Extra question orthology and domain composition
Below are three sequences from the species Methanobacterium thermoautotrophicum, Archaeoglobus fulgidus and Synechocystis. The last of the three is phytochrome, a histidine kinase that is involved in light perception. No function information is available for the first two sequences.

```
>MTH444 (Methanotermobacter thermoautotrophicus delta H)
MEFFAVSLDS HVFSHIKEIF LYNHNQKFYV VMMKHDDVLM ILAAAALMFS FSWVHGMLLV
LPLLAIPQLS SRTLKWGVLV LAVFVALLIL QPMTRAVGLN ISDELFRLVL LLLFTFMVAL
LIERIEKVRS LRELNLELKK QAEKLEDANK ELEAFAYSVS HDLRVPLRAI DGFSRILVED
YEDKLDDEGV RILGIIRDNT RKMGQLIDDI LLLSRAGRQE MNLAMLDMRE LAESTYRELA
SQEEGRSIEF SVADLPPAMA DRALMGQVMG NLLSNAIKFT RDRDPAVIEV GYMDGGDEHT
YYVKDNGAGF DMKYASKLFG LFQRLHSQEE FEGTGVGLSI VQRIIKRHGG RVWGEGKVDG
GATIYFTLPK VVK

>AF1483 (Archaeoglobus fulgidus)
MVLEEMRIRI DISNEQNRKM LVDFLGKRYE IAEDNFDLLI IDGVTLKRKW REIERIKAES
RAFLPVLLVT TRKDLKIAEK HLWKRVDELI IEPVDKLELL ARIEILLRAR KQALQLEEHA
RIMEIELGTL FETIAHPIVV ISPEFEILHA NRYAQKIFRE QGIENAIGKK CYKVFHGREE
PAENCPCVAT FRNHKPETRE IEIFGRMYAV STTPIFIDGE LRKVVHLAFD ITDFKRMERR
LERLYKANLL LHEVERAILS ADETEEILKM TAEKLAEMLP VRGVGITVFE NGRARVVAVT
DKKMPGFREG EMIAGEDVAK VMQTLSQGKP WVKRVEGRGE GERRLMELGI KSYALIPIVS
DSLLGSINVP SEEEDAFDEE TIQILMEVAH SVALAIRSAR MREELEESEE KFRKLAEHSQ
VGIDIIQEGV FVYVNEKFAE ILGYEREELI GKSPVDFIHP DDREKFERNY RARILGEKNH
VNYRLRVLTK SGEVRIIDAY GSRVILRGKP AIVGVSVDIT EREKMRQELE KYTQELEKLV
EERTKQLAES EKRYRLLVES PIVAFWEADS NGVFRFVNDR LLEMSGYSRD EVVGKMTMFD
PIAPEQREWL AERIRLHKEH KLYGDVVEAE LVKKDGSRFH VLVSPAPIYD EKGNLVRIVG
AMIDITDRKM AEEKLKQTLE ELRKANEELE AYVHAISHDL RAPLRNLQGY VSALVEDYGE
KLEEDARFYL SRLKALTEKM DGLINDLLEY ARVSKAKAEV RRVDLNLIVE DVLDYLKDEI
RGKSAVIEIE KLPAVKGDRK LLFTVMLNLI SNAIKFVEEG VRPEVKVWAE DVNGKVRVYV
KDNGIGIPEE YHEKIFNIFE RLHGEEVYPG TGVGLAIVKK AMEVMGGRYG VRSKPGEGSI
FWIELERG

>slr0473 (Synechocystis)
MATTVQLSDQ SLRQLETLAI HTAHLIQPHG LVVVLQEPDL TISQISANCT GILGRSPEDL
LGRTLGEVFD SFQIDPIQSR LTAGQISSLN PSKLWARVMG DDFVIFDGVF HRNSDGLLVC
ELEPAYTSDN LPFLGFYHMA NAALNRLRQQ ANLRDFYDVI VEEVRRMTGF DRVMLYRFDE
NNHGDVIAED KRDDMEPYLG LHYPESDIPQ PARRLFIHNP IRVIPDVYGV AVPLTPAVNP
STNRAVDLTE SILRSAYHCH LTYLKNMGVG ASLTISLIKD GHLWGLIACH HQTPKVIPFE
LRKACEFFGR VVFSNISAQE DTETFDYRVQ LAEHEAVLLD KMTTAADFVE GLTNHPDRLL
GLTGSQGAAI CFGEKLILVG ETPDEKAVQY LLQWLENREV QDVFFTSSLS QIYPDAVNFK
SVASGLLAIP IARHNFLLWF RPEVLQTVNW GGDPNHAYEA TQEDGKIELH PRQSFDLWKE
IVRLQSLPWQ SVEIQSALAL KKAIVNLILR QAEELAQLAR NLERSNADLK KFAYIASHDL
QEPLNQVSNY VQLLEMRYSE ALDEDAKDFI DFAVTGVSLM QTLIDDILTY AKVDTQYAQL
TFTDVQEVVD KALANLKQRI EESGAEIEVG SMPAVMADQI QLMQVFQNLI ANGIKFAGDK
SPKIKIWGDR QEDAWVFAVQ DNGIGIDPQF FERIFVIFQR LHTRDEYKGT GMGLAICKKI
IEGHQGQIWL ESNPGEGSTF YFSIPIGN
```

Are the sequences orthologs of each other? Approach this question using the following criteria:

1. Are they consistent bi-directional "best" (e.g. in terms of Blast scores) hits in the three genomes? (Use Blastp and search specifically per genome.)

2. Is there conservation of gene order? i.e. are their neighbors also Best-Bi-directional Best Hits (You can examine the local gene context of a gene with the microbial genome viewer http://www.cmbi.ru.nl/MGV. You can examine whether the neighboring genes are orthologs by blast searches and best-bidirectional hits approaches. I am particularly interested in neighboring proteins with a CheY-like receiver domain that functions in the same pathway as histidine kinases and that is often in the same operon as the histidine kinase) **Alternatively,** you can visualize the local genomic context via the "GENE ID" entry from your Blast output.
3. Do they form a monophyletic branch in a phylogenetic tree? In order to establish this you will have to include some other, similar sequences from at least three other species (e.g. the most similar ones that you can detect when you Blast against other species **and (at least) one extra homologous sequence from the species themselves (***M. thermo*, *A.fulgidus*, *Synechocystis*) into a sequence alignment, and derive a phylogeny. You can run ClustalW at the EBI site (www.ebi.ac.uk/clustalw) to obtain a multiple sequence alignment and get a tree. Notice that you first have to make the alignment and then paste it in the phylogeny tool. Put the tree type on "nj" for neighbor joining and the "Correct dist" on "on" to correct for multiple substitutions. Once you have a phylogeny (at the bottom of the results page) click on "Show Phylogram Tree" to get the best "representation" of the tree.
4. Is the domain composition of these genes conserved? How about the domain composition other proteins that in the tree are close to these three proteins (form a monophyletic group with these three proteins)? Can you indicate where in evolution the changes in the domain architecture of this protein family have occurred? Plot the tree on paper, and indicate where the events occurred. (Check a domain prediction server for the domain composition of the proteins in the tree).
5. Histidine Kinases are signaling proteins. The N-terminal domains of Histidine Kinases determine on which signal they react. To what extent can you regard the function of these histidine kinases conserved?