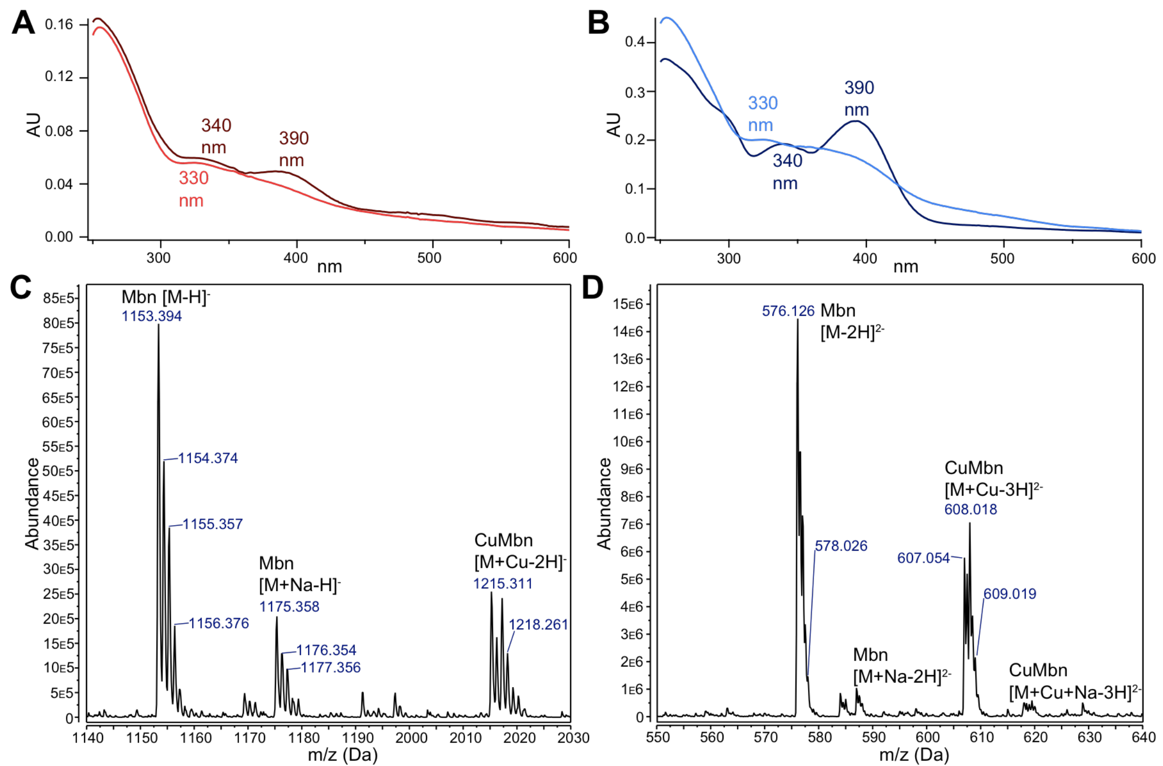



**Fig. S3.** Apo and holo Mbns from wildtype *Ms. trichosporium* OB3b and the PP358 strain. UV-visible light spectra for spent medium from the wildtype (*A*) (dark red: 0 m and 0µM $CuSO_4$, light red, 15 m and 12.5µM $CuSO_4$) and mutant (*B*) (dark blue, 0m and 0µM $CuSO_4$, light blue; 15m and 12.5µM $CuSO_4$). Prior to copper addition, two peaks at approximately 340 and 390nm are observed, corresponding to unchelated oxazolones; after copper addition, both features (especially the 390nm feature) diminish and a 330nm feature becomes more visible. Differing amounts of total Mbn, oxazolone hydrolysis, C-terminal methionine loss and presence of non-Mbn molecules account for the differences in the spectra. Mbn isolated from PP358 on a Diaion HP-20 column exhibits the same apo and holo masses as Mbn from the wildtype species when measured on an Amazon X ESI ion trap mass spectrometer in negative mode: (*C*) singly charged species and (*D*) doubly charged species. In the holo form, the characteristic copper-derived isotope splitting is clearly visible.

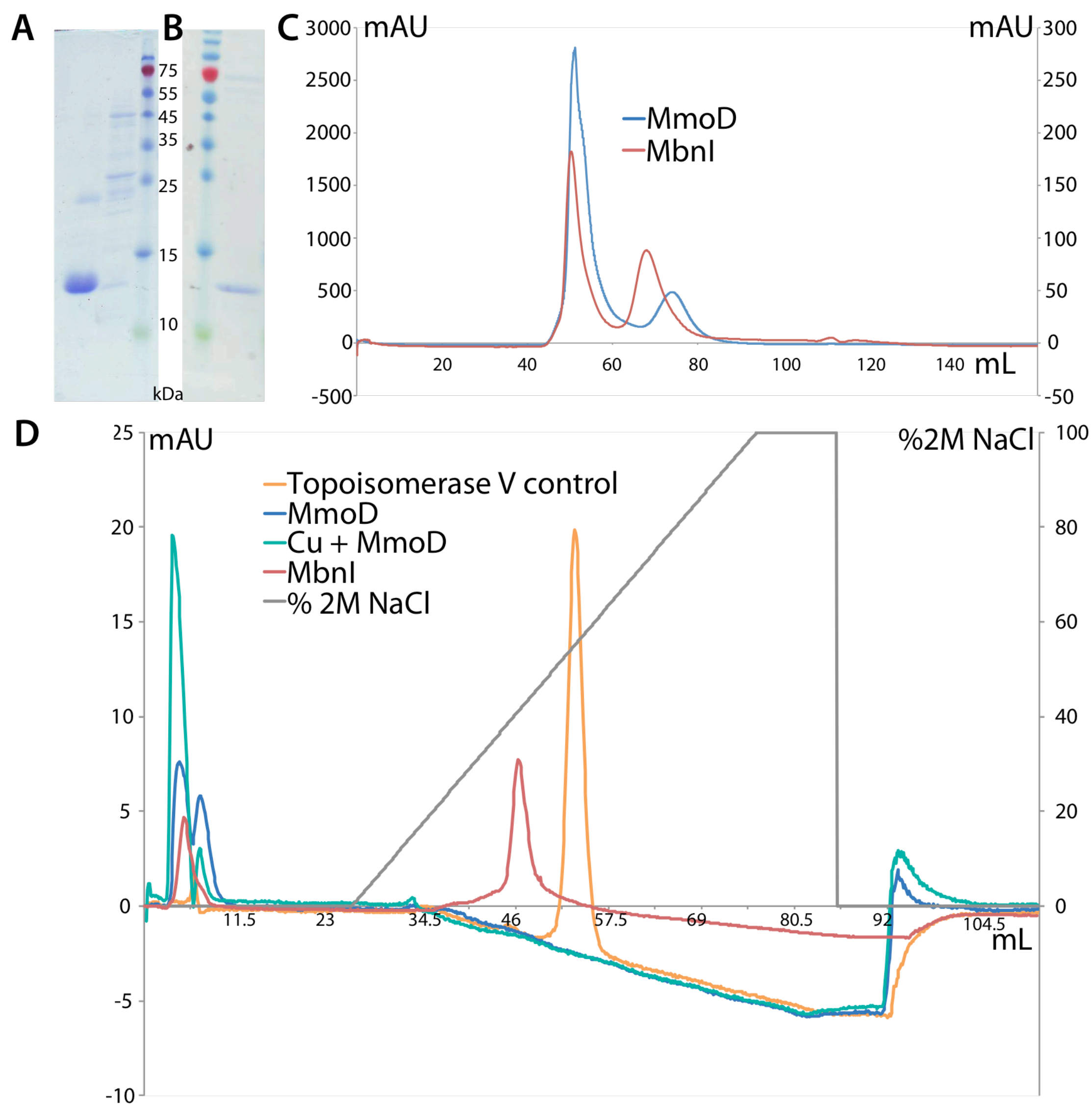



**Fig. S4.** (*A*) MbnI SDS-PAGE. Degradation to form a C-terminal fragment of ~13 kDa is observed over time. (*B*) MmoD SDS-PAGE. (*C*) Size exclusion chromatography of MmoD and MbnI. After initial purification, primarily monomeric MmoD is observed. (*D*) Chromatographic analysis of MmoD, MbnI, and a control protein loaded onto a heparin column and eluted over a gradient of 0-100% B (2 M NaCl). MmoD does not bind to a heparin column with or without copper. MbnI does bind a heparin column and elutes at approximately 0.8 M NaCl.

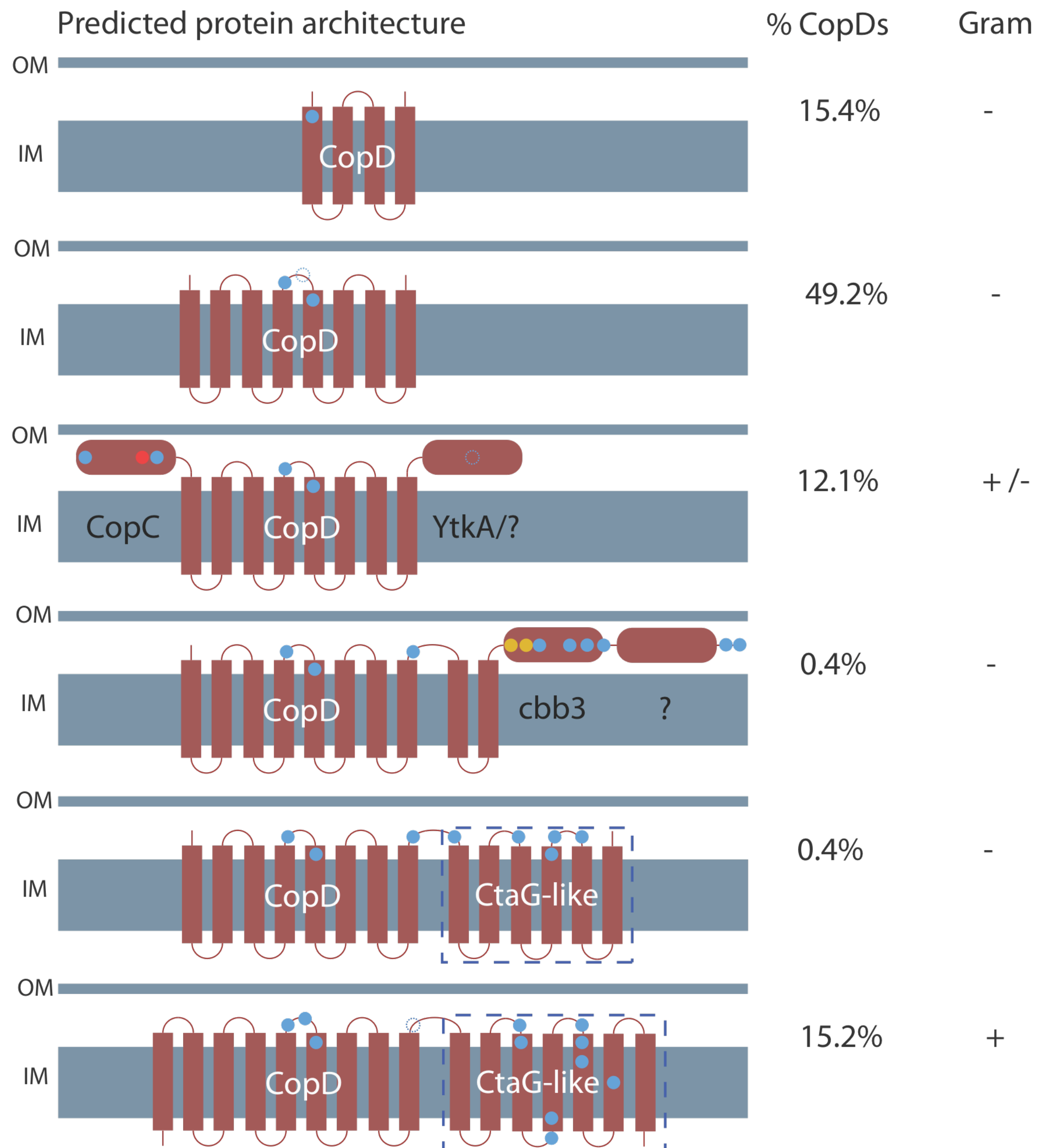



**Fig. S5.** Common predicted CopD architectures. Almost half of CopD sequences are predicted to consist solely of an 8-transmembrane-helix protein. A significant minority are truncated (containing only 4 transmembrane helices), contain soluble domains (including CopC fusions, YtkA-like domains, and cytochrome *cbb3*-like domains), or contain membrane protein fusions (primarily CtaG-like C-terminal domains.) Conserved histidine, cysteine, and aspartic acid residues are indicated by blue, yellow, and red circles.

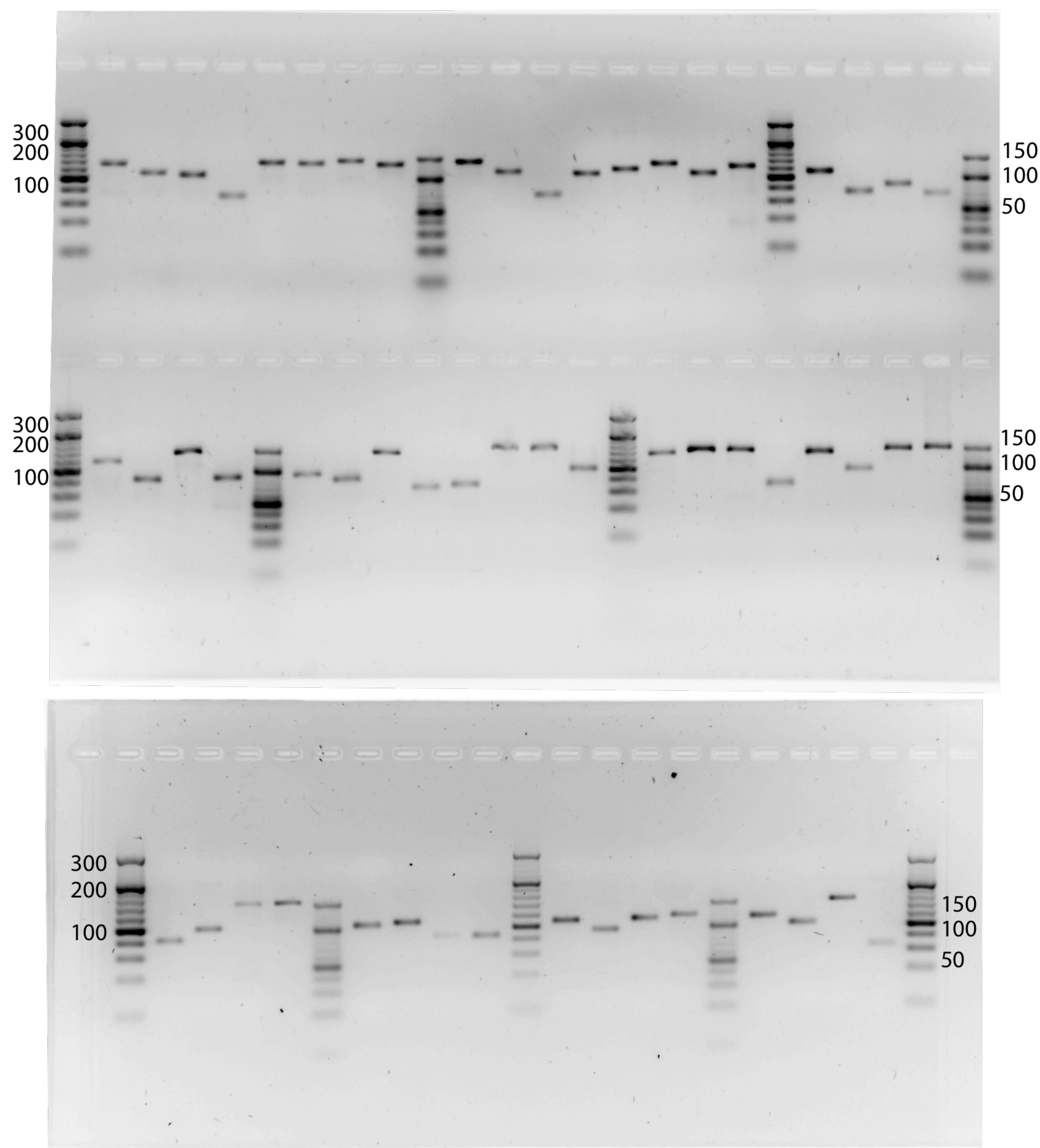
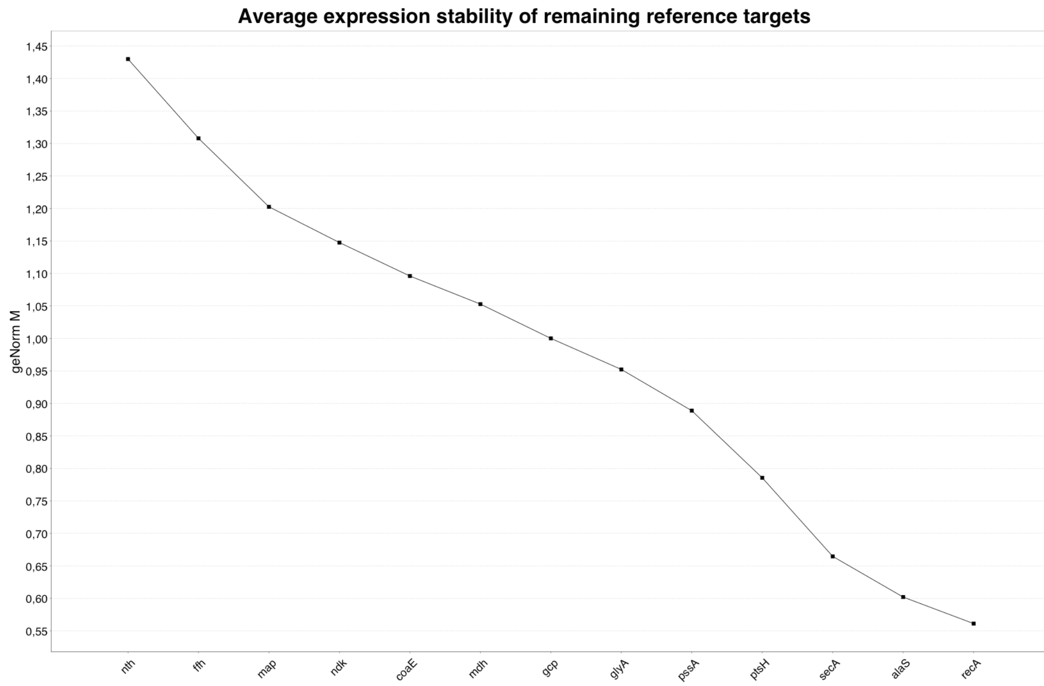



**Fig. S6.** 4% NuSieve 3:1 agarose TBE gel of qRT-PCR products. *Top*: 20bp O'RangeRuler ladder, coaE, ffh, map, mmoX, alaS, secA, glyA, gmk, 10bp O'RangeRuler ladder, ptsH, mdh, mmoX, recA, pssA, ndk, gcp, rpe, 20bp ladder, 16s rRNA, pmoB, pmoC, mmoX, 10bp ladder. *Middle*. 20bp ladder, mmoD, mbnB D1, ferritin, 10bp ladder, pmoC v2, mmoG, mbnH, mbnC, mmoX, mbnA, mbnM, mbnT, 20bp ladder, pmoA, mbnB D2, mmoY, mmoC, mmoX, mmoZ, mbnN, mmoR, mmoB, 10bp ladder. *Bottom*: 20bp ladder, mbnP, pmoD, nth, MettrDRAFT_0733, 10bp ladder,mbnI, map, tpiA, mbnR, 20bp ladder, duf461, copC, mbnE, sco1, 10p ladder, pmoAG, copD, mmoXG, 16sG, 20bp ladder.



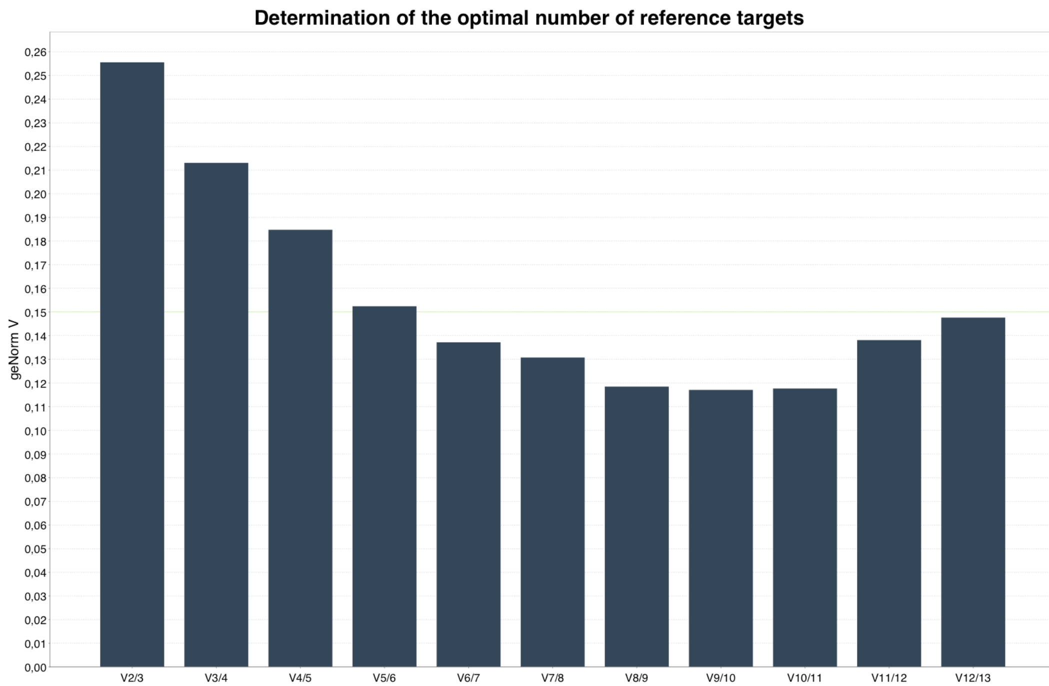




**Fig. S7**. Identification of optimal reference genes. (*A*) GeNorm M value graph, indicating expression stability of reference gene candidates. (*B*) GeNorm V value graph, indicating the optimal number of reference genes.

## Supplemental Tables

**Table S1.** Variants observed in the PP358 genome of interest due to significant mutations combined with unassigned functions or functions likely to be relevant to methane or copper metabolism.

| Gene | Locus Tag | Notes |
|---|---|---|
| N/A | MettrDRAFT_0214 | Hypothetical protein |
| *copD* | MettrDRAFT_0379 | Putative periplasmic copper importer |
| N/A | MettrDRAFT_0988 | Histidine kinase (2-component signaling) |
| *mxaL* | MettrDRAFT_1969 | Part of methanol dehydrogenase assembly |
| N/A | MettrDRAFT_2026 | Hypothetical protein |
| *Surf1* | MettrDRAFT_2859 | Cytochrome c oxidase copper loading |
| N/A | MettrDRAFT_2787 | Hypothetical protein |
| N/A | MettrDRAFT_2979 | Hypothetical protein |
| *mbnB* | MettrDRAFT_3422 | Mbn biosynthesis |
| N/A | MettrDRAFT_3481 | Hypothetical protein |
| N/A | MettrDRAFT_4334 | Histidine kinase (2-component signaling) |

**Table S2**. Final qRT-PCR primers and product sizes for genes analyzed in these studies. Bolded genes are reference gene candidates. Additional reference gene candidates that failed screening at various stages include *gmk*, *rimM*, *rnpA*, *rpe*, *rplI*, *rpoD*, and *rpoN* (primer sequences not shown.)

| Locus tag | Gene | Product | Forward | Reverse |
|---|---|---|---|---|
| r0051 | ***16s*** | 116 | TAGGCGGATTGTTAAGTCAGG | ATTTCACCTCTACACTCGCAG |
| r0051 | ***16s (G)*** | 66 | GCAGAACCTTACCAGCTTTTGAC | CCCTTGCGGGAAGGAAGTC |
| MettrDRAFT_2641 | ***alaS*** | 139 | TGATCCGACCTTGATGTTCAC | GTCGAGATCATTGTGCTTGC |
| MettrDRAFT_2299 | ***coaE*** | 136 | TCGTGCTGTTCGATATTCCG | AATTTCTCCTCCGTCATGCC |
| MettrDRAFT_0380 | *copC* | 93 | GTCGGCGAAAGACCATGT | AGGATCGAGATCGAGGAATAGG |
| MettrDRAFT_0379 | *copD* | 102 | CTGACTTATGGTGGGCAGAG | GTTCGACAGAACGGTGAGATAG |
| MettrDRAFT_0378 | *DUF461* | 109 | GCATGAATACGATCTCGGGAAG | ATCGCCCTTGTTATGGATGAAG |
| MettrDRAFT_0806 | ***ffh*** | 115 | GGAAGATGAAGGAGCAGATCG | TTGAGGATATCGGGATTGCG |
| MettrDRAFT_2542 | ***gcp*** | 111 | CGGTTTGCGTTTCGTCATG | AAATCGAGGTCGTTGGTGAG |
| MettrDRAFT_3587 | ***glyA*** | 146 | TGCTCACCAATGACGAAGAG | TTGCTGATAGGCCTTGAACTC |
| MettrDRAFT_3782 | ***map*** | 110 | CCAATATTCTGCATTACGGCG | GAGAATCTTCACCTGCGGAC |
| N/A | *mbnA* | 75 | TGACTGTCAAGATTGCTCAGAAG | GATAGCACGAACCGCAAAG |
| MettrDRAFT_3422 | *mbnB D1* | 87 | GAAAGAGATCGGGAGGAATTGG | ATGATCAGAGACATAGACGGGC |
| MettrDRAFT_3422 | *mbnB D2* | 152 | TTCCACGTTGCAGGATATGG | TCCCGTTCATAAGTGATCGTCG |
| MettrDRAFT_3423 | *mbnC* | 148 | CTACTGGCACACATTGTTCGAC | CCAGACATAGACCCAGAGGAAA |
| MettrDRAFT_3413 | *mbnE* | 113 | GAGAGCCTCAATCGCTTCAA | CCATAATAGGTCTTCCGCTCATC |
| MettrDRAFT_3427 | *mbnH* | 92 | CGCCTGGAATTGGGACTTAC | TGGAACTTCTCCTCCGACAT |
| MettrDRAFT_3419 | *mbnI* | 105 | GCGTTCATCCTCTACGTTTCT | CTGAAGAGCGAGACGGATATG |
| MettrDRAFT_3424 | *mbnM* | 161 | GCTTCCTTCGTCCGATCTATTG | GAACAGCATCAGCATCAAGGTC |
| MettrDRAFT_3425 | *mbnN* | 97 | CCATCCTTTCCGATGTATGCTC | CCACTTTCGAAGACAAGGAGAG |
| MettrDRAFT_3426 | *mbnP* | 83 | CTCTACGTCTATGGCTTCGAG | GCGTATTGCCAATCGTTCTG |
| MettrDRAFT_3420 | *mbnR* | 86 | CGTGTTCGATGCGGATGAT | GGTATTTGGTCAGAACCGTCAT |
| MettrDRAFT_3421 | *mbnT* | 160 | GATCGGTTACGATTGGACCTTC | GTAAGTCGGTGTCATGAAGAGG |
| MettrDRAFT_1083 | ***mdh*** | 131 | TCGTCATCTGCATCACCAAC | GGCGAGGAAATAACGGAAAC |
| MettrDRAFT_2364 | *mmoB* | 149 | GGTGGATCAAGGCCGATG | GGTGATGGTGAATTTGGTGC |
| MettrDRAFT_2367 | *mmoC* | 148 | CCTATGATCGCATCTCCTTCG | GGAACGAAATTGAGCGAGATG |
| MettrDRAFT_2366 | *mmoD* | 124 | TGTTGCGATGGGAGATTCTG | GACGTAGCATCCTGTCGG |
| MettrDRAFT_2361 | *mmoG* | 92 | GCCTCTTGCTCAATCATATCCG | ATCGTGCAGGAACTCATAGC |
| MettrDRAFT_2359 | *mmoR* | 146 | GCAAGGAGCATTTCAACCG | GATCTTGAGACAGCCGAGC |
| MettrDRAFT_2362 | *mmoX (G)* | 153 | TCAACACCGATCTGAACAACG | TCCAGATTCCACCCCAATCC |
| MettrDRAFT_2362 | *mmoX* | 73 | TTCAAAGAGAACCGGACGAAG | TGACCTTGAACTGCTCCTTG |
| MettrDRAFT_2363 | *mmoY* | 147 | GTGAAGGACAAGTCGGAAGAG | GCCGTATTCGCTGTAGAGTAG |
| MettrDRAFT_2365 | *mmoZ* | 124 | AGAACCCATCCACGACAATTC | GATATCGTAGCTCTTGCGGAAG |
| MettrDRAFT_0355 | ***ndk*** | 134 | CTGATCGAGAGCATCACCTC | AATTCCTTGCGGATCGTACC |
| MettrDRAFT_3054 | ***nth*** | 149 | AACATCTTCACTCTGCTCGTC | TTGATATAGTCCTTCAGCCGC |
| MettrDRAFT_0383 | *pmoA (G)* | 94 | TTCTGGGGCTGGACCTATTTC | CCGACAGCAGCAGGATGATG |
| MettrDRAFT_0383 | *pmoA* | 135 | GGAATATATCCGCATGGTCGAG | GAATACCACTTACCGACGAACC |
| MettrDRAFT_0382 | *pmoB* | 103 | GGAATATATCCGCATGGTCGAG | GTAGACCATCATCGACACGAAG |
| MettrDRAFT_0384 | *pmoC* | 131 | GACCCGCATTCCGTATTTTG | ACAGCTCTTCCATGAACCAG |
| MettrDRAFT_1032 | *pmoC2* | 88 | CGTCGATATGAAGCCCTTATGG | CCAGCCATAGATCTGCTCATAG |
| MettrDRAFT_0381 | *pmoD* | 101 | CGCTCACCATCATCCATCTT | CGTGACAAAGCCGATGAAATG |
| MettrDRAFT_4197 | ***pssA*** | 121 | GTCGCCTTCTACATCGTCTATG | CGAACAGGACCAGGATCAC |

| | | | | |
|---|---|---|---|---|
| MettrDRAFT_2387 | ***ptsH*** | 142 | ATCTCGTGATCTGCAACCG | GTCAGTATGCCCATGATCGAG |
| MettrDRAFT_2640 | ***recA*** | 111 | GTTGAAATTCTACGCTTCGGTG | ACCTTGTTCTTGACCACTTTGA |
| MettrDRAFT_2356 | *sco1* | 119 | TTCGGCTTCACGCAATGT | GGATCGAGGCTGACGAATAGA |
| MettrDRAFT_1638 | ***secA*** | 138 | CATGATCGAGAAGCAGGTGG | TCGGACTTGTATTCGTTGAGC |
| MettrDRAFT_0356 | ***tpiA*** | 86 | TCTCCGCTGAAATGCTGG | GTCGCTCTCGCCATGTTC |
| MettrDRAFT_2360 | *N/A* | 150 | CGGCACGCTCGATTTCG | TTGGATTTGTCATAGGCGGAG |

**Table S3.** RNA quantities and RNA Integrity Numbers (RINs) for samples used in qRT-PCR experiments.

| | WT concentrations (ng/μL) | | | WT RIN | | | PP358 concentrations (ng/μL) | | | PP358 RIN | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| 0m | 516 | 505 | 586 | 8.2 | 9.2 | 6.9 | 608 | 340 | 239 | 8.6 | 8.9 | 8.6 |
| 15m | 550 | 580 | 661 | 8.4 | 9.9 | 7.5 | 654 | 385 | 264 | 8.5 | 9.1 | 7.9 |
| 5h | 536 | 250 | 693 | 8.6 | 8.2 | 9.1 | 448 | 175 | 307 | 8.3 | 8.5 | 8.1 |
| 24 | 232 | 295 | 666 | 7.7 | 9.0 | 9.5 | 722 | 200 | 380 | 7.8 | 8.3 | 7.4 |

**Table S4.** Primer efficiency data, calculated in qBase$^{PLUS}$

| Target gene | E | E (SE) | R^2 | Slope | Slope error | Intercept | Intercept error |
|---|---|---|---|---|---|---|---|
| *16s* | 2.054 | 0.024 | 0.996 | -3.199 | 0.053 | 24.998 | 0.253 |
| *alaS* | 1.896 | 0.081 | 0.937 | -3.598 | 0.241 | 47.98 | 1.201 |
| *coaE* | 2.152 | 0.082 | 0.976 | -3.005 | 0.15 | 44.756 | 0.814 |
| *copC* | 1.925 | 0.059 | 0.976 | -3.515 | 0.165 | 43.176 | 0.808 |
| *copD* | 1.977 | 0.055 | 0.974 | -3.378 | 0.137 | 44.542 | 0.594 |
| *duf461* | 1.999 | 0.019 | 0.997 | -3.324 | 0.045 | 38.347 | 0.213 |
| *ffh* | 1.735 | 0.05 | 0.976 | -4.177 | 0.216 | 57.422 | 1.158 |
| *gcp* | 2 | 0.06 | 0.982 | -3.323 | 0.143 | 46.564 | 0.75 |
| *glyA* | 1.906 | 0.011 | 0.999 | -3.569 | 0.032 | 42.229 | 0.157 |
| *map* | 1.808 | 0.042 | 0.985 | -3.889 | 0.154 | 48.016 | 0.802 |
| *mbnA* | 1.901 | 0.04 | 0.991 | -3.584 | 0.116 | 36.016 | 0.503 |
| *mbnBD1* | 1.846 | 0.024 | 0.997 | -3.755 | 0.081 | 39.095 | 0.325 |
| *mbnBD2* | 1.939 | 0.011 | 0.999 | -3.479 | 0.031 | 38.107 | 0.146 |
| *mbnC* | 1.911 | 0.028 | 0.996 | -3.555 | 0.08 | 39.607 | 0.333 |
| *mbnE* | 1.867 | 0.021 | 0.996 | -3.687 | 0.066 | 45.144 | 0.343 |
| *mbnH* | 1.975 | 0.031 | 0.993 | -3.383 | 0.078 | 39.899 | 0.359 |
| *mbnI* | 1.801 | 0.054 | 0.968 | -3.914 | 0.198 | 49.379 | 1.001 |
| *mbnM* | 1.909 | 0.063 | 0.965 | -3.56 | 0.182 | 43.139 | 0.867 |
| *mbnN* | 2.056 | 0.032 | 0.994 | -3.195 | 0.07 | 39.104 | 0.333 |
| *mbnP* | 1.965 | 0.019 | 0.998 | -3.409 | 0.049 | 39.761 | 0.216 |
| *mbnR* | 2.079 | 0.071 | 0.975 | -3.147 | 0.147 | 42.549 | 0.761 |
| *mbnT* | 1.992 | 0.015 | 0.998 | -3.341 | 0.036 | 37.785 | 0.173 |
| *mdh* | 1.941 | 0.059 | 0.972 | -3.472 | 0.159 | 44.47 | 0.706 |
| *mmoB* | 1.969 | 0.006 | 1 | -3.398 | 0.016 | 33.97 | 0.074 |
| *mmoC* | 2 | 0.013 | 0.999 | -3.321 | 0.032 | 36.98 | 0.146 |
| *mmoD* | 1.821 | 0.015 | 0.997 | -3.842 | 0.054 | 42.454 | 0.238 |
| *mmoG* | 1.957 | 0.029 | 0.994 | -3.429 | 0.077 | 41.038 | 0.356 |
| *mmoR* | 2.121 | 0.074 | 0.969 | -3.063 | 0.141 | 43.861 | 0.669 |
| *mmoX* | 1.938 | 0.022 | 0.997 | -3.481 | 0.058 | 35.567 | 0.273 |
| *mmoXG* | 1.96 | 0.009 | 0.999 | -3.422 | 0.024 | 33.124 | 0.113 |
| *mmoY* | 1.955 | 0.007 | 1 | -3.433 | 0.018 | 35.079 | 0.085 |
| *mmoZ* | 1.96 | 0.005 | 1 | -3.422 | 0.013 | 34.697 | 0.057 |
| *ndk* | 2.03 | 0.035 | 0.993 | -3.252 | 0.08 | 41.642 | 0.369 |
| *nth* | 1.987 | 0.161 | 0.857 | -3.354 | 0.395 | 46.539 | 1.968 |
| *pmoA* | 1.961 | 0.015 | 0.998 | -3.42 | 0.038 | 38.127 | 0.175 |
| *pmoAG* | 2.011 | 0.014 | 0.998 | -3.296 | 0.033 | 35.609 | 0.156 |
| *pmoB* | 1.94 | 0.018 | 0.997 | -3.475 | 0.048 | 39.441 | 0.238 |
| *pmoC* | 2.018 | 0.005 | 1 | -3.28 | 0.012 | 34.452 | 0.058 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *pmoCr* | 1.896 | 0.047 | 0.988 | -3.599 | 0.141 | 45.91 | 0.733 |
| *pmoD* | 1.938 | 0.04 | 0.981 | -3.479 | 0.11 | 42.152 | 0.51 |
| *pssA* | 2.041 | 0.166 | 0.906 | -3.227 | 0.368 | 44.015 | 1.775 |
| *ptsH* | 1.786 | 0.06 | 0.951 | -3.97 | 0.232 | 47.272 | 1.067 |
| *recA* | 2.146 | 0.124 | 0.935 | -3.016 | 0.229 | 43.673 | 1.083 |
| *sco1* | 1.986 | 0.089 | 0.963 | -3.355 | 0.219 | 39.948 | 0.931 |
| *secA* | 1.857 | 0.059 | 0.967 | -3.719 | 0.192 | 45.946 | 0.899 |
| *tpiA* | 1.793 | 0.377 | 0.524 | -3.943 | 1.419 | 55.513 | 8.12 |
| *MettrDRAFT_2360* | 1.673 | 0.061 | 0.934 | -4.472 | 0.317 | 47.238 | 1.536 |

**Table S5.** Cloning and sequencing primers.

| Primer name | Forward | Reverse |
| --- | --- | --- |
| OB3b_MbnIR_nest | TTTTTGCGCCTTCCAAAGAGC | GAGATCGAGGGAGAGGAGGTC |
| OB3b_MbnI_pPR-IBA1 | ATCGTAGGTCTCTAATGCCGG ACCAAAAATCGAATTTC | ATCGTAGGTCTCAGCGCTTTC CAGCGCCATCCGAC |
| OB3b_CopD_out_seq | CGCATTTCTCGACGGGAGGCC | CGACTGGCAAGCTCGTCGTCG |
| OB3b_CopD_in_seq | GCCTGGTATGCGAAGATCGGCG | TCGAAACCGCGCTTGACCGAC |

# SUPPLEMENTAL TEXT SEQUENCES

## **Text 1.** pPR-IBA1 MbnI and MmoD sequences

```
LOCUS       pPR-IBA1-MmoD-genscript.xdna    3109 bp      DNA     circular   15.09.08
DEFINITION
ACCESSION
VERSION
SOURCE
  ORGANISM
COMMENT     Serial Cloner Genbank Format
COMMENT     SerialCloner_Type=DNA
COMMENT     SerialCloner_Comments=Ligation of : pPR-IBA1.xdna [2782  nt] : (#BsaI[172] / #BsaI
            [104]) to Sequence Window #3 [319  nt] : (#BsaI[7] / #BsaI[3
            42])
COMMENT     SerialCloner_Ends=0,0,,0,
FEATURES             Location/Qualifiers
     primer_bind     20..39
                     /label=T7F
                     /SerialCloner_Color=&h0C7900
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     CDS             427..456
                     /label=Strep tag
                     /SerialCloner_Color=&hC07E00
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    507..526
                     /label=T7R
                     /SerialCloner_Color=&hC00400
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     rep_origin      668..1106
                     /label=f1 origin
                     /SerialCloner_Color=&h85009C
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=False
     CDS             1254..2113
                     /label=ampR
                     /SerialCloner_Color=&hBCC000
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     rep_origin      2291..2963
                     /label=Col E1 origin
                     /SerialCloner_Color=&h0006C0
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=False
     misc_feature    100..456
                     /label=mmoD-strep
                     /SerialCloner_Color=&h007A00
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=note="direct strand translation" \

/SerialCloner_Desc=translation="MDQQTAQPEVRQTLIHADERYQAYTMDLEYMLRWEILRDGEFVQEGCSLSQESAREAVAHVLSHFRRQDATSQNDGGKSEAIRALLREIGTPEPLKD
ENGAAKPAHISAWSHPQFEK*" \
ORIGIN
        1 GATCTCGATC CCGCGAAATT AATACGACTC ACTATAGGGA GGCCACAACG GTTTCCCTCT
       61 AGAAATAATT TTGTTTAACT TTAAGAAGGA GATATACAaa tgGACCAACA AACCGCCCAA
      121 CCGGAAGTGC GCCAAACGCT GATTCATGCT GACGAACGCT ATCAAGCCTA CACGATGGAT
      181 CTGGAATATA TGCTGCGTTG GGAAATTCTG CGCGATGGCG AATTTGTCCA GGAAGGTTGC
      241 AGCCTGTCTC AAGAAAGCGC GCGTGAAGCG GTGGCCCATG TTCTGAGCCA CTTCCGTCGC
      301 CAGGATGCAA CCAGTCAAAA CGACGGCGGT AAATCCGAAG CAATCCGTGC TCTGCTGCGC
      361 GAAATTGGCA CCCCGGAACC GCTGAAGGAC GAAAACGGCG CTGCTAAACC GGCTCACATC
      421 AgcgcTTGGA GCCACCCGCA GTTCGAAAAA TAATAAGCTT GATCCGGCTG CTAACAAAGC
      481 CCGAAAGGAA GCTGAGTTGG CTGCTGCCAC CGCTGAGCAA TAACTAGCAT AACCCCTTGG
      541 GGCCTCTAAA CGGGTCTTGA GGGGTTTTTT GCTGAAAGGA GGAACTATAT CCGGATCTGG
      601 CGTAATAGCG AAGAGGCCCG CACCGATCGC CCTTCCCAAC AGTTGCGCAG CCTGAATGGC
      661 GAATGGGACG CGCCCTGTAG CGGCGCATTA AGCGCGGCGG GTGTGGTGGT TACGCGCAGC
      721 GTGACCGCTA CACTTGCCAG CGCCCTAGCG CCCGCTCCTT TCGCTTTCTT CCCTTCCTTT
      781 CTCGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC GGGGGCTCCC TTTAGGGTTC
      841 CGATTTAGTG CTTTACGGCA CCTCGACCCC AAAAAACTTG ATTAGGGTGA TGGTTCACGT
      901 AGTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTTGA CGTTGGAGTC CACGTTCTTT
      961 AATAGTGGAC TCTTGTTCCA AACTGGAACA ACACTCAACC CTATCTCGGT CTATTCTTTT
     1021 GATTTATAAG GGATTTTGCC GATTTCGGCC TATTGGTTAA AAAATGAGCT GATTTAACAA
     1081 AAATTTAACG CGAATTTTAA CAAAATATTA ACGCTTACAA TTTAGGTGGC ACTTTTCGGG
     1141 GAAATGTGCG CGGAACCCCT ATTTGTTTAT TTTTCTAAAT ACATTCAAAT ATGTATCCGC
     1201 TCATGAGACA ATAACCCTGA TAAATGCTTC AATAATATTG AAAAAGGAAG AGTATGAGTA
     1261 TTCAACATTT CCGTGTCGCC CTTATTCCCT TTTTTGCGGC ATTTTGCCTT CCTGTTTTTG
     1321 CTCACCCAGA AACGCTGGTG AAAGTAAAAG ATGCTGAAGA TCAGTTGGGT GCACGAGTGG
     1381 GTTACATCGA ACTGGATCTC AACAGCGGTA AGATCCTTGA GAGTTTTCGC CCCGAAGAAC
     1441 GTTTTCCAAT GATGAGCACT TTTAAAGTTC TGCTATGTGG CGCGGTATTA TCCCGTATTG
     1501 ACGCCGGGCA AGAGCAACTC GGTCGCCGCA TACACTATTC TCAGAATGAC TTGGTTGAGT
     1561 ACTCACCAGT CACAGAAAAG CATCTTACGG ATGGCATGAC AGTAAGAGAA TTATGCAGTG
```

```
     1621 CTGCCATAAC CATGAGTGAT AACACTGCGG CCAACTTACT TCTGACAACG ATCGGAGGAC
     1681 CGAAGGAGCT AACCGCTTTT TTGCACAACA TGGGGGATCA TGTAACTCGC CTTGATCGTT
     1741 GGGAACCGGA GCTGAATGAA GCCATACCAA ACGACGAGCG TGACACCACG ATGCCTGTAG
     1801 CAATGGCAAC AACGTTGCGC AAACTATTAA CTGGCGAACT ACTTACTCTA GCTTCCCGGC
     1861 AACAATTAAT AGACTGGATG GAGGCGGATA AAGTTGCAGG ACCACTTCTG CGCTCGGCCC
     1921 TTCCGGCTGG CTGGTTTATT GCTGATAAAT CTGGAGCCGG TGAGCGTGGT TCTCGCGGTA
     1981 TCATTGCAGC ACTGGGGCCA GATGGTAAGC CCTCCCGTAT CGTAGTTATC TACACGACGG
     2041 GGAGTCAGGC AACTATGGAT GAACGAAATA GACAGATCGC TGAGATAGGT GCCTCACTGA
     2101 TTAAGCATTG GTAACTGTCA GACCAAGTTT ACTCATATAT ACTTTAGATT GATTTAAAAC
     2161 TTCATTTTTA ATTTAAAAGG ATCTAGGTGA AGATCCTTTT TGATAATCTC ATGACCAAAA
     2221 TCCCTTAACG TGAGTTTTCG TTCCACTGAG CGTCAGACCC CGTAGAAAAG ATCAAAGGAT
     2281 CTTCTTGAGA TCCTTTTTTT CTGCGCGTAA TCTGCTGCTT GCAAACAAAA AAACCACCGC
     2341 TACCAGCGGT GGTTTGTTTG CCGGATCAAG AGCTACCAAC TCTTTTTCCG AAGGTAACTG
     2401 GCTTCAGCAG AGCGCAGATA CCAAATACTG TCCTTCTAGT GTAGCCGTAG TTAGGCCACC
     2461 ACTTCAAGAA CTCTGTAGCA CCGCCTACAT ACCTCGCTCT GCTAATCCTG TTACCAGTGG
     2521 CTGCTGCCAG TGGCGATAAG TCGTGTCTTA CCGGGTTGGA CTCAAGACGA TAGTTACCGG
     2581 ATAAGGCGCA GCGGTCGGGC TGAACGGGGG GTTCGTGCAC ACAGCCCAGC TTGGAGCGAA
     2641 CGACCTACAC CGAACTGAGA TACCTACAGC GTGAGCTATG AGAAAGCGCC ACGCTTCCCG
     2701 AAGGGAGAAA GGCGGACAGG TATCCGGTAA GCGGCAGGGT CGGAACAGGA GAGCGCACGA
     2761 GGGAGCTTCC AGGGGGAAAC GCCTGGTATC TTTATAGTCC TGTCGGGTTT CGCCACCTCT
     2821 GACTTGAGCG TCGATTTTTG TGATGCTCGT CAGGGGGGCG GAGCCTATGG AAAAACGCCA
     2881 GCAACGCGGC CTTTTTACGG TTCCTGGCCT TTTGCTGGCC TTTTGCTCAC ATGTTCTTTC
     2941 CTGCGTTATC CCCTGATTCT GTGGATAACC GTATTACCGC CTTTGAGTGA GCTGATACCG
     3001 CTCGCCGCAG CCGAACGACC GAGCGCAGCG AGTCAGTGAG CGAGGAAGCG GAAGAGCGCC
     3061 CAATACGCAA ACCGCCTCTC CCCGCGCGTT GGCCGATTCA TTAATGCAG
//

LOCUS       pPR-IBA1_mbnI_OB3b.xdna    3286 bp      DNA     circular   15.09.08
DEFINITION
ACCESSION
VERSION
SOURCE
  ORGANISM
COMMENT     Serial Cloner Genbank Format
COMMENT     SerialCloner_Type=DNA
COMMENT     SerialCloner_Comments=Ligation of : pPR-IBA1.xdna [2782  nt] : (#BsaI[172] / #BsaI
            [104]) to PCR(Methylosinus-trichosporium-OB3b
_Mbn.xdna) [496  nt] : (#BsaI[7] / #BsaI[519])
COMMENT     SerialCloner_Ends=0,0,,0,
FEATURES             Location/Qualifiers
     primer_bind     20..39
                     /label=T7F
                     /SerialCloner_Color=&h0C7900
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     misc_feature    684..703
                     /label=T7R
                     /SerialCloner_Color=&hC00400
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     rep_origin      845..1283
                     /label=f1 origin
                     /SerialCloner_Color=&h85009C
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=False
     CDS             1431..2290
                     /label=ampR
                     /SerialCloner_Color=&hBCC000
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
     rep_origin      2468..3140
                     /label=Col E1 origin
                     /SerialCloner_Color=&h0006C0
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=False
     misc_feature    100..633
                     /label=mbnI-strep
                     /SerialCloner_Color=&h84A4C0
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
                     /SerialCloner_Desc=note="direct strand translation" \

/SerialCloner_Desc=translation="MPDQKSNFIRELFLRNKRELLDYLKRRVGPDDAADLLQETFLRALRREDHGAIADPPAFLQQIAINLTRDFARRRRTEANCLVFGDLPDEAPSKEAT
PGDLFEADEESRLLRQAVDALPPRCREAFILYVSARLSPAEIAKRLGISTNMAQKHIRLALQRCRMALESAWSHPQFEK*" \
     CDS             604..633
                     /label=Strep tag
                     /SerialCloner_Color=&hC07E00
                     /SerialCloner_Show=True
                     /SerialCloner_Protect=True
                     /SerialCloner_Arrow=True
ORIGIN
        1 GATCTCGATC CCGCGAAATT AATACGACTC ACTATAGGGA GGCCACAACG GTTTCCCTCT
       61 AGAAATAATT TTGTTTAACT TTAAGAAGGA GATATACAaa tgCCGGACCA AAAATCGAAT
      121 TTCATTCGCG AGCTCTTTCT GCGCAACAAG CGCGAACTGC TCGACTATCT GAAACGCAGG
      181 GTGGGCCCGG ACGATGCTGC GGACCTATTG CAGGAGACGT TCCTCAGAGC GCTGCGGCGC
      241 GAGGATCACG GCGCAATCGC CGATCCGCCG GCCTTCCTGC AACAGATTGC GATCAACCTC
      301 ACCAGGGACT TCGCGCGCCG CCGACGGACG GAGGCCAATT GCCTCGTCTT CGGTGACCTG
      361 CCGGACGAGG CGCCCTCCAA AGAGGCGACG CCGGGAGACC TCTTCGAAGC CGACGAAGAA
      421 TCGCGTCTCC TCCGCCAAGC GGTGGACGCG CTGCCGCCAC GCTGCCGCGA GGCGTTCATC
      481 CTCTACGTTT CTGCGAGGCT GTCTCCTGCG GAGATCGCGA AGCGGCTGGG CATTTCGACG
      541 AATATGGCGC AAAAGCATAT CCGTCTCGCT CTTCAGCGAT GTCGGATGGC GCTGGAAAgc
```

```
      601 gcTTGGAGCC ACCCGCAGTT CGAAAAATAA TAAGCTTGAT CCGGCTGCTA ACAAAGCCCG
      661 AAAGGAAGCT GAGTTGGCTG CTGCCACCGC TGAGCAATAA CTAGCATAAC CCCTTGGGGC
      721 CTCTAAACGG GTCTTGAGGG GTTTTTTGCT GAAAGGAGGA ACTATATCCG GATCTGGCGT
      781 AATAGCGAAG AGGCCCGCAC CGATCGCCCT TCCCAACAGT TGCGCAGCCT GAATGGCGAA
      841 TGGGACGCGC CCTGTAGCGG CGCATTAAGC GCGGCGGGTG TGGTGGTTAC GCGCAGCGTG
      901 ACCGCTACAC TTGCCAGCGC CCTAGCGCCC GCTCCTTTCG CTTTCTTCCC TTCCTTTCTC
      961 GCCACGTTCG CCGGCTTTCC CCGTCAAGCT CTAAATCGGG GGCTCCCTTT AGGGTTCCGA
     1021 TTTAGTGCTT TACGGCACCT CGACCCCAAA AAACTTGATT AGGGTGATGG TTCACGTAGT
     1081 GGGCCATCGC CCTGATAGAC GGTTTTTCGC CCTTTGACGT TGGAGTCCAC GTTCTTTAAT
     1141 AGTGGACTCT TGTTCCAAAC TGGAACAACA CTCAACCCTA TCTCGGTCTA TTCTTTTGAT
     1201 TTATAAGGGA TTTTGCCGAT TTCGGCCTAT TGGTTAAAAA ATGAGCTGAT TTAACAAAAA
     1261 TTTAACGCGA ATTTTAACAA AATATTAACG CTTACAATTT AGGTGGCACT TTTCGGGGAA
     1321 ATGTGCGCGG AACCCCTATT TGTTTATTTT TCTAAATACA TTCAAATATG TATCCGCTCA
     1381 TGAGACAATA ACCCTGATAA ATGCTTCAAT AATATTGAAA AAGGAAGAGT ATGAGTATTC
     1441 AACATTTCCG TGTCGCCCTT ATTCCCTTTT TTGCGGCATT TTGCCTTCCT GTTTTTGCTC
     1501 ACCCAGAAAC GCTGGTGAAA GTAAAAGATG CTGAAGATCA GTTGGGTGCA CGAGTGGGTT
     1561 ACATCGAACT GGATCTCAAC AGCGGTAAGA TCCTTGAGAG TTTTCGCCCC GAAGAACGTT
     1621 TTCCAATGAT GAGCACTTTT AAAGTTCTGC TATGTGGCGC GGTATTATCC CGTATTGACG
     1681 CCGGGCAAGA GCAACTCGGT CGCCGCATAC ACTATTCTCA GAATGACTTG GTTGAGTACT
     1741 CACCAGTCAC AGAAAAGCAT CTTACGGATG GCATGACAGT AAGAGAATTA TGCAGTGCTG
     1801 CCATAACCAT GAGTGATAAC ACTGCGGCCA ACTTACTTCT GACAACGATC GGAGGACCGA
     1861 AGGAGCTAAC CGCTTTTTTG CACAACATGG GGGATCATGT AACTCGCCTT GATCGTTGGG
     1921 AACCGGAGCT GAATGAAGCC ATACCAAACG ACGAGCGTGA CACCACGATG CCTGTAGCAA
     1981 TGGCAACAAC GTTGCGCAAA CTATTAACTG GCGAACTACT TACTCTAGCT TCCCGGCAAC
     2041 AATTAATAGA CTGGATGGAG GCGGATAAAG TTGCAGGACC ACTTCTGCGC TCGGCCCTTC
     2101 CGGCTGGCTG GTTTATTGCT GATAAATCTG GAGCCGGTGA GCGTGGTTCT CGCGGTATCA
     2161 TTGCAGCACT GGGGCCAGAT GGTAAGCCCT CCCGTATCGT AGTTATCTAC ACGACGGGGA
     2221 GTCAGGCAAC TATGGATGAA CGAAATAGAC AGATCGCTGA GATAGGTGCC TCACTGATTA
     2281 AGCATTGGTA ACTGTCAGAC CAAGTTTACT CATATATACT TTAGATTGAT TTAAAACTTC
     2341 ATTTTTAATT TAAAAGGATC TAGGTGAAGA TCCTTTTTGA TAATCTCATG ACCAAAATCC
     2401 CTTAACGTGA GTTTTCGTTC CACTGAGCGT CAGACCCCGT AGAAAAGATC AAAGGATCTT
     2461 CTTGAGATCC TTTTTTTCTG CGCGTAATCT GCTGCTTGCA AACAAAAAAA CCACCGCTAC
     2521 CAGCGGTGGT TTGTTTGCCG GATCAAGAGC TACCAACTCT TTTTCCGAAG GTAACTGGCT
     2581 TCAGCAGAGC GCAGATACCA AATACTGTCC TTCTAGTGTA GCCGTAGTTA GGCCACCACT
     2641 TCAAGAACTC TGTAGCACCG CCTACATACC TCGCTCTGCT AATCCTGTTA CCAGTGGCTG
     2701 CTGCCAGTGG CGATAAGTCG TGTCTTACCG GGTTGGACTC AAGACGATAG TTACCGGATA
     2761 AGGCGCAGCG GTCGGGCTGA ACGGGGGGTT CGTGCACACA GCCCAGCTTG GAGCGAACGA
     2821 CCTACACCGA ACTGAGATAC CTACAGCGTG AGCTATGAGA AAGCGCCACG CTTCCCGAAG
     2881 GGAGAAAGGC GGACAGGTAT CCGGTAAGCG GCAGGGTCGG AACAGGAGAG CGCACGAGGG
     2941 AGCTTCCAGG GGGAAACGCC TGGTATCTTT ATAGTCCTGT CGGGTTTCGC CACCTCTGAC
     3001 TTGAGCGTCG ATTTTTGTGA TGCTCGTCAG GGGGGCGGAG CCTATGGAAA AACGCCAGCA
     3061 ACGCGGCCTT TTTACGGTTC CTGGCCTTTT GCTGGCCTTT TGCTCACATG TTCTTTCCTG
     3121 CGTTATCCCC TGATTCTGTG GATAACCGTA TTACCGCCTT TGAGTGAGCT GATACCGCTC
     3181 GCCGCAGCCG AACGACCGAG CGCAGCGAGT CAGTGAGCGA GGAAGCGGAA GAGCGCCCAA
     3241 TACGCAAACC GCCTCTCCCC GCGCGTTGGC CGATTCATTA ATGCAG
//
```

## File 2. Full HHPred results for aligned proteobacterial 14-helix CopD sequences, indicating CtaG homology in the C-terminal region

```
Query         2585137231/1-548 EK23DRAFT_00619 Putative copper export protein [Methylococaceae sp. 73a :
EK23DRAFT_scaffold00001.1]
Match_columns 548
No_of_seqs    328 out of 1128
Neff          6.8
Searched_HMMs 68562
Date          Mon Aug 17 20:01:21 2015
Command       /cluster/toolkit/production/bioprogs/hhsuite/bin/hhsearch -cpu 4 -v 1 -i
/cluster/toolkit/production/tmp/production/2063359/4077494.hhm -d
/cluster/toolkit/production/databases/hhpred/new_dbs/pdb70_15Aug15/db/pdb.hhm
/cluster/toolkit/production/databases/hhpred/new_dbs/tigrfam_16Sep14/db/tigrfam.hmm
/cluster/toolkit/production/databases/hhpred/new_dbs/supfam_16Sep14/db/supfam.hmm
/cluster/toolkit/production/databases/hhpred/new_dbs/pfam_14Jan15/db//pfam.hhm -o
/cluster/toolkit/production/tmp/production/2063359/4077494.hhr -p 20 -P 20 -Z 100 -B 100 -seq 1 -aliw 80 -local -ssm 2 -norealign
-sc 1 -dbstrlen 10000 -cs /cluster/toolkit/production/bioprogs/hhblits/context_data.lib

 No Hit                             Prob E-value P-value  Score    SS Cols Query HMM  Template HMM
  1 pfam09678 Caa3_CtaG Cytochrome  99.9 1.1E-26 1.6E-31  233.3  17.2  161  373-548     1-192 (241)
  2 TIGR03304 TIGR03304 outer memb  99.6 9.6E-20 1.4E-24    5.1   0.0    5  168-172     1-5   (10)
  3 TIGR04294 TIGR04294 prepilin-t  99.6 1.1E-18 1.6E-23    8.3   0.0    4  144-147    13-16  (16)
  4 TIGR01167 TIGR01167 LPXTG cell  99.4 2.3E-16 3.4E-21   11.0   0.0   25  361-385     9-34  (34)
  5 pfam05425 CopD Copper resistan  99.5 6.6E-13 9.6E-18  115.9  13.1   99  196-297     2-101 (101)
  6 TIGR03057 TIGR03057 X-X-X-Leu-  98.2 1.8E-09 2.7E-14    5.7   0.0    5  213-217    18-22  (28)
  7 TIGR03501 TIGR03501 GlyGly-CTE  97.5 3.9E-07 5.7E-12    8.9   0.0   14  245-258     9-22  (22)
  8 pfam03653 UPF0093 Uncharacteri  97.5  0.0063 9.2E-08   57.7  15.8  135  154-294     3-142 (147)
  9 pfam09980 DUF2214 Predicted me  96.9    0.19 2.8E-06   47.9  17.6  134  158-299     6-146 (150)
 10 TIGR02737 TIGR02737 cytochrome  95.4  0.0012 1.8E-08   62.5   0.0  138  396-548    54-209 (281)
 11 TIGR00756 TIGR00756 pentatrico  95.0  0.0028   4E-08    6.9   0.0   10  479-488    16-25  (35)
 12 pfam10027 DUF2269 Predicted in  93.0      12 0.00017   36.0  17.7  143  158-301     2-150 (150)
 13 pfam13664 DUF4149 Domain of un  86.5      16 0.00024   32.1  10.8   88  165-265     1-93  (100)
 14 pfam07185 DUF1404 Protein of u  63.1 1.3E+02  0.0019   29.9  10.3  118  395-548     2-129 (170)
 15 pfam10355 Ytp1 Protein of unkn  48.6 1.3E+02  0.0018   32.2   8.3  115  424-545   102-228 (268)
 16 pfam11346 DUF3149 Protein of u  20.7 2.9E+02  0.0043   21.1   4.0   29  227-255     3-31  (42)

No 1
>pfam09678 Caa3_CtaG Cytochrome c oxidase caa3 assembly factor (Caa3_CtaG). Members of this family are the CtaG protein required
for assembly of active cytochrome c oxidase of the caa3 type, as found in Bacillus subtilis.
Probab=99.95  E-value=1.1e-26  Score=233.33  Aligned_cols=161  Identities=20%  Similarity=0.297  Sum_probs=0.0

Q ss_pred             HHHHHHHHHHHhhccCCCchh-----HHHHHHHHHHHcCccccCCCHHHHHHHhCcHHHHHHHHHHHHHHHHHHHH----
Q 2585137231/1-5  373 GIFAMLSYVRRFHWAQYWPLG-----FIGLSIFLFFRSDAETWPLGPIGFWESTFGNGEVFQHRIATLLAFVLGYM----  443 (548)
Q Consensus       373 ~~~y~~g~~r~~~~~~~~wp~~-----~~G~~~~~~~~~~~~~~~~w~~g~~~~~~~~~lfs~hm~qH~~l~~~~~~~l~~~----  443 (548)
                        +++|..|.+|+|||+++||..     .+|+++++++++|        |++.|+|.+||+||+||+++++++|||+++
```

```
T Consensus         1 ~~~Y~~g~~~lrr~~~~wp~~r~~~f~~G~~~l~~a~~s-------pl~~~~~~lfs~HmvqhllL~~vaPpLlvlg~P~   73 (241)
T pfam09678         1 AVLYLLGVRRLRRRGDRWPVWRTVAFLLGLALLYLATSS-------PLDAYGHALFSAHMVQHLLLSMVAPPLLVLGAPV   73 (241)
T ss_pred             CchhHHHHHHHHhcCCCCCHHHHHHHHHHHHHHHHHHhC-------cHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHhhCcH


Q ss_pred               ---------HHHHHhhhhhh-----hhhhHHHHHHHHHHHHHHHcchHhHHHHHHhhcchHHHHHHHHHHHHHHHHHHHHh
Q 2585137231/1-5  444 ---------ELKARTRPDAKK-----LQHIFPMLCAFGGILLLTHAHAQFELKSEFLIQSTHTTMGLLAVIMASGRWLEL  509 (548)
Q Consensus       444 ---------~~r~~~~~~~~~~~~-----~~~~~p~~~~~~~~~~~~~~wH~p~~~~~~a~~~~~~~~~~H~~~~~~~~~~~~~~~~~W~~l  509 (548)
                              +.|++.++..++      ..+|..++.++++.+|+||.|..||++++  |+..|.+||.+++++|++|||++
T Consensus        74 ~l~~~~l~~~~~~~~~~~~~~~~~~~~r~l~~P~va~~lf~~~~~~~h~p~~~~~al~--~~~~h~~~~~~~l~sg~lfw~~v  151 (241)
T pfam09678        74 TLLLRALPPRGRRRLGAALHSRPLRFLTHPLVALALFAGSFYLWYLPPLFDAALS--SHWGHWLMHVHFLLSGLLFWWPV  151 (241)
T ss_pred             HHHHhhCCccccchhhhHHhhhhHHHHHccHHHHHHHHHHHHHHHHHHHHHHHHHH--hHHHHHHHHHHHHHHHHHHHHHHHH


Q ss_pred               -ccCcccccccccccccch-------HHHHHHHHHHHHHhcCCCCC
Q 2585137231/1-5  510 -RLVTPDGQSAVEGRIAGF-------VAILAMFIIGNFLMFYREPLY  548 (548)
Q Consensus       510 -~~~~~~~~~~~~~~~~~~~~~~~~~-------~~~~~~~~~~~~~g~~l~f~~~~~~~y  548 (548)
                       ++||+      .+|....       +..+..+.+|+.++|.++|+|
T Consensus       152 l~~~p~------~~~~~~~~~~~~~r~~~l~~~~~~~~~~~~~g~~l~~s~~~ly  192 (241)
T pfam09678       152 LGPDPV------PRRLSYLGRLLYLFAAMPLHAFLGALLMFAPTPLY  192 (241)
T ss_pred             hCCCCC------ccccChHHHHHHHHHHHHHHHHHHHHHhCCCchh
```

No 2
>TIGR03304 TIGR03304 outer membrane insertion C-terminal signal; InterPro: IPR017690 <p>Beta-barrel outer membrane proteins (OMP) rely on Omp85-like proteins for insertion into the outer membrane. The targeting signal for Omp85 is usually found in the C terminus of OMPs, and the target sequences differ somewhat by species. Observations in Escherichia coli mutant PhoE protein and neisserial OMPs suggest that there are alternative, less-efficient recognition sites in the protein that mediate binding to Omp85 [].</p> <p>This entry represents a targeting signal for outer membrane insertion.</p>.
Probab=99.62  E-value=9.6e-20  Score=5.14  Aligned_cols=5  Identities=20%  Similarity=0.354  Sum_probs=0.0

```
Q ss_pred               HHHHH
Q 2585137231/1-5  168 AVWFG  172 (548)
Q Consensus       168 avWvG  172 (548)
                     .+++|
T Consensus         1 ~~~~g    5 (10)
T TIGR03304         1 TFGLG    5 (10)
```

No 3
>TIGR04294 TIGR04294 prepilin-type processing-associated H-X9-DG domain; InterPro: IPR027558 <p>This entry describes a region of approximately 16 residues found typically about 30 residues away from the C terminus of large numbers of proteins in the Planctomycetes, Lentisphaerae, and Verrucomicrobia, on proteins with a prepilin-type N-terminal cleavage/methylation domain (see TIGR02532 from TIGRFAMs). The motif H-X(9)-D-G is nearly invariant. Single genomes may encode over 200 such proteins. </p>.
Probab=99.56  E-value=1.1e-18  Score=8.33  Aligned_cols=4  Identities=25%  Similarity=0.244  Sum_probs=0.0

```
Q ss_pred               cCcc
Q 2585137231/1-5  144 VHAV  147 (548)
Q Consensus       144 GHAa  147 (548)
                     ||.+
T Consensus        13 GHV~   16 (16)
T TIGR04294        13 GHVA   16 (16)
```

No 4
>TIGR01167 TIGR01167 LPXTG cell wall anchor domain; InterPro: IPR019931 <p>This LPXTG motif-containing region is found at the C terminus of many surface proteins of Streptococcus and Streptomyces species. Cleavage between the Thr and Gly by sortase or a related enzyme leads to covalent anchoring at the new C-terminal Thr to the cell wall. Proteins that do not lie at the C terminus or are not found in Gram-positive bacteria are probably false-positive. A common feature of proteins containing this domain appears to be a high proportion of charged and zwitterionic residues immediatedly upstream of the LPXTG motif.</p>.
Probab=99.37  E-value=2.3e-16  Score=11.04  Aligned_cols=25  Identities=20%  Similarity=0.170  Sum_probs=0.0

```
Q ss_pred               cccH-HHHHHHHHHHHHHHHHHHhh
Q 2585137231/1-5  361 NHNV-AGLFLTGMGIFAMLSYVRRFH  385 (548)
Q Consensus       361 ~~~~-~~~~~~~~~~~~~y~~g~~r~~~  385 (548)
                     +.++ ..+.+++++...+...+|.++
T Consensus         9 ~~~~~~~~~~G~~l~~~~~~~~~~~~~rk~~   34 (34)
T TIGR01167         9 SNSLLLLLGLLLLGLAGLLLRKRKKK   34 (34)
```

No 5
>pfam05425 CopD Copper resistance protein D. Copper sequestering activity displayed by some bacteria is determined by copper-binding protein products of the copper resistance operon (cop). CopD, together with CopC, perform copper uptake into the cytoplasm.
Probab=99.49  E-value=6.6e-13  Score=115.92  Aligned_cols=99  Identities=30%  Similarity=0.453  Sum_probs=0.0

```
Q ss_pred               HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHcCchhhhcCCchHHHHHHHHHHHHHHHHHHHHHHHH-hhhHhhccCcchhHHH
Q 2585137231/1-5  196 VAIGRFANMGIASVVLLLLTGLPLAWVYIGSWDGLFGTGYGSLVVTKVWLLLVALGFAYLNNR-AGRRWRENPGDNTLIT  274 (548)
Q Consensus       196 ~~l~RFS~~A~~av~~l~~sGv~~a~~~l~~~~~~l~~T~YG~lLlvKl~lv~~ll~la~~~~r~-~~p~l~~~~~~~~~~~~~~~~~  274 (548)
                     ...+|||+++..|+.++++||+++++.+.+ ++++++|+||+.+.+|++++.++.++++.||+ ..|+++++++..   .+
T Consensus         2 ~~~~rf~~~~~~~~~~~~~~~~sG~~~~~~~~~~~~~~~-~~~l~~t~yg~~l~~Kl~lv~~~~~~~~~~~~~~~~l~~~~~~~~-~~   78 (101)
T pfam05425         2 AVLRRFSRLALVAVALLLVTGVVNALVLVG-LGGLLTTAYGRLLLAKLALVVLMLALAAANRFRLVPRLRRAWGRA--LG   78 (101)
T ss_pred             HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHc-ccccccChHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHhcccchh--HH


Q ss_pred               HHHHHHHHHHHHHHHHHHHHHHH
Q 2585137231/1-5  275 KVPYYIEAETFILIGILFIAATL  297 (548)
Q Consensus       275 ~l~r~i~~E~~l~~~vl~vaa~L  297 (548)
                     +++++...|.+++++++..+++.+
T Consensus        79 ~~~~~~~~~~~e~~l~~~~il~~~~~~~~l  101 (101)
T pfam05425        79 RLRRSVAIELVLGLLVLGLAAVL  101 (101)
T ss_pred             HHHHHHHHHHHHHHHHHHHHHC
```

No 6
>TIGR03057 TIGR03057 X-X-X-Leu-X-X-Gly heptad repeats; InterPro: IPR023908 <p> This entry represents a repeat of seven residues each, in which the fourth residue tends to be Leu and the seventh tends to be Gly in each set. This heptad periodicity,

corresponding to two turns of an alpha helix, suggests alpha-helical structure. Arrangement of these sequences on a helical wheel would show a strict alternation of Leu and Gly residues on one side of the helix, that is, an extremely bulky side chain alternating with the virtual absence of one. This suggests an extended zippering of one alpha helix to another, analogous to the shorter leucine zippers found in many dimerizing transcription factors. Proteins in which these heptad repeats occur often have higher order repeats of a unit comprised of several heptads. </p>.
Probab=98.25  E-value=1.8e-09  Score=5.69  Aligned_cols=5  Identities=40%  Similarity=0.647  Sum_probs=0.0

```
Q ss_pred              HHHHH
Q 2585137231/1-5  213 LLTGL  217 (548)
Q Consensus       213 ~~sGv  217 (548)
                      +.+|+
T Consensus        18 L~~G~   22 (28)
T TIGR03057        18 LASGA   22 (28)
```

No 7
>TIGR03501 TIGR03501 GlyGly-CTERM domain; InterPro: IPR020008 <p> This homology domain, GlyGly-CTERM, shares a species distribution with rhombosortase (TIGR03902 from TIGRFAMs), a subfamily of rhomboid-like intramembrane serine proteases []. It is probably a recognition sequence for protein sorting and then cleavage by rhombosortase. Shewanella species have the largest number of target proteins per genome, up to thirteen. The domain occurs at the extreme carboxyl-terminus of a diverse set of proteins, most of which are enzymes with conventional signal sequences and with hydrolytic activities: nucleases, proteases, agarases, etc. The agarase AgaA from Vibro sp. strain JT0107 is secreted into the medium, while the same protein heterologously expressed in E. coli is retained in the cell fraction []. This suggests cleavage and release in species with this domain. Both this suggestion, and the chemical structure of the domain (motif, hydrophobic predicted transmembrane helix, cluster of basic residues) closely parallels that of the LPXTG/sortase system and the PEP-CTERM/exosortase(EpsH) system. For this reason, the putative processing enzyme is designated rhombosortase. </p>.
Probab=97.52  E-value=3.9e-07  Score=8.92  Aligned_cols=14  Identities=21%  Similarity=0.233  Sum_probs=0.0

```
Q ss_pred              HHHHHHHHHHHHHH
Q 2585137231/1-5  245 LLLVALGFAYLNNR  258 (548)
Q Consensus       245 lv~~ll~la~~~r~  258 (548)
                      .+..+..++.+.|+
T Consensus         9 ~LllL~~l~~~Rr~   22 (22)
T TIGR03501         9 SLLLLLLLGLRRRK   22 (22)
```

No 8
>pfam03653 UPF0093 Uncharacterized protein family (UPF0093).
Probab=97.55  E-value=0.0063  Score=57.70  Aligned_cols=135  Identities=14%  Similarity=0.166  Sum_probs=0.0

```
Q ss_pred              hHHHHHHHHHHHHHHHHHHHHHHHHHHHHHhhccCcchh----hHHHHHHHH-HHHHHHHHHHHHHHHHHHHHHHHcCchh
Q 2585137231/1-5  154 ALLMATTVVHQFAAAVWFGCVAQLLALWRAARKDADAL----ALWPVAIGR-FANMGIASVVLLLLTGLPLAWVYIGSWD  228 (548)
Q Consensus       154 ~~a~~~~~lHllaaavWvGgL~~l~~l~~~~~~~~~~~~~~~----~~~~~~l~R-FS~~A~~av~~l~~sGv~~a~~~l~~~~  228 (548)
                      +.-...-.+|++++..|+||++++-.+.-...+.++..    +......|| |..+...+..+..++|+.+....-+
T Consensus         3 ~~y~wlKalHIi~vv~W~aGlfylprl~v~~~~~~~~~~~~~~~~~~l~~~~~rl~~~~i~~Pami~a~~~G~~ll~~~~g---   79 (147)
T pfam03653         3 DYYLWIKALHIIFVISWMAGLFYLPRLFVYHAEAAPGSDESDEIFKIMERRLLRYIMNPAMILTWVFGLLLLAITPG---   79 (147)
T ss_pred              hhHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHcccccchhHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHhhh---


Q ss_pred              hhcCCchHHHHHHHHHHHHHHHHHHHHHhhhHhhccCcchhHHHHHHHHHHHHHHHHHHHHHHHHH
Q 2585137231/1-5  229 GLFGTGYGSLVVTKVWLLLVALGFAYLNNRAGRRWRENPGDNTLITKVPYYIEAETFILIGILFIA  294 (548)
Q Consensus       229 ~l~~T~YG~lLlvKl~lv~~ll~la~~~r~~~p~l~~~~~~~~~~~~~~~~l~r~i~~E~~l~~~vl~va  294 (548)
                        .....+-.+-+|++++.++.+.-.+..+..+++++++....- +.+|..-++-.+++++++.++
T Consensus        80 --~~~~~~~~Wl~~KL~lV~lL~~~H~~~~~~lk~~~~~~~~~~~s~-~~~R~~ne~p~i~li~Iv~lv  142 (147)
T pfam03653        80 --LLDWSSGWLHVKLALVLLLTGYHGWLGKYRKDFAAGENTRSG-RFYRILNEVPTLLMIVIVILV  142 (147)
T ss_pred              --HHHhcCcHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCH-HHHHHHHHHHHHHHHHHhHhe
```

No 9
>pfam09980 DUF2214 Predicted membrane protein (DUF2214). This domain, found in various hypothetical bacterial proteins, has no known function.
Probab=96.86  E-value=0.19  Score=47.86  Aligned_cols=134  Identities=19%  Similarity=0.083  Sum_probs=0.0

```
Q ss_pred              HHHHHHHHHHHHHHHHHHHHHHHHHhhccCcchhhHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHcCchhhhcCCchHH
Q 2585137231/1-5  158 ATTVVHQFAAAVWFGCVAQLLALWRAARKDADALALWPVAIGRFANMGIASVVLLLLTGLPLAWVYIGSWDGLFGTGYGS  237 (548)
Q Consensus       158 ~~~~lHllaaavWvGgL~~l~~l~~~~~~~~~~~~~~~~~~~~~~~l~RFS~~A~~av~~l~~sGv~~a~~~l~~~~~l~~T~YG~  237 (548)
                      ..-.+|.++...=++.+..=..++|     ++-.....+.+.|.-.+-..|..++++||+.......-+.+--.+.+
T Consensus         6 llAy~H~lai~~l~~~L~~E~~llr-----~~~~~~~~~~rL~~~D~~yGlaA~~vl~tG~lRv~~~~KG~~fY~~nP---   77 (150)
T pfam09980         6 LLAYVHYLAIFVLFAALTAERLLLR-----RDMSAAELRRLVIADIVYGLAALLLLITGLLRVFYFGKGSEFYLHNP---   77 (150)
T ss_pred              HHHHHHHHHHHHHHHHHHHHHHHhc-----CCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHhcchhHHHHhCh---


Q ss_pred              HHHHHHHHHHHHHHHHHHHHH-hhhHhhccCcc------hhHHHHHHHHHHHHHHHHHHHHHHHHHHHHcc
Q 2585137231/1-5  238 LVVTKVWLLLVALGFAYLNNR-AGRRWRENPGD------NTLITKVPYYIEAETFILIGILFIAATLSS  299 (548)
Q Consensus       238 lLlvKl~lv~~ll~la~~~r~-~~p~l~~~~~~~~------~~~~~~~~l~r~i~~E~~l~~~vl~vaa~Lt~  299 (548)
                      +.-.|+.+++++..+....-. .++..++.+.+      .+..++.++++..|+.+...+-.+++.|++
T Consensus        78 ~F~~K~~lFvligllSi~PT~~firWr~~l~~~~~~p~~~~~~~~~~~r~~~~v~~E~~ll~lipl~A~lMAR  146 (150)
T pfam09980        78 LFHLKVTLFVLIGLLSLYPTITFIRWRIPLRKGPAPVVSIGLAKRVRRIVNIELLLFALIPLLATLMAR  146 (150)
T ss_pred              HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHhcCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHhc
```

No 10
>TIGR02737 TIGR02737 cytochrome c oxidase assembly factor CtaG; InterPro: IPR014108 <p>CtaG is required for assembly of active cytochrome c oxidase of the caa3 type, as found in Bacillus subtilis []. This entry represents CtaG from B.subtilis and other Bacilli.</p>.
Probab=95.41  E-value=0.0012  Score=62.46  Aligned_cols=138  Identities=18%  Similarity=0.262  Sum_probs=0.0

```
Q ss_pred              HHHHHHHHHcCccccCCCHHHHHHHHhCcHHHHHHHHHHHHHHHHHHH----------HHHHHhhhhhhhhhhhHHHHHH
Q 2585137231/1-5  396 GLSIFLFFRSDAETWPLGPIGFWESTFGNGEVFQHRIATLLAFVLGYM----------ELKARTRPDAKKLQHIFPMLCA  465 (548)
Q Consensus       396 G~~~~~~~~~~~~~~~~w~~g~~~~~~~~~~lfs~hm~qH~~l~~~~~~~~l~~~----------~~r~~~~~~~~~~~~~~~~~~~p~~~~  465 (548)
                      ++..+.....+       |++..+...++.||+|-.+..++++.+...          ..+.......+....|.-...+
T Consensus        54 ~~~~ly~~~g~-------p~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~g~p~~~~~~~~~~~~~~~~~~~~~~~~~p~~~l~~  126 (281)
T TIGR02737        54 GLLLLYIVKGS-------PLDLLGHLILTAHMVQMAVLYLVVPPLLLLGIPAWLYEKILERPFVKAVLKLLTKPLIALLL  126 (281)


Q ss_pred              HHHHHHHHcchHhHHHHHhhcchHHHHHHHHHHHHHHHHHHHHh-ccCccccccccccccccch-------HHHHHHHHHH
Q 2585137231/1-5  466 FGGILLLTHAHAQFELKSEFLIQSTHTTMGLLAVIMASGRWLEL-RLVTPDGQSAVEGRIAGF-------VAILAMFIIG  537 (548)
```

```
Q Consensus       466 ~~~~~~~wH~p~~~~~a~~~~~~~~H~~~~~~~~~~~~~W~~l-~~~~~~~~~~~~~~~~~~-------~~~~~~~~~g  537 (548)
                      +++..-+.|.|..||...+  ++..|.......+..+...||++ ...++      .++..+.       ...+..+..+
T Consensus       127 f~g~f~~y~~p~~f~~~~~-~~~~~~~~~~~~~~~~~a~~~w~~~~~~~~~~------~~~l~~~~~~gy~~a~~~l~~pac  198 (281)
T TIGR02737       127 FNGLFSLYHVPLVFDTVKQ--SPLLHTLITLLLLVAAFLMWWPLLAKLKE------LQKLSDLKKLGYIFANGILLTPAC  198 (281)


Q ss_pred              HHHHhcCCCCC
Q 2585137231/1-5  538 NFLMFYREPLY  548 (548)
Q Consensus       538 ~~l~f~~~~~~y  548 (548)
                      .++.|.++|+|
T Consensus       199 ~l~~f~~~pl~  209 (281)
T TIGR02737       199 ALIIFASAPLY  209 (281)


No 11
>TIGR00756 TIGR00756 pentatricopeptide repeat domain; InterPro: IPR002885 <p>This entry represents the PPR repeat.</p>
<p>Pentatricopeptide repeat (PPR) proteins are characterised by tandem repeats of a degenerate 35 amino acid motif []. Most of
PPR proteins have roles in mitochondria or plastid []. PPR repeats were discovered while screening Arabidopsis proteins for those
predicted to be targeted to mitochondria or chloroplast [, ]. Some of these proteins have been shown to play a role in post-
transcriptional processes within organelles and they are thought to be sequence-specific RNA-binding proteins [, , ]. Plant
genomes have between one hundred to five hundred PPR genes per genome whereas non-plant genomes encode two to six PPR
proteins.</p> <p>Although no PPR structures are yet known, the motif is predicted to fold into a helix-turn-helix structure
similar to those found in the tetratricopeptide repeat (TPR) family (see PDOC50005 from PROSITEDOC) []. </p> <p>The plant PPR
protein family has been divided in two subfamilies on the basis of their motif content and organisation [, ].</p> <p>Examples of
PPR repeat-containing proteins include PET309 P32522 from SWISSPROT, which may be involved in RNA stabilisation [], and crp1,
which is involved in RNA processing []. The repeat is associated with a predicted plant protein O49549 from SWISSPROT that has a
domain organisation similar to the human BRCA1 protein.</p>.
Probab=94.96  E-value=0.0028  Score=6.85  Aligned_cols=10  Identities=20%  Similarity=0.132  Sum_probs=0.0

Q ss_pred              HHHHHhhcch
Q 2585137231/1-5  479 FELKSEFLIQ  488 (548)
Q Consensus       479 ~~~a~~~~~~  488 (548)
                      ++-|.+++++
T Consensus        16 ~~~a~~~~~~   25 (35)
T TIGR00756        16 VEEALELFDE   25 (35)


No 12
>pfam10027 DUF2269 Predicted integral membrane protein (DUF2269). Members of this family of bacterial hypothetical integral
membrane proteins have no known function.
Probab=92.96  E-value=12  Score=36.03  Aligned_cols=143  Identities=17%  Similarity=0.082  Sum_probs=0.0

Q ss_pred              HHHHHHHHHHHHHHHHHHHHH-HHHhhccCcchh-hHHHHHHHHHHHH-HHHHHHHHHHHHHHHHHcCchh-hhcCC
Q 2585137231/1-5  158 ATTVVHQFAAAVWFGCVAQLLA-LWRAARKDADAL-ALWPVAIGRFANM-GIASVVLLLLTGLPLAWVYIGSWD-GLFGT  233 (548)
Q Consensus       158 ~~~~lHllaaavWvGgL~~l~~-l~~~~~~~~~~~~~-~~~~~~~l~RFS~~-A~~av~~l~~sGv~~a~~~l~~~~~-~l~~T  233 (548)
                      ....+|++++.+|+|+...... ..+..+.++++. ....+..++..++ ...+..+..++|+.+....--++. ..+..
T Consensus         2 ~lk~lHV~aaill~G~~~~~~a~~~~~~a~~~~~~~~~~~~~~~~v~~~d~~f~~p~~i~~~v~Gl~l~~~~~~~~~~~~~Wl~~   81 (150)
T pfam10027         2 LLKFLHILSAVLLFGTGIGTAFYMLRARRTGDPAVIARAARRVVLADWLFTAPAVILQPITGLWLAHLAGWPLGQPWLLG   81 (150)
T ss_pred              hHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHhcCCHhHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHhcCCCcCChHHHH


Q ss_pred              chHHHHHHHHHHHHHHHHHHHHHH-hhhHhhccCc-chhHHHHHHHHHHHHHHHHHHHHHHHHHHHHcCC
Q 2585137231/1-5  234 GYGSLVVTKVWLLLVALGFAYLNNR-AGRRWRENPG-DNTLITKVPYYIEAETFILIGILFIAATLSSQP  301 (548)
Q Consensus       234 ~YG~lLlvKl~lv~~ll~la~~~r~-~~p~l~~~~~~-~~~~~~~~l~r~i~~E~~l~~~vl~vaa~Lt~~p  301 (548)
                      ++...+++-...+.++. .--..++ .-...+++++ .++..|..|++...+.......+.+-..|..=|
T Consensus        82 s~~L~~ia~~~~l~vv~-~q~~~~~~~~~~~~~~~~~g~~l~~~y~r~~~~~w~~~g~~a~~~~~i~i~~LMV~KP  150 (150)
T pfam10027        82 SLVLYLVAGACWLPVVW-LQIRLRRLARAAAETGAPLPPEYHRLFRIWFALGVPAFIALVAIFWLMVAKP  150 (150)
T ss_pred              HHHHHHHHHHHHHHHH-HHHHHHHHHHHHHhcCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHcCc


No 13
>pfam13664 DUF4149 Domain of unknown function (DUF4149).
Probab=86.51  E-value=16  Score=32.08  Aligned_cols=88  Identities=28%  Similarity=0.254  Sum_probs=0.0

Q ss_pred              HHHHHHHHHHHHH----HHHHHhhccCcchhhHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHcCchhhhcCCchHHHHH
Q 2585137231/1-5  165 FAAAVWFGCVAQL----LALWRAARKDADALALWPVAIGRFANMGIASVVLLLLTGLPLAWVYIGSWDGLFGTGYGSLVV  240 (548)
Q Consensus       165 laaavWvGgL~~l----~~l~~~~~~~~~~~~~~~~~~~~~~~~l~RFS~~A~~av~~l~~sGv~~a~~~l~~~~~~l~~T~YG~lLl  240 (548)
                      +..+.|+|+.+..    .....+.-++++..+......++|..+...|..+++++....-             ..+..--
T Consensus         1 l~~~~w~G~~~~~~~~fv~ap~lF~~L~~~~~ag~l~~~lFp~y~~~~~~~~~~~i~l~~~~~~~~~-------------------~~~~~~~~   67 (100)
T pfam13664         1 LLLALWLGGQLFVGFVVAPVLFKALPREQAGALQGKLFPVYFLLGLALALLLLLTELLLH-------------LSGKLAR   67 (100)
T ss_pred              ChHHHHHHHHHHHHHHHHHHHHHhCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHhc-------------ccchHHH


Q ss_pred              HHHHHHHHHHHHHHHHH-hhhHhhc
Q 2585137231/1-5  241 TKVWLLLVALGFAYLNNR-AGRRWRE  265 (548)
Q Consensus       241 vKl~lv~~ll~la~~~r~-~~p~l~~  265 (548)
                      .+..++..+......|.+ +.|+..+
T Consensus        68 ~~~~~l~~~~~~~~~~n~~~l~P~~~~~   93 (100)
T pfam13664        68 LQLLLLAVVLLLTLLNLFYLTPRITA   93 (100)
T ss_pred              HHHHHHHHHHHHHHHHHHHHhHHHHH


No 14
>pfam07185 DUF1404 Protein of unknown function (DUF1404). This family consists of several archaeal proteins of around 180
residues in length. Members of this family seem to be found exclusively in Sulfolobus tokodaii and Sulfolobus solfataricus. The
function of this family is unknown.
Probab=63.08  E-value=1.3e+02  Score=29.93  Aligned_cols=118  Identities=16%  Similarity=0.185  Sum_probs=0.0

Q ss_pred              HHHHHHHHHcCccccCCCHHHHHHHhCcHHHHHHHHHHHHHHHHHHHHHHHHHHhhhhhhhhhhHHHHHHHHHHHHHc
Q 2585137231/1-5  395 IGLSIFLFFRSDAETWPLGPIGFWESTFGNGEVFQHRIATLLAFVLGYMELKARTRPDAKKLQHIFPMLCAFGGILLLTH  474 (548)
Q Consensus       395 ~G~~~~~~~~~~~~~~~w~~g~~~~~~~~~~~lfs~hm~qH~~l~~~~~~~~~l~~~~~~~r~~~~~~~~~~~~~~~~~~~~~~p~~~~~~~~~~~~~~~wH  474 (548)
                      .|...++...+-       =-..++...=-.-|.-|+.+-.....++.--.|..               ...+..-+...-+||
T Consensus         2 ~~i~L~i~svNP-------~te~l~~~~~~~~~~~M~sHYaL~~~G~llGy~~~~~~~-----------~~~~~~ig~~~~~v~WH   63 (170)
T pfam07185         2 LGLALIIISVNP-------FTESLEFKNPIVYMLSHYALYIGGILIGYKLFKGK-----------YISLIIGGILAIFWH   63 (170)
T ss_pred              ccHHHHHHhch-------hHHHHhccChHHHHHHHHHHHHHHHHHHHHHHHhccc-----------hhHhhhcchhhheee
```

```
Q ss_pred             chHhHHHHHhhcchHHHHHHHHHHHHHHHHHHHHHhccCcccccccccccch----------HHHHHHHHHHHHHHhcC
Q 2585137231/1-5  475 AHAQFELKSEFLIQSTHTTMGLLAVIMASGRWLELRLVTPDGQSAVEGRIAGF----------VAILAMFIIGNFLMFYR  544 (548)
Q Consensus       475 ~p~~~~~a~~~~~~~~~H~~~~~~~~~~~~~~~~W~~l~~~~~~~~~~~~~~~~~~~~~~~----------~~~~~~~~~~g~~l~f~~  544 (548)
                      .|..||+.-.  ++....++-+++++.|.+-                ...+...          .|.+.=+.+..++...
T Consensus        64 lP~fF~l~a~--~~~~R~i~elsl~lgGiL~---------------Gss~~~m~~~~Ki~Lf~lwM~gDT~LsI~liig-  125 (170)
T pfam07185        64 LPYFFDLGAG--SLSYRILLEISLLLGGILI---------------GSSIRKMSLVFKISLFALWMIGDTLLSIFLILG-  125 (170)
T ss_pred             cchhHhhhhh--hHHHHHHHHHHHHHHHHHh---------------cccHHHHHHHHHHHHHHHHHHHHHHHHHHHHHhc-


Q ss_pred             CCCC
Q 2585137231/1-5  545 EPLY  548 (548)
Q Consensus       545 ~~~y  548 (548)
                      .|.|
T Consensus       126 ~p~Y  129 (170)
T pfam07185       126 SPLY  129 (170)
T ss_pred             Cccc
```

No 15

>pfam10355 Ytp1 Protein of unknown function (Ytp1). This is a family of proteins found in fungi. The region appears to contain regions similar to mitochondrial electron transport proteins. The C-terminal domain is hydrophobic and negatively charged. There are consensus sites for both N-linked glycosylation and cAMP-dependent protein kinase phosphorylation.

Probab=48.63  E-value=1.3e+02  Score=32.16  Aligned_cols=115  Identities=17%  Similarity=0.165  Sum_probs=0.0

```
Q ss_pred             cHHHHHHHHHHHHHHHHHHH-------HHHHHhhhhhh-hhhhHHHHHHHHHHHHHHHccchHhHHHHHHhhcchHHHHHHH
Q 2585137231/1-5  424 NGEVFQHRIATLLAFVLGYM-------ELKARTRPDAKK-LQHIFPMLCAFGGILLLTHAHAQFELKSEFLIQSTHTTMG  495 (548)
Q Consensus       424 s~hm~qH~~l~~~~~~~l~~~~-------~~r~~~~~~~~~~~-~~~~~~p~~~~~~~~~~~~~wH~p~~~~~~a~~~~~~~~~~H~~~~  495 (548)
                      +.+=+||-...++----+..       ..|.+....... .++|+|+.+++   +..|-.-.-.|..+-  +..+|..-+
T Consensus       102 s~~DlQH~si~v~~~~gGL~G~~le~~~~~~~~~~~~~~~~~s~N~~PalvI~---~TG~~MS~H~Q~~~~--St~VH~~wG  176 (268)
T pfam10355       102 SAKDLQHTSIGIMFWGGGLCGMLLERKRVRWRLNSASYGFSRNPIPALTII---LTGILMSSHHQHSMI--STKIHKQWG  176 (268)
T ss_pred             ChhhHHHHHHHHHHHHhHhHheeeccccchhhhcCCCCCCCCCccHHHHH---HHHHHHHhcCchhHH--HHHHHHHHH


Q ss_pred             HHHHHHHHHHHHHh---ccCcccccccccccccc-hHHHHHHHHHHHHHHhcCC
Q 2585137231/1-5  496 LLAVIMASGRWLEL---RLVTPDGQSAVEGRIAG-FVAILAMFIIGNFLMFYRE  545 (548)
Q Consensus       496 ~~~~~~~~~~~~W~~l---~~~~~~~~~~~~~~~~~~~~~~~~~~-~~~~~~~~~~~~~g~~l~f~~~  545 (548)
                      ......+..|..++   -.||+  .+...+|-.- .+.+.+.+.-|++.|-+.|
T Consensus       177 ~lL~~ag~~Riit~~~l~~~~~~--~s~~ps~p~~e~l~~F~Li~gG~vFM~Ste  228 (268)
T pfam10355       177 YLLMGAGLFRIIEILFLLLDPP--SSSLPSRPPTEYLTPFCLIAGGLVFMGSTE  228 (268)
T ss_pred             HHHHHHHHHHhheeeEEccCC--cccCCCCCchhHHHHHHHHHhHheeecCH
```

No 16

>pfam11346 DUF3149 Protein of unknown function (DUF3149). This bacterial family of proteins has no known function.

Probab=20.72  E-value=2.9e+02  Score=21.10  Aligned_cols=29  Identities=21%  Similarity=0.471  Sum_probs=0.0

```
Q ss_pred             hhhhcCCchHHHHHHHHHHHHHHHHHHH
Q 2585137231/1-5  227 WDGLFGTGYGSLVVTKVWLLLVALGFAYL  255 (548)
Q Consensus       227 ~~~l~~T~YG~lLlvKl~lv~~ll~la~~  255 (548)
                      |.+|++|+||..-++-++....+...-.+
T Consensus         3 W~~LF~t~iGL~Sl~vI~~~lgm~~f~~~   31 (42)
T pfam11346         3 WLDLFGNDIGLMSLLVIFFTIGMMAFFGR   31 (42)
T ss_pred             HHHHhcCcHHHHHHHHHHHHHHHHHHHHH
```

Done!

## Supplemental References


1 S. A. Bustin, V. Benes, J. A. Garson, J. Hellemans, J. F. Huggett, M. Kubista, R. Mueller, T. Nolan, M. W. Pfaffl and G. L. Shipley, *Clin Chem*, 2009, **55**, 611.
2 S. Fleige and M. W. Pfaffl, *Mol Aspects Med*, 2006, **27**, 126–139.
3 J. Hellemans, G. Mortier, A. de Paepe, F. Speleman and J. Vandesompele, *Genome Biol*, 2007, **8**, R19.
4 I. Rieu and S. J. Powers, *Plant Cell*, 2009, **21**, 1031–1033.
5 J. Vandesompele, K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. de Paepe and F. Speleman, *Genome Biol*, 2002, **3**, research0034.1.
6 C. W. Knapp, D. A. Fowle, E. Kulczycki, J. A. Roberts and D. W. Graham, *Proc Nat Acad Sci USA*, 2007, **104**, 12040–12045.
7 J. D. Semrau, S. Jagadevan, A. A. DiSpirito, A. Khalifa, J. Scanlan, B. H. Bergman, B. C. Freemeier, B. S. Baral, N. L. Bandow, A. Vorobev, D. H. Haft, S. Vuilleumier and J. C. Murrell, *Environ Microbiol*, 2013, **15**, 3077–3086.
8 G. Giannoukos, D. M. Ciulla, K. Huang, B. J. Haas, J. Izard, J. Z. Levin, J. Livny, A. M. Earl, D. Gevers, D. V. Ward, C. Nusbaum, B. W. Birren and A. Gnirke, *Genome Biol*, 2012, **13**, R23.
9 T. Theis, R. A. Skurray and M. H. Brown, *J Microbiol Methods*, 2007, **70**, 355–362.
10 R. Gil, F. J. Silva, J. Pereto and A. Moya, *Microbiol Mol Biol R*, 2004, **68**, 518–537.
11 P. A. Nieto, P. C. Covarrubias, E. Jedlicki, D. S. Holmes and R. Quatrini, *BMC Mol Biol*, 2009, **10**, 63.
12 N. L. Bandow, W. H. Gallagher, L. A. Behling, D. W. Choi, J. D. Semrau, S. C. Hartsel, V. S. Gilles and A. A. DiSpirito, in *Methods in Methane Metabolism, Part B, 1, (2011) 1-14. doi:10.1016/B978-0-12-386905-0.00001-2*, Methods in enzymology, 2011, vol. 495, pp. 259–269.
13 A. Krogh, B. Larsson, G. von Heijne and E. L. Sonnhammer, *J Mol Bio*, 2001, **305**, 567–580.
14 K. D. Tsirigos, C. Peters, N. Shu, L. Käll and A. Elofsson, *Nucleic Acids Res*, 2015, **43**, W401–7.
15 J. Soding, A. Biegert and A. N. Lupas, *Nucleic Acids Res*, 2005, **33**, W244–W248.
16 K. Katoh and D. M. Standley, *Mol Biol Evol*, 2013, **30**, 772–780.
17 R. D. Finn, J. Clements, W. Arndt, B. L. Miller, T. J. Wheeler, F. Schreiber, A. Bateman and S. R. Eddy, *Nucleic Acids Res*, 2015, **43**, W30–8.
18 K. Katoh and D. M. Standley, *Methods Mol Biol*, 2014, **1079**, 131–146.