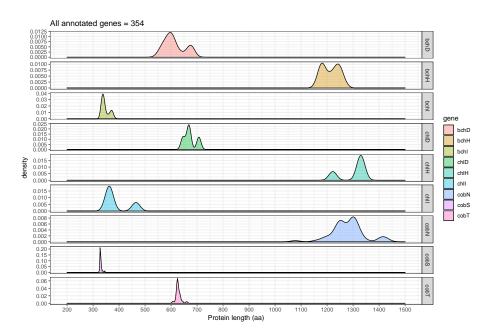
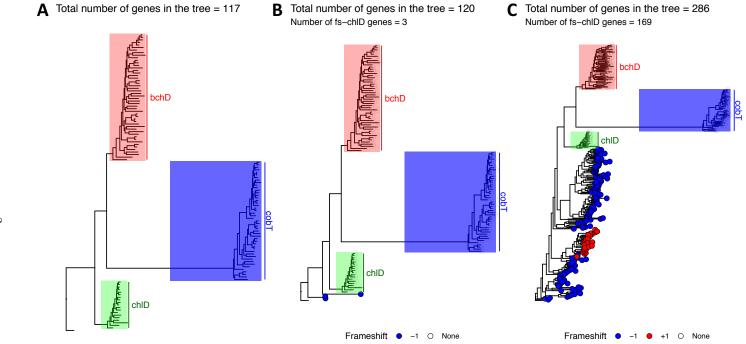
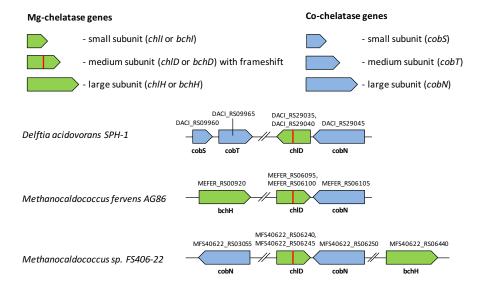
1 Supplementary figures



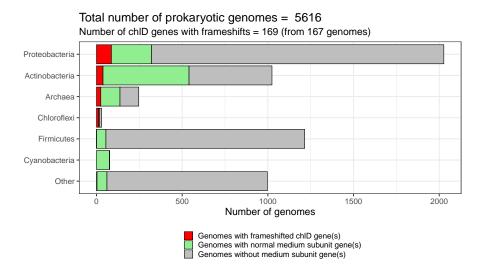
Supplementary Figure 1: Lengths of the annotated subunits of the cobNST, chlIDH and bchIDH chelatases.



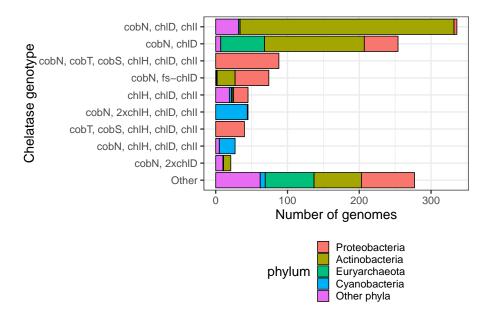
Supplementary Figure 2: The reference phylogenetic tree of the annotated medium chelatase subunit proteins and the location of the long products from the fs-chlD genes. The only annotated archaeal bchD gene (METFOR_RS04410 from Methanoregula formicica SMSP) was used as the outgroup. The -1 and +1 frameshift types in the fs-chlD genes are indicated by the blue and red colors, respectively.



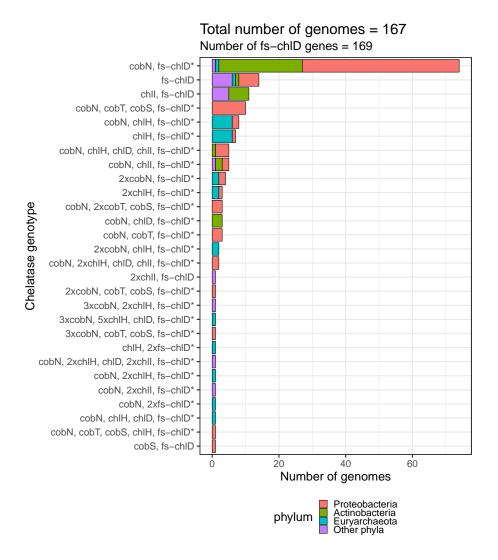
Supplementary Figure 3: Schematic locations of the genes encoding the Mg-/Co-chelatase subunits in the genomes of *D. acidovorans SPH-1*, *M. fervens AG86* and *M. sp. FS406-22*. The frameshifts in the chlD genes are indicated by the red bars. The corresponding GenBank IDs are shown above each gene. 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 "chelatase genotypes" among the 1207 genomes containing chlD genes. The chelatase genotype indicates the presence of the genes encoding the magnesium (chlH, chlD, chlI) or cobalt (cobN, cobT, cobS) chelatase 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.

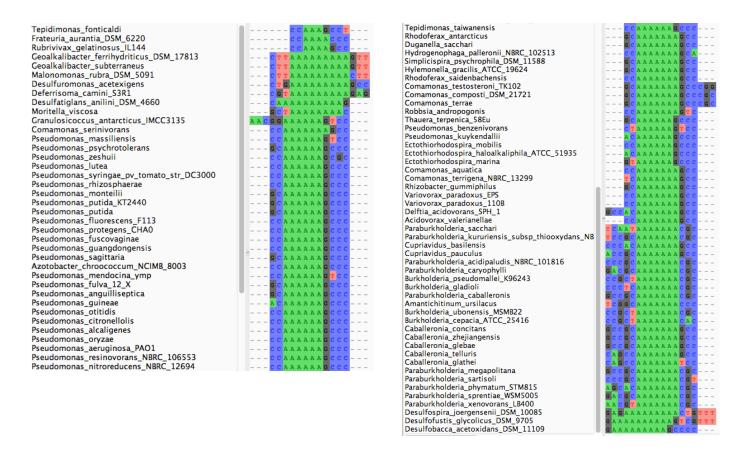


Supplementary Figure 6: The 166 genomes with frameshifted chlD genes groupped by the chelatase genotypes. The genotypes with at least one large chelatase subunit gene (chlH, bchH or cobN) are marked by '*'.

		Small subunit gene is absent in the genome		
		YES (reduced genotype)	NO (complete genotype)	
Frameshift in chID gene	YES (fs-chID genotype)	53 genomes	2 genomes	
	NO (chID genotype)	48 genomes	31 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.

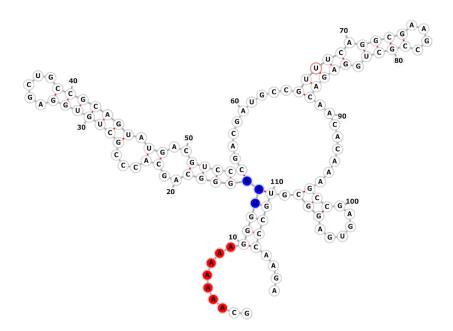


Supplementary Figure 8: The codon alignment of the putative slippery sites in the fs-chlD genes from Proteobacteria. The conserved A_AAA_AA slippery site upstream of the premature stop codon is present in 84 out of the 87 sequences.

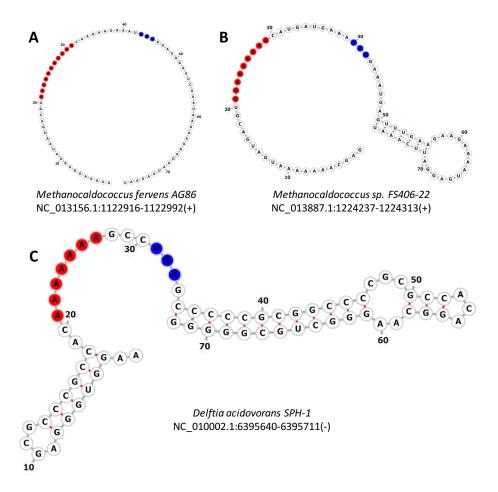
		chID gene from the reduced genotype		
		YES	NO	
Slippery site in chID gene	YES	60 chID genes	2 chID genes	
	NO	41 chID genes	110 chID genes	

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

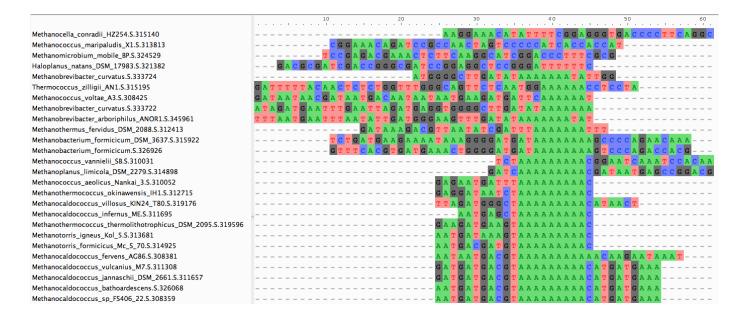
Supplementary Figure 9: The putative slippery sites are enriched in the proteobacterial chlD genes from the reduced genotypes. To avoid ambiguous cases the reduced genotypes contained exactly one chlD gene (frameshifted or not), at least one large subunit gene and zero small subunit genes. The complete 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 consequitive T's or A's). The chlD 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 (Archaea, Euryarchaeota)	NC_013156.1:1122916-1122992(+)	AAA AAG GAA AAT AAT GAC GTA AAA AAA AAA ACA AGA ATA AAT TAA ATA ATG AAT CAA ATA ATA ATG AGA ATA ATA AT	77 bp	63%
Methanocaldococcus sp. FS406-22 (Archaea, Euryarchaeota)	NC_013887.1:1224237-1224313(+)	GAG CAA AAA AAT GAT GAC GTA AAA AAA AAC ATG ATG AAA TAA GAA ATG AGT TTG AAG AAA AATG AGG ATT CAA AT	77 bp	40%
Delftia acidovorans SPH-1 (Bacteria; Betaproteobacteria)	NC_010002.1:6395640-6395711(-)	AAG_GTG_GGA_GCG_CCC_GCC_ACA_AAA_AAA_GCC_TGA_GCC_CCC_GCG GCC_CCG_CGC_CAC_AGG_CAA_GGG_CTG_CGG_GGG	72 bp	10%

Supplementary Table 2: The sequences that were used to predict the domains structures for the Figure 2 in the main text. The slippery sites and the corresponding amino acids are highlighted by the red color.

Gene/protein ID and description	Sequence	Pfam domains
MFS40622_RS06240_fs_MFS40622_RS06245 – fs-chID gene with validated frameshifting signal from <i>Methanocaldococcus sp. FS406-22</i> (2039 bp)	AND CHART ATTENTION TO A CONTROL TO GOLD COMMAND AND CHART AT TO A TO A THAT TO CANADA TEST COMMITTENT TO A COMMAND AND CHART	
MFS40622_RS06240 – the short product from the fs-chID gene (362 aa)	MQYIYPTAIVQEMKKALILMAINPKIGGVLIRGEKGTAKSTAVRALALLPELEVVEGCFPNCDPNGELCDICKERKERGELRVTKKNMKVWLPIGATEDRVIGTLDIEKSIKEGIKA LEPGILABANKRIIYIDBWLLDDHIIDVLLDVAANGWRIISBGVYKIKHFRFILVGTWAPEGGELRPQILDBRGLMVDVEGLADVKDRVEVIRRVEKPERGEKIREKI IKARBILMKYESISELLETISKUT LEIGIGTWRADITVYRATALAANNGRVYTVIDDWKSABELAPHRARRKFPERPOLNEKELERNIKKONNEKEKEEGIKMEKI	
MFS40622_RS06240_fs_MFS40622_RS06245 – the long product from the fs-chID gene (679 aa)	MQYIYPFTAIVOGEMKKALILMAINPKIGGVLIRGEKGTAKSTAVPALADLIPEIEVVEGCFPNCDPNGELCDICKENKERGELKVIKKIMKVVNLPIGATEDRVISTLDIEKAIKEGIKA LEPGILABANKNILYIDENULLODHIIDVLLUVAAMGWNIIERGVYKIKHPERFILKVTMAPEGGELRPQILDBRGIMVDVEGLADVKDRVAVIKRVEFRFEKGEKIKEKI IKAREILMKYESISELEETISKUT LEIGIGTMARDITVATARALAANNGAVYTIDONKEANEEDIAHDENERFEKGEKIKEKIL FERENDENNOMNNINNSNOMEDFOGFERTGIESFKYNPELIGKKLKKHINIKYGSGRHIRSYSKGRIVKYKLYKKUTSDIAFDATFRAAIHOKRREKEGIKALIYLEKEDIVEKKA GEKISHILTVYDVOSSIMAANBHRAAGARIISLLUDAVYRREKIGHIRSKESKELEKLIKUDPTGGKTPLADAFIKSYEVFREEKKHLATYLEKEDIVEKR DYVKEVFDACEKIAEKCINVILLOTERGSFRIGIGRELANRSFKYYKIEELSEEKILDICKSLEINF	WA-2
BAA17166.1 – Chll protein (small Mg-chelatase subunit) from <i>Synechocystis sp. PCC 6803</i> (357 aa)	MTATLAAPSKTRRVVFPFTAIVQDEMKLALLLNVIDPKIGGWIMGDRGTCKSTIFAALADLLPEIEVVANDPENSSPSDPEMSEEVRIKVDSQEPLSIVKRVTMVDLPLGATEDRVCS TIDIERALSSGVRAFEFGLAKANGGILVEVSHLODHLODHLDSAAGGANTVERRGISIMFAREVLUSSGNEEGELRGYLLDKFGMABLIKTVRSPEKEPLKVINGEKEFGNIFHFT DOYTTEGRALDRIVANGKULLPYTIDDTRKYNSEVALADLODGLAGDOTTHRAKALAARGEREVTVODISKVIVLCLHRHIKKOFELSIGGSNYKKFFRKVOVEDA	
BAA16787.1 – ChID protein (medium Mg-chelatase subunit) from Synechocystis sp. PCC 6803 (676 aa)	MTILTPFIPINFPITAIVQGEAIKLALLIGAIDPGLGGIVIAGRAGTAKSWARAIHTLIPPIEIINGNRYQCDFKNPGSWDDDTLEXFAUPPLDQLETQVIFAPFIQIPLGVTEDRILIGSV DVEKSWKGGEAVFORGLIAQAHRGVLYIDEINLILDQIANQLIFVLTGKKNJIEREGMSFQHRYQFLIATNNEEGFIRRHLIGA IAIASAGGILGLORWAAVDQVLAYADSFISFIGN YDABLIDLIKATI IARENKIKSVITEPBGYSVILVEBIAGGIQHGELFRAWAIA TAALOSGAVYAQDIAQAWAYUPRSDIMENPPSPPPPPNDGWSENDERDGDDQDDECED KONEPERAQOPSIREFFIPPBGVSLDFSUYFAQMAKGGISSGSSVIFSDDRGRYLKFILPKGKVRRAVADTLRAASFYQKSBRLRHDRQVIVEQGIRKKHLVRRAGALIVEIVA SGSMALNRWQAAKGAWQLILTEAYENDQVSLIPPGGENEVLIPPTRSIAMAKRIETIPCGGSPISHGLWQAVNVGNAKRSGDIGQVVIVAITDGGGIFLARSLGEIPEGEKFDIK ABLEITAKIRGLGQDLVINTEKKFVSTGFRELBARKAGKYQLFRATUGGISMAAQAIAMQ	WA.2

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

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_CG T_TTT_AAA_AAA_A CA_GTG_CCC_ATT
2	Marinomonas mediterranea MMB-1 (NC_015276.1)	GGA_CGT_ACC_G TT_TTA_AAT_ATT_ATT_AT G_CTC_GAC_ACC
3	Marinomonas posidonica IVIA-Po-181 (NC_015559.1)	GTT_GAT_GCA_CC T_AAA_AAA_A GT_CTG_AGA
4	Marinomonas spartinae (NZ_FLOB01000019.1)	CGA_AGT_CTG_AC A_AAA_AAT _CCG_CAA_ACC
5	Marinomonas ushuaiensis DSM 15871 (NZ_JAMB01000006.1)	GAA_TGG_CCG_CC A_AAA_AAA _CTT_GCT_CAT
6	Neptuniibacter caesariensis (NZ_CH724125.1)	GCA_TTA_CAG_ TTT_AAA_AA G_CAG_GGT_AAA
7	Psychromonas arctica DSM 14288 (NZ_KI519487.1)	GAA_TTA_CCA_AC A_TTA_AAA_AAA_A GT_ACA_TTA_AGT
8	Psychromonas hadalis ATCC BAA-638 (NZ_ATUO01000011.1)	AGT_CAT_AAT_GAA_AAT_AAT_GAA_CAA_ACG
9	Zymobacter palmae DSM 10491 (NZ_KK211056.1)	CCT_CAG_GGA_C TT_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 chID gene (with frameshift correction)				
Species (sequence iu)	N-terminal part	C-terminal part			
Methanocaldococcus fervens AG86 (NC_013156.1)	MQYIYPFTAIVGQEKMKKALILNAINPKIGGVLIRGEKGTAKSTAVRALADLLP EIEVVEGCSFNCDPNGELCDICKEKKDKGGLKITKKKMKVVNLPIGATEDRVIG TLDIEKAIKEGIKALEPGILABANRNILYIDEVNLLDDHIVDVLLDAAMGWNI IEREGVKIKHPSRFILVGTMNPEEGELRPQILDRFGLMVDVEGLTDIKDRVEVI KRVEEFNNNPEEFYKKFEEEQNKLREKIIKAREILKDVEISDELLEFISKVCIE LGIQTNRADITVVRTAKALAAYNGRTKVTLDDVKEAMELALPHRMRRKPFEPPQ LNKEKLEQMINEFEQQINKNEEEKEEKKENNDVKKKTRIN	LNNESNNNENNNPNNEHENNNQNQDENTGDFEQTFGIDESVKVNPKLIQFKIKD NIHRYSSGRHIKSYSMRGRYIKFKPANDKIVDIAFDATFRRAAIHQKKRREKAN KKLAIYLEREDIVEKVRQRKISSHILFVVDASGSMGAMKRMEAAKGAIISLLLD AYQKRNKIGMIAFRKDRAELILPFTSSVELGEKLLKDLPTGGKTPLADAFIKSY EVFDREIRKNPNIIPIMIVISDFKPNVAVKGDYIKEVFDACEKIVEKGINVVLI DTEPQSFIKIGIGKEIADRFGFKYYKIEELSKDKLLDICKGFVEY			
Methanocaldococcus sp. FS406-22 (NC_013887.1)	MQYIYPFTAIVGQEKMKKALILNAINPKIGGVLIRGEKGTAKSTAVRALADLLP EIEVVEGCPFNCDPNGELCDICKEKKERGELKVTKKKMKVVNLPIGATEDRVIG TLDIEKAIKEGIKALEPGILABANRNILYIDEVNLLDDHIIDVLLDVAAMGWNI IEREGVKIKHPSRFILVGTMNPEEGERPQILDRFGLMVDVEGLNDVKDRVEVI KRVEEFNNNPEEFYKKFEEEQKKLREKIIKARBILNKVEISDELLEFISKVCIE LGIQTNRADITVVRTAKALAAYNGRTYVTIDDVKEAMELALPHRMRRKPFEPPQ LNKEKLEQMINEFKQQKNNNEEEKEEQKNDD	RKKKHDEIRNEFEEENEDSNNQNNNNNSNNQNEDTPGDFERTFGIDESFKVNPK LIQFKLKDNIHRYGSGRHIKSYSKRGRYVKFKLPKDKTSDIAFDATFRRAAIHQ KRRREKSNKKLAIYLEKEDIVEKVRQRKISSHILFVVDVSGSMGAMRRMEAAKG AIISLLLDAYQKRNKIGMIAFRKDRAELILPFTSSVELGEKLLKDLPTGGKTPL ADAFIKSYEVFDRELRKNPNIIPIMIVISDFKPNVAVKGDYVKEVFDACEKIAE KCINVILIDTEPQSFIKIGIGREIANRFGFKYYKIEELSEEKILDICKSLEINF			
Delftia acidovorans SPH-1 (NC_010002.1)	MTASPDLGSGAQPALPPLFPFAAIAGQPLLCQALLLAAIDPQLGGVLIEGPRGT AKSTAARALAELIDAAPFVTLPLGASLEHLAGSLDLGQAMAGHALKFAPGILAR AHGGVLYVDEINLLPDALVDVLLDAAASGVNVVERDGISHRHAARFVLVGTMNP EEGQLRPQLLDRLGLCVRLANVQDFAQRQAIVRARLLFDADPAAFRARHADAQA ALAAQLRTARERLQPAGALPWSDAVLQAAGALCIAAQVDGLRADLVLLRAARAW AAWLGDAEVTFEHVQAVAELVLVHRRKPGAPQPQPQSQPQPQQPSPQQAFSNPP SKSPAQSRPPAESTANGAGDSPPSEADWGAMPPEPVGLLRVAVPSLLAREAAKV GAP	RHKKSLSPRGPAPQARAAGAGQQATGQPRLAQAQGGQTTGSWDWPRTLAARGAQ ALAPPHLQRRREPPQATRLHCFVLDCSASMAGSGALARAKGVLLSLMEEAYRRR EQVALICFAGTRVELSLPPRKAAAWNDDWIAPIGGGGTPLQAAVEQAGQLLQR HCGADASCQGWLWLLTDGRTRERPARPAAAQEAGIVDFESGRVRLGRAAALAAA WDARHVRADDFAG			