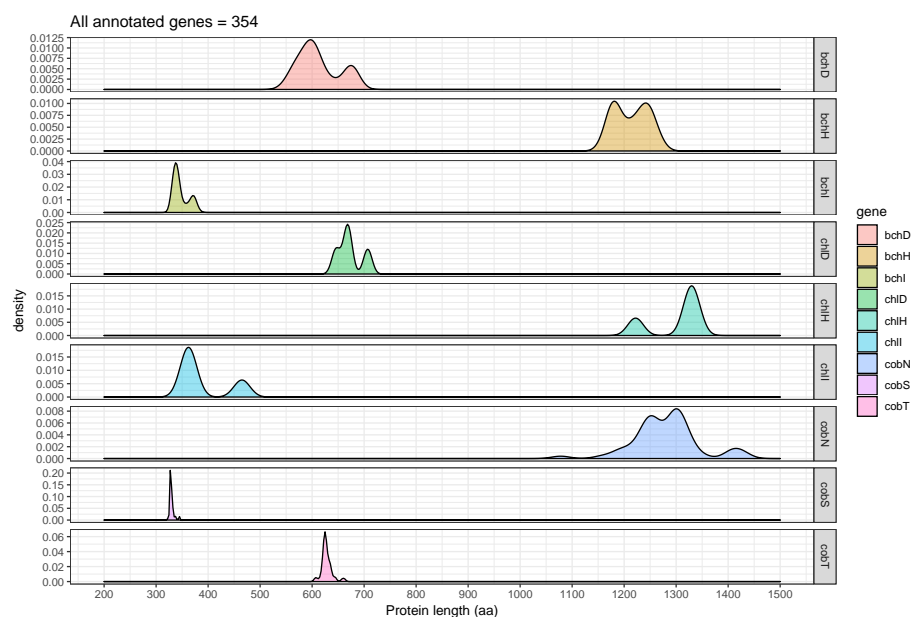
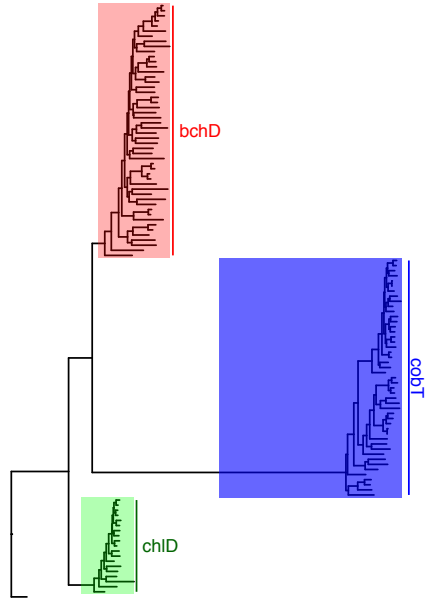


# 1 Supplementary figures

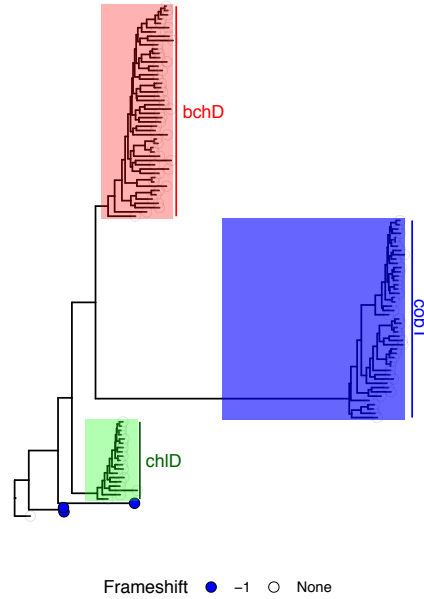


Supplementary Figure 1: Lengths of the annotated subunits of the cobNST, chlIDH and bchIDH chelatases.

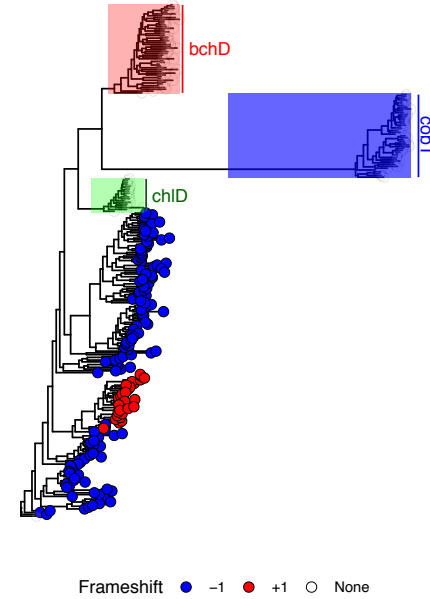
**A** Total number of genes in the tree = 117



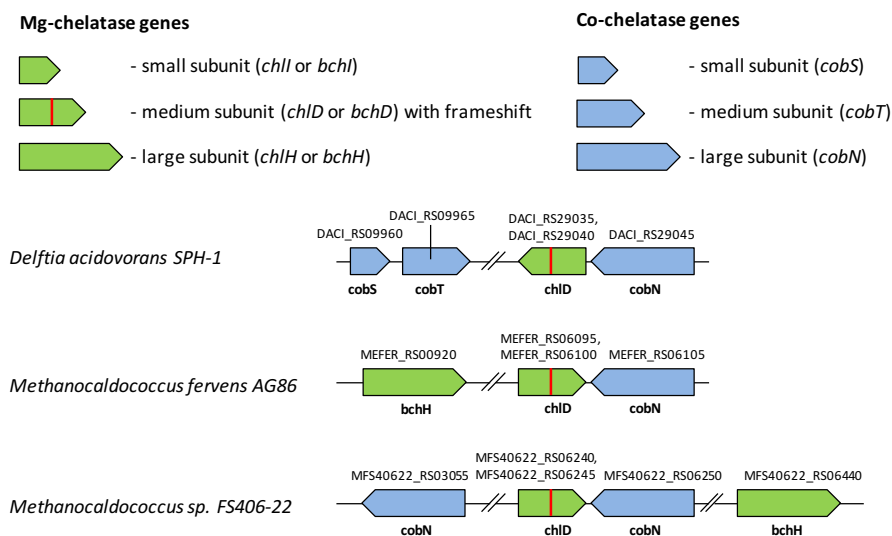
**B** Total number of genes in the tree = 120  
Number of fs-chID genes = 3



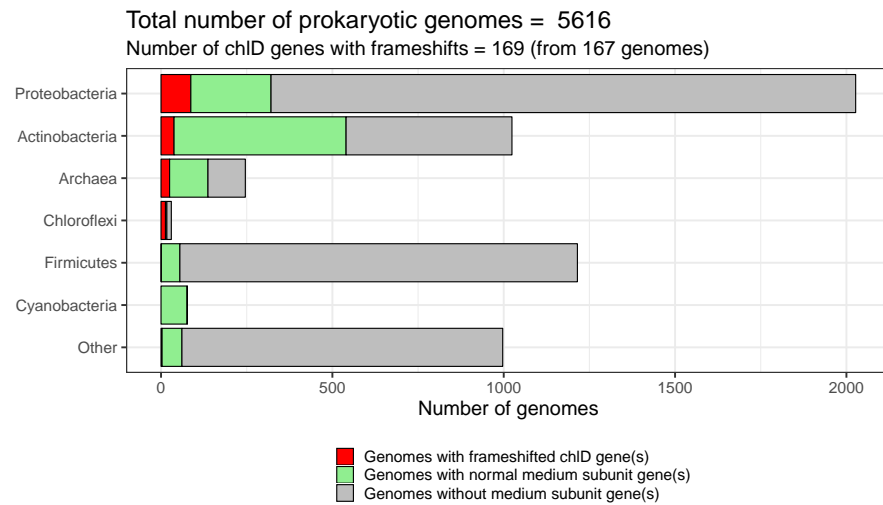
**C** Total number of genes in the tree = 286  
Number of fs-chID genes = 169



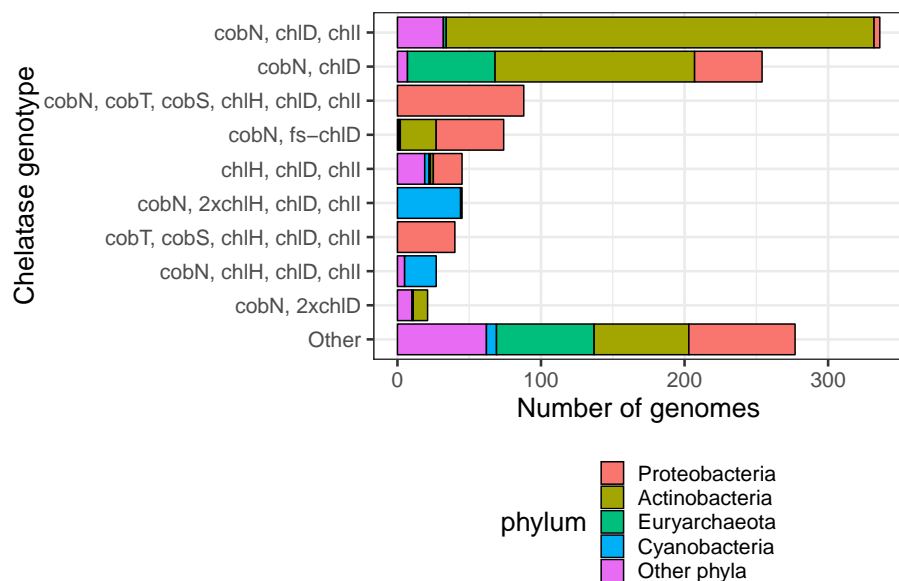
Supplementary Figure 2: The reference phylogenetic tree of the annotated medium chelatase subunit proteins and the location of the long products from the fs-chID genes. The only annotated archaeal bchD gene (METFOR\_RS04410 from *Methanoregula formica* SMSP) was used as the outgroup. The -1 and +1 frameshift types in the fs-chID genes are indicated by the blue and red colors, respectively.



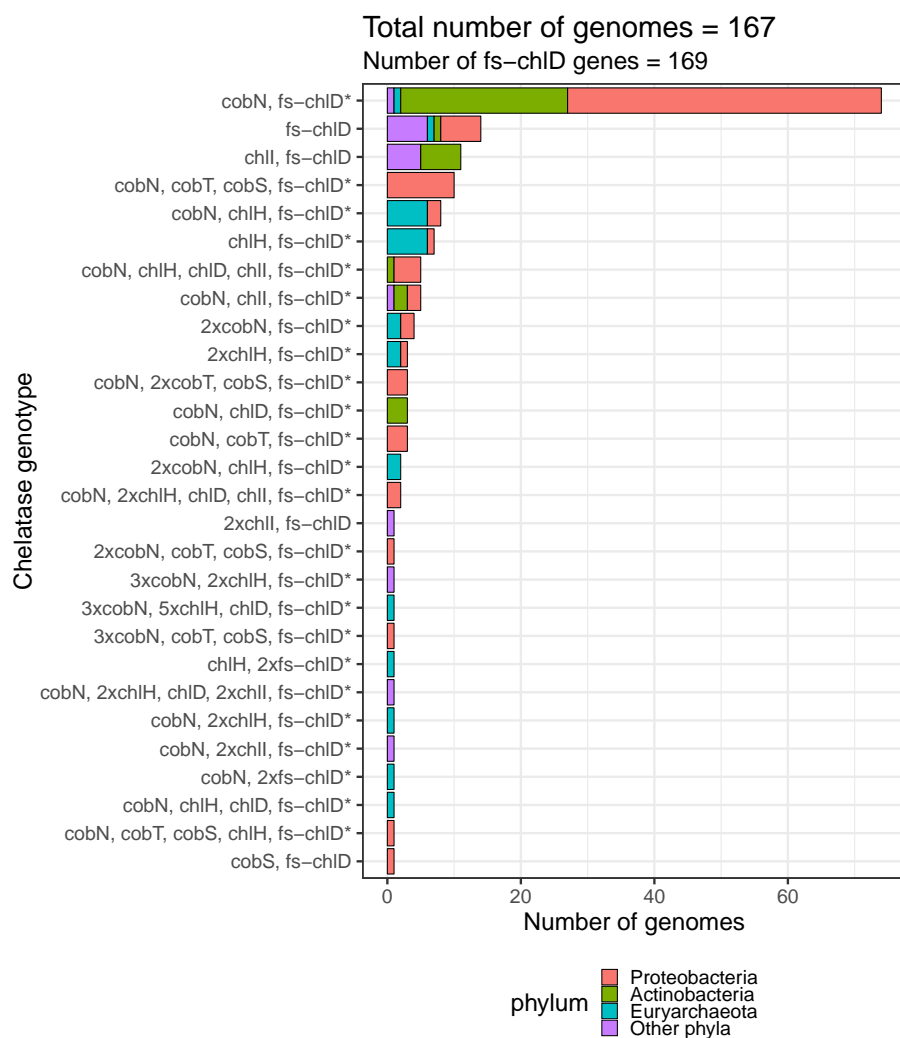
Supplementary Figure 3: Schematic locations of the genes encoding the Mg-/Co-chelatase subunits in the genomes of *D. acidovorans* SPH-1, *M. fervens* AG86 and *M. sp.* FS406-22. The frameshifts in the *chlD* genes are indicated by the red bars. The corresponding GenBank IDs are shown above each gene. The fs-*chlD* genes have IDs because they are currently annotated as two separate genes (before and after the frameshift). In all genomes the frameshifted *chlD* genes are co-localized with the *cobN* genes. The *cobS* and *cobT* genes are absent in the two *Methanocaldococcus* genomes and in the *Delftia acidovorans* they are located away from the *cobN* gene.



Supplementary Figure 4: The number of the identified chlD genes in the genomes from different prokaryotic phyla. In each group the fraction of the chlD genes with predicted -1 and +1 frameshift is indicated by the red and green colors, respectively.



Supplementary Figure 5: The most frequent "chelataase genotypes" among the 1207 genomes containing chlD genes. The chelataase genotype indicates the presence of the genes encoding the magnesium (chlH, chlD, chlI) or cobalt (cobN, cobT, cobS) chelataase subunits. The frameshifted chlD gene is denoted as 'fs-chlD'. The '2x' prefix is added to a gene name if genome contains two copies of the same gene.



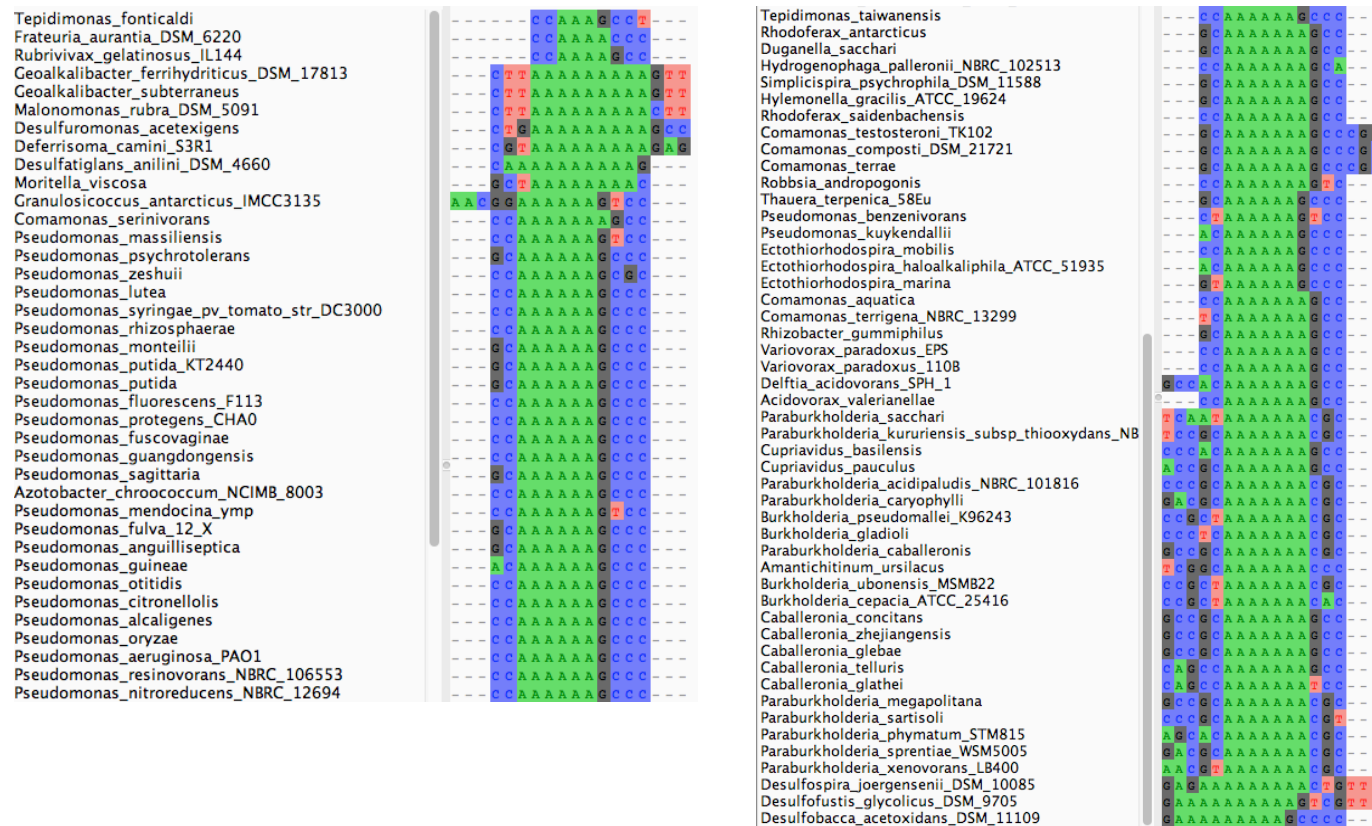
Supplementary Figure 6: The 166 genomes with frameshifted chlD genes grouped by the chelatase genotypes. The genotypes with at least one large chelatase subunit gene (chlH, bchH or cobN) are marked by '\*'.

|                         |                           | Small subunit gene is absent in the genome |                           |
|-------------------------|---------------------------|--|---------------------------|
|                         |                           | YES<br>(reduced genotype)                  | NO<br>(complete genotype) |
| Frameshift in chlD gene | YES<br>(fs-chlD genotype) | 53<br>genomes                              | 2<br>genomes              |
|                         | NO<br>(chlD genotype)     | 48<br>genomes                              | 31<br>genomes             |

**Fisher's Exact Test p-value = 1.036e-06**

Supplementary Figure 7: The fs-chlD genes are enriched in the genomes with the reduced genotypes. To avoid ambiguous cases the 134 proteobacterial genomes with exactly one medium subunit gene and at least one large subunit gene were considered only.

∞



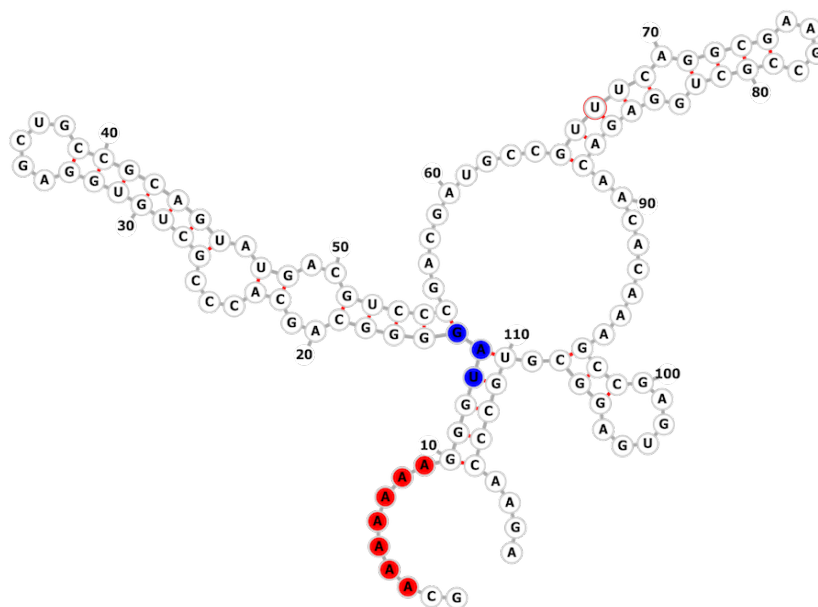
Supplementary Figure 8: The codon alignment of the putative slippery sites in the fs-chlD genes from Proteobacteria. The conserved A.AAA.AA slippery site upstream of the premature stop codon is present in 84 out of the 87 sequences.



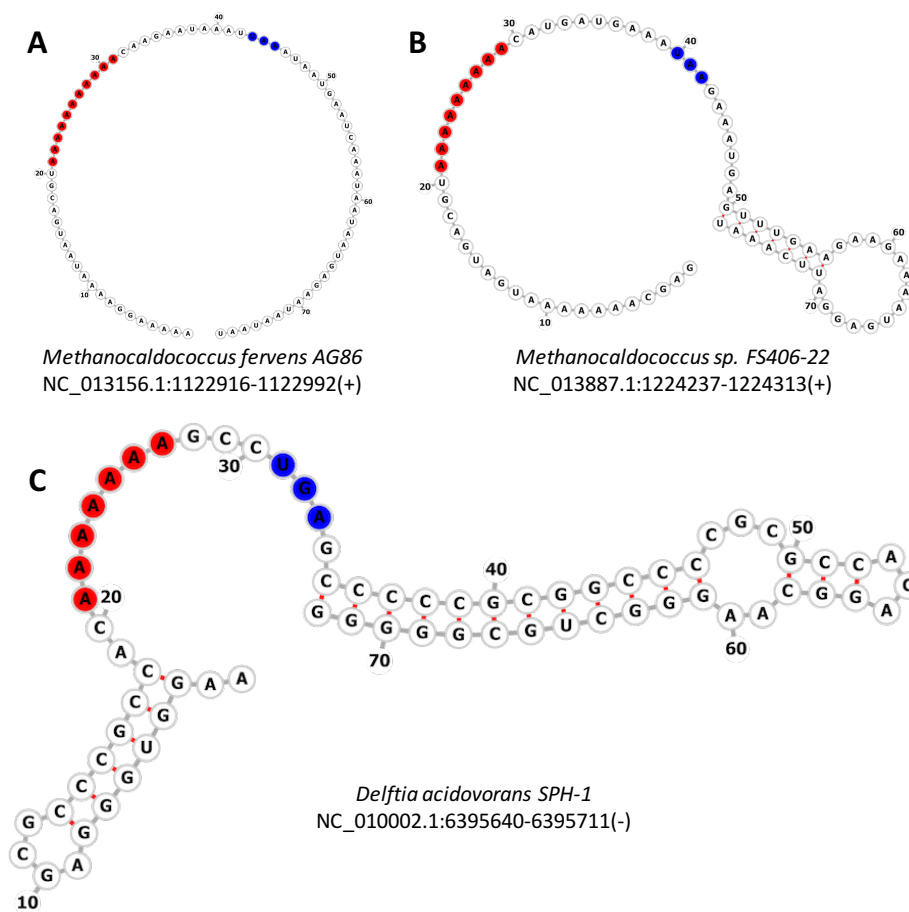
|                            |     | chID gene from the reduced genotype |                   |
|----------------------------|-----|-------------------------------------|-------------------|
|                            |     | YES                                 | NO                |
| Slippery site in chID gene | YES | 60<br>chID genes                    | 2<br>chID genes   |
|                            | NO  | 41<br>chID genes                    | 110<br>chID genes |

**Fisher's Exact Test p-value = 1.1e-22**

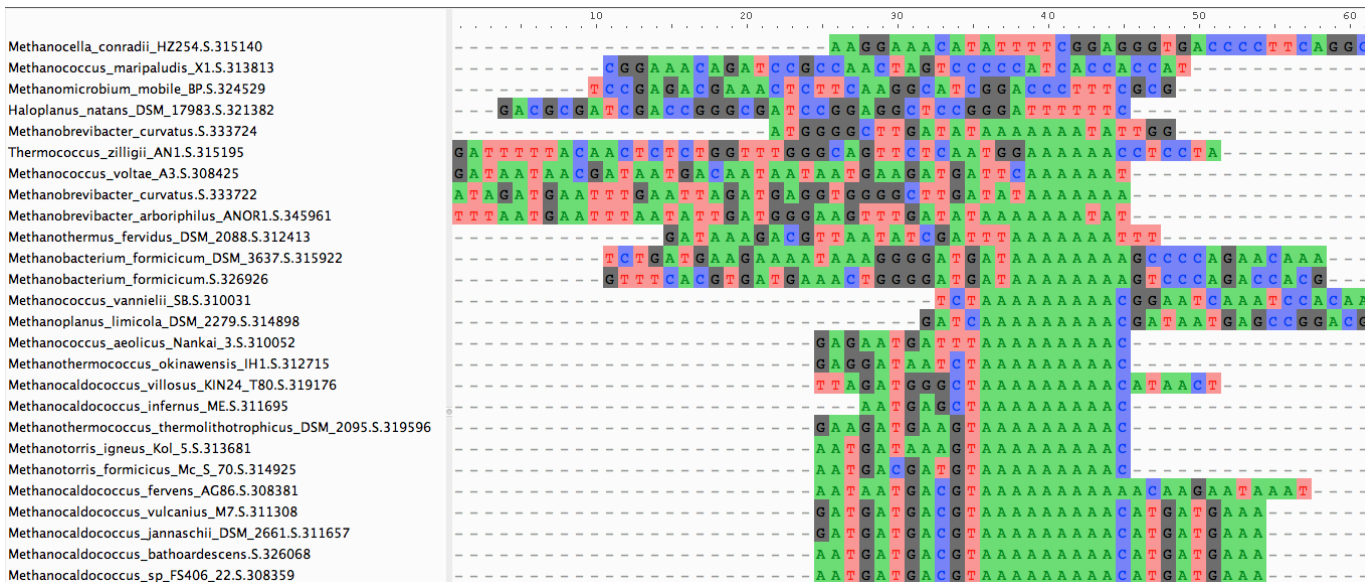
Supplementary Figure 9: The putative slippery sites are enriched in the proteobacterial chID genes from the reduced genotypes. To avoid ambiguous cases the reduced genotypes contained exactly one chID gene (frameshifted or not), at least one large subunit gene and zero small subunit genes. The complete genotypes were '1xcobN, 1xchlD, 1xchlI', '1xchlH, 1xchlD, 1xchlI' and '1xcobN, 1xcobT, 1xcobS, 1xchlH, 1xchlD, 1xchlI'. The slippery site patterns were [TA]\_AAA\_AA (underscores indicate the original reading frame) and [TA]{10,} (i.e. 10 or more consecutive T's or A's). The chID regions between 900-1350 bp were analyzed.



Supplementary Figure 10: The putative -1 frameshifting signal in the chlD gene from *Slackia heliotrinireducens*. Genomic coordinates are NC\_013165.1:1584991-1584874(-).



Supplementary Figure 11: The predicted RNA secondary structures of the sequences from Supplementary Table 1. The minimal free energy (MFE) folding was performed by the RNAfold web server. The putative poly-A slippery sites and in-frame stop codons are in red and blue colors, respectively.



Supplementary Figure 12: The alignment of the putative slippery sites in the archaeal fs-chlD genes.

## 2 Supplementary tables

Supplementary Table 1: The -1 frameshifting signals that have been validated in the E.coli model system. The underscores in the sequences mark the main (i.e. before the frameshifting) reading frame; the putative poly-A slippery sites are in red color; the in-frame stop codons that can be bypassed by the -1 frameshifting are in blue color.

| Species (kingdom, phylum)                                   | Genomic location (strand)      | Frameshifting signal sequence  | Length | Efficiency |
|---|--------------------------------|--|--------|------------|
| Methanocaldococcus fervens AG86<br>(Archaea, Euryarchaeota) | NC_013156.1:1122916-1122992(+) | AAA_AAG_GAA_AAT_AAT_GAC_GTA_AAA_AAA_AAA_ACA_AGA_ATA_AAT_<br>TAA_ATA_ATG_AAT_CAA_ATA_ATA_ATG_AGA_ATA_ATA_AT | 77 bp  | 63%        |
| Methanocaldococcus sp. FS406-22<br>(Archaea, Euryarchaeota) | NC_013887.1:1224237-1224313(+) | GAG_CAA_AAA_AAT_GAT_GAC_GTA_AAA_AAA_AAC_ATG_ATG_AAA_TAA_<br>GAA_ATG_AGT_TTG_AAG_AAG_AAA_ATG_AGG_ATT_CAA_AT | 77 bp  | 40%        |
| Delftia acidovorans SPH-1<br>(Bacteria; Betaproteobacteria) | NC_010002.1:6395640-6395711(-) | AAG_GTG_GGA_GCG_CCC_GCC_ACA_AAA_AAA_GCC_TGA_GCC_CCC_GCG_<br>GCC_CCG_CGC_CAC_AGG_CAA_GGG_CTG_CGG_GGG        | 72 bp  | 10%        |



| Domain       | Phylum         | Total number of genomes | Genomes with chlD genes | Number of chlD genes | Genomes with fs-chlD genes | Number of fs-chlD genes |
|--------------|----------------|-------------------------|-------------------------|----------------------|----------------------------|-------------------------|
| Archaea      | Euryarchaeota  | 220                     | 134                     | 150                  | 25                         | 27                      |
|              | Crenarchaeota  | 24                      | 3                       | 3                    | 0                          | 0                       |
|              | Thaumarchaeota | 2                       | 0                       | 0                    | 0                          | 0                       |
| Bacteria     | Proteobacteria | 2027                    | 321                     | 337                  | 87                         | 87                      |
|              | Actinobacteria | 1024                    | 540                     | 572                  | 38                         | 38                      |
|              | Chloroflexi    | 30                      | 17                      | 21                   | 13                         | 13                      |
|              | Spirochaetes   | 67                      | 4                       | 4                    | 2                          | 2                       |
|              | Firmicutes     | 1215                    | 55                      | 61                   | 1                          | 1                       |
|              | Bacteroidetes  | 569                     | 19                      | 27                   | 1                          | 1                       |
|              | Cyanobacteria  | 77                      | 76                      | 78                   | 0                          | 0                       |
|              | Other          | 361                     | 38                      | 47                   | 0                          | 0                       |
| <b>TOTAL</b> |                | <b>5616</b>             | <b>1207</b>             | <b>1300</b>          | <b>167</b>                 | <b>169</b>              |

Supplementary Table 3: The statistics on the number of identified chlD and fs-chlD genes in the genomes from different prokaryotic phyla.

Supplementary Table 4: The putative "in-frame" slippery sites in the chlD genes from the Proteobacteria with the "1 x cobN, 1 x chlD" genotype.

| # | Species (sequence ID)                                 | Putative slippery site                          |
|---|---|---|
| 1 | Marinagarivorans algicola (NZ_LGAK01000029.1)         | TTA_GCC_CCT_CGT_TTT_AAA_AAA_ACA_GTG_CCC_ATT     |
| 2 | Marinomonas mediterranea MMB-1 (NC_015276.1)          | GGA_CGT_ACC_GTT_TTA_AAT_ATT_ATT_ATG_CTC_GAC_ACC |
| 3 | Marinomonas posidonica IVIA-Po-181 (NC_015559.1)      | GTT_GAT_GCA_CCT_AAA_AAA_AGT_CTG_AGA             |
| 4 | Marinomonas spartinae (NZ_FLOB01000019.1)             | CGA_AGT_CTG_ACA_AAA_AAT_CCG_CAA_ACC             |
| 5 | Marinomonas ushuaiensis DSM 15871 (NZ_JAMB01000006.1) | GAA_TGG_CCG_CCA_AAA_AAA_CTT_GCT_CAT             |
| 6 | Neptuniibacter caesariensis (NZ_CH724125.1)           | GCA_TTA_CAG_TTT_AAA_AAG_CAG_GGT_AAA             |
| 7 | Psychromonas arctica DSM 14288 (NZ_KI519487.1)        | GAA_TTA_CCA_ACA_TTA_AAA_AAA_AGT_ACA_TTA_AGT     |
| 8 | Psychromonas hadalis ATCC BAA-638 (NZ_ATUO01000011.1) | AGT_CAT_AAT_GAA_AAT_AAT_AAT_GAA_CAA_ACG         |
| 9 | Zymobacter palmae DSM 10491 (NZ_KK211056.1)           | CCT_CAG_GGA_CTT_AAA_AAA_GTG_CTG_CGC             |



Supplementary Table 5: The N- and C-terminal parts (i.e. before and after the predicted frameshift location) of the translation of the *chlD* genes with validated frameshifting signals. The amino acids corresponding to the poly-A slippery sites are highlighted by the red color.

| Species (sequence id)                         | Translation of the <i>chlD</i> gene (with frameshift correction)  |  |
|---|---|--|
|   | N-terminal part   | C-terminal part  |
| Methanocaldococcus fervens AG86 (NC_013156.1) | MQYIYPFTAIVGQEKMKKALILNAINPKIGGVLIRGEKGTAKSTAVRALADLLP<br>EIEVVEGCSFNCDPNGLCDICKEKKDKGGLKITKKMKVNLPIGATEDRVIG<br>TLDIEKAIKEGIKALEPGILAEANRNILYIDEVNLLDDHIVDVLDAAMGWNI<br>IEREGVKIKHPSRFILVGTMTNPEEGELRPQILDRFGLMVDVEGLTDIKDRVEVI<br>KRVEEFNNNPEEFYKKFEEQNKLRKI IKAREILKDVEISDELLEFISKVCIE<br>LGIQTNRADITVVRTAKALAAAYNGRTKVTLDVKEAMELALPHMRMRKPFEPPO<br>LNKEKLEQMINEFEQQINKNEEEKEKEKKENNNDV <b>KKK</b> TRIN          | LNNESENNEENNNPNEHENNNQNDENTGDFEQTFGIDESVKVNPCLIQFKID<br>NIHRYSSGRHIKYSYMRGRYIKFKPANDKIVDIAFDATFRRAAIHQKKREKAN<br>KKLAIYLEREDIVEKVRQRKISSHILFVVDASGSMGAMKRMEAAKGAIISLLD<br>AYQKRNKIGMIAFRKDRAELILPFTSSVELGEKLLKDLPTGGKTPLADAFIKSY<br>EVFDREIRKNPNII PIMIVISDFKPNVAVKGDYIKEVFDACEKIVEKGINVLI<br>DTEPQSFIKIGIGKEIADRFGFKYKIEELSKDKLLDICKGFVEY               |
| Methanocaldococcus sp. FS406-22 (NC_013887.1) | MQYIYPFTAIVGQEKMKKALILNAINPKIGGVLIRGEKGTAKSTAVRALADLLP<br>EIEVVEGCPFNCDPNGLCDICKEKKERGLKVTKKKMKVNLPIGATEDRVIG<br>TLDIEKAIKEGIKALEPGILAEANRNILYIDEVNLLDDHII DVLLDVAAMGWNI<br>IEREGVKIKHPSRFILVGTMTNPEEGELRPQILDRFGLMVDVEGLNDVKDRVEVI<br>KRVEEFNNNPEEFYKKFEEQKKLRKI IKAREILNKVEISDELLEFISKVCIE<br>LGIQTNRADITVVRTAKALAAAYNGRTYVTIDDVKEAMELALPHMRMRKPFEPPO<br>LNKEKLEQMINEFEQQKNNNEEEKEEQKND                           | <b>RKKK</b> HDEIRNEFEENEDSNQNNNNNSNNQNEDTFGDFERTFGIDESFKVNP<br>LIQFKLKDNIHRYSSGRHIKYSYKRGYVVKFLPKDKTSDIAFDATFRRAAIHQ<br>KRRREKSNKKLAIYLEKEDIVEKVRQRKISSHILFVVDVSGSMGAMRMEAAKG<br>AIISLLDAYQKRNKIGMIAFRKDRAELILPFTSSVELGEKLLKDLPTGGKTPL<br>ADAFIKSYEVFDRELKKNPNII PIMIVISDFKPNVAVKGDYVKEVFDACEKIAE<br>KGINVILIDTEPQSFIKIGIGREIANRFGFKYKIEELSEEKILDKSLEINF |
| Delftia acidovorans SPH-1 (NC_010002.1)       | MTASPDLGSAQPALPPLFPFAAIAQGPLLCQALLAAIDPQLGGVLIIEGPRGT<br>AKSTAARALAEIIDAAPFVTPLPGASLEHLAGSLDLQAMAGHALKFAPGLLAR<br>AHGGVLYVDEINLLPDALVDVLLDAAASGVNVVERDGISHRHAARFVLVGTMTNP<br>EEGQLRPQLLDRLGLCVRLANVQDPAQRQAIVRARLLFDADPAAFRARHADAQA<br>ALAAQLRTARERLQAGALPWSDAVLQAAGALCIAAQVDGLRADVLVLLRAARAW<br>AAWLGDAEVTEHVQVAELVLVHRRKPGAPQPQPQSQPQPQPSPQAPSNPP<br>SKSPAQSRPPAESTANGAGDSPSEADWGAMPPEFVGLLRVAVPSSLAREAAKV<br>GAP | RH <b>KK</b> SLSPRGPAPOARAAGAGQQATGQPRLAQAQGGQTGSDWDPRTLAARGAQ<br>ALAPEHLQRRREPPQATRLHCFVLDCASMAGSGALARAKGVLLSLMEEAYRRR<br>EQVALICFAGTRVELSLPPKAAAWNDWIAPIGGGGTPLQAAVEQAGQLLQR<br>HCGADASCQGLWLLTDGRTREPARPAAAEAGIVDFESGRVRLGRAALAAA<br>WDARHVRADDFAG  |