Evolution et biodiversité des microorganismes : travaux dirigés

L'évolution des protéines universelles Sua5/TsaC : quand la simplicité est l'ultime sophistication ?

Objectifs : Savoir chercher les séquences à partir de la base de données NCBI. Savoir utiliser les logiciels d'alignement de séquences. Savoir construire les arbres phylogénétiques et les interpréter.

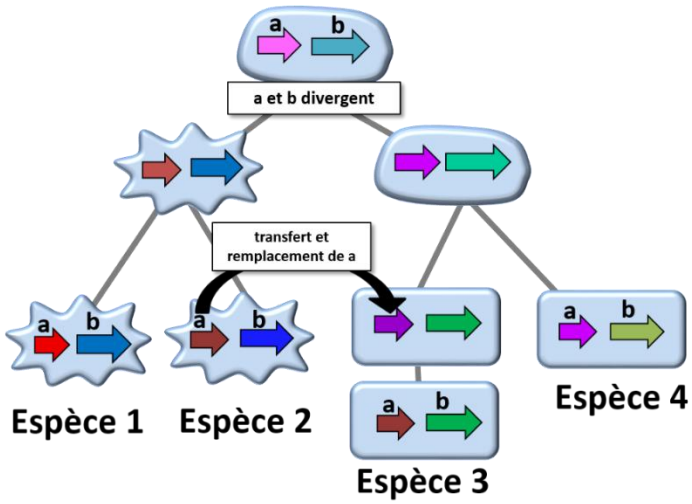
Consignes : Donnez vos réponses en couleur pour mieux les repérer, utilisez la police Courier New (taille de lettres constante) pour écrire les séquences.

PARTIE II

Le transfert horizontal de gènes (HGT) a été découvert grâce à l'étude des microorganismes à la fin des années 1940. Depuis, l'amélioration des méthodes de détection d'HGT a mis en évidence l'étendue surprenante de l'HGT dans l'évolution du contenu génomique chez les virus, les procaryotes et les eucaryotes (Soucy et al., 2015, Nature Reviews Genetics). Par exemple, de nombreuses duplications de gènes sont maintenant connus pour être le résultat de l'HGT, et non pas de la duplication des gènes autochtones.

Dans cette deuxième partie du TD nous allons tenter d'identifier l'HGT au sein de gènes codant pour la famille Sua5/TsaC. En effet, la mobilité des gènes entre espèces pourrait expliquer en partie la distribution étonnante de Sua5 et de TsaC au sein du vivant.

1. Les gènes « voyageurs » qui ont subi un HGT sont généralement éliminés lors des analyses phylogénétiques dont l'objectif est d'inférer la relation de parenté entre différents organismes. Expliquez pourquoi à l'aide du schéma ci-dessous.



Sur le schéma on observe un transfert du gène a de l'espèce 2 à l'espèce 3 suivi par la perte du gène a endogène. Cela implique que les gènes a des espèces 2 et 3 sont plus proches que les gènes a des espèces 3 et 4. En effet l'arbre fait avec les séquences des gènes a va regrouper les espèces 2 et 3 en groupe sœur ; cet arbre représente l'histoire évolutive du gène a mais ne représente pas l'histoire évolutive des espèces qui le portent.

2. La question précédente nous montre qu'un HGT peut être identifié si la topologie de l'arbre d'un gène ne correspond pas à celle des espèces porteuses de ce gène. Nous avons vu également qu'avec le temps le gène endogène peut être remplacé par le gène acquis par HGT. Cependant, quand le transfert est relativement récent les deux copies de gène (endogène et exogène) peuvent co-exister. L'analyse de distribution de Sua5 et TsaC a ainsi mis en évidence la présence de gènes codant pour Sua5 et pour TsaC chez les champignons du genre Aspergillus. Or, l'analyse de distribution de *tsaC* et *sua5* au sein du vivant montre que tous les autres champignons codent seulement pour la Sua5. Cela indique que le gène codant pour la TsaC a été introduit par HGT chez le genre Aspergillus. Votre mission est de tester cette hypothèse grâce à l'analyse phylogénétique.

Voici les séquences protéiques pour le Sua5 (UniProt ID Q4X1Q3) et TsaC (Uniprot ID A4D9M1) *d'Aspergillus fumigatus* Z5.

```
>tr|Q4X1Q3|Q4X1Q3_ASPFU Threonylcarbamoyl-AMP synthase OS=Aspergillus
fumigatus (strain ATCC MYA-4609 / CBS 101355 / FGSC A1100 / Af293)
OX=330879 GN=AFUA_2G09160 PE=3 SV=1
MTPRETRILSVRRLNKEDPGTRSLAEWWASERSQKTPESSAIEEAAHLLRTSDIPVAFPT
ETVYGLGADATRSAAVQGIYKAKQRPSDNPLIVHIDSIEMLERLLNPASSSPTRTTRTAK
NTIPAIYQPLIERFWPGPLTILLHNPSGSLLADEVTANLTTFGVRMPASPLARLLIHVAD
RPLAAPSANASTKPSPTAAEHVYHDLEGRINLILDGGPCGVGVESTVVDGLSKPPAILRP
GGVGIEELRTVPGWENVQIAYHDGNLDVKEVPRAPGMKYRHYSPKARVVLFCAGSSEEEI
AKYVYKDLEDTAIRAHMIGVVRTRQWKRGLGLVSEEDIQKTLKPIPSLVDDLVCFSVPVK
GRINNREIARAAFDCHLGDNIESIARGLFSALRAMDEMEVDVIYVEGVSDSQGDLAAAVM
NRLRKAAGTVLKL
```

```
>tr|A4D9M1|A4D9M1_ASPFU Threonylcarbamoyl-AMP synthase OS=Aspergillus
fumigatus (strain ATCC MYA-4609 / CBS 101355 / FGSC A1100 / Af293)
OX=330879 GN=AFUA_4G13765 PE=3 SV=1
MQTTIDIPTDAARVFSILAQGGIGIVPSSVGYGIIATEPPALQRIYTVKRRQPHKRHAII
GSYALHREIHVLPPDKMALVRWLTVDLNLPLGVIARYRRDHPLLARLDEETRAASSMDGT
MAMLVNGGPFQEELVRVAAAAGRAVLGSSANLTGQGTKTVVEAIEEEIREAADIVVDYGR
VRDGWPRASSTMVDFEAMRVVRVGACYEAIHDVVKRFAGLNWPDPSV
```

3. Vérifiez par alignement multiple de séquences que AfSua5 et AfTsaC contiennent les motifs conservés (P^{227}-G^{228}-M^{229} et H^{234}-Y^{235} et K^{56}-X-R^{58}/S^{143}-X-N^{145}). Utilisez la séquence de Sua5 de *P. abyssi* (ci-dessous) comme référence. Copiez-collez le résultat d'alignement et soulignez les motifs trouvés en jaune. Que pouvez-vous conclure ?

```
>WP_010868714.1 threonylcarbamoyl-AMP synthase [Pyrococcus abyssi]
MTIIINVRERIEEWKIRIAAGFIREGKLVAFPTETVYGLGANALDENAVKRIFEAKGRPADNPLIIHIAS
FEQLEVLAKEIPEEAEMLAKRFWPGPLTLVLPKSEVVPRVITGGLDTVAVRMPAHEIALKLIELSERPIA
APSANISGKPSPTSAHHVAEDFYGKIECIIDGGETRIGVESTVIDLTEWPPVLLRPGGLPLEEIEKVIGE
IRIHPAVYGKSVDTAKAPGMKYRHYAPSAEVIVVEGPRDKVRRKIEELIAKFKEEGKKVGVIGSGSYDAD
EVFYLGDTVEEIARNLFKALRHMDRTGVDVILAEGVEEKGLGLAVMNRLRKASGYRIIKV
```

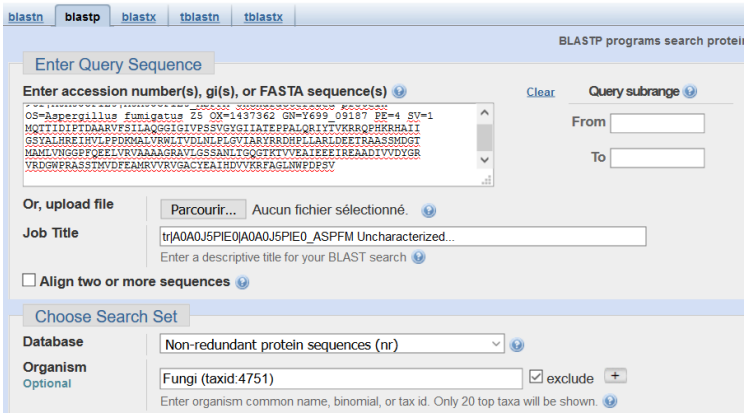
```
tr|A0A0J5PIE0|A0A0J5PIE0_ASPFM      ----------------------------MQTTIDIPTDA--ARVFSIL-AQGGIGIVPS  28
WP_010868714.1                      ----MTIIINV---------RERIEEWK-------------IRIAAGFI-REGKLVAFPT  33
tr|A0A0J5Q2G4|A0A0J5Q2G4_ASPFM      MTPRETRILSVRRLNKEDPGTRSLAEWWASERSQKTPESSAIEEAAHLLRTSDIPVAFPT  60
                                                                              .  ::  ..    .*:

tr|A0A0J5PIE0|A0A0J5PIE0_ASPFM      --SVGYGIIATEPPALQRIYTVKRRQPHKRHAI-IGSYALHREIHVLPPDKMALVRWL-T  84
WP_010868714.1                      ETVYGLGANALDENAVKRIFEAKGRPADNPLIIHIASFEQLE--VLAK-----------  79
tr|A0A0J5Q2G4|A0A0J5Q2G4_ASPFM      ETVYGLGADATRSAAVQGIYKAKQRPSDNPLIVHIDSIEMLE--RLLNPASSSPTRTTRT  118
                                        * *  *    *:: *: .* *  .:   : * *    .   :

tr|A0A0J5PIE0|A0A0J5PIE0_ASPFM      VDLNLPL---GVIARYRRDHPLLARLDEE--TRAASSMDG---TMAMLVNGGPFQEELVR  136
WP_010868714.1                      ---EIPEEAEMLAKRFW-PGPLTLVLP--KSEVVPRVITGGLDTVAVRMPAHEIALKLI-  132
tr|A0A0J5Q2G4|A0A0J5Q2G4_ASPFM      AKNTIPAIYQPLIERFW-PGPLTILLHNPSGSLLADEVTANLTTFGVRMPASPLARLLI-  176
                                        :*     :  *:    **   *           : .   *..: : .  :   *:

tr|A0A0J5PIE0|A0A0J5PIE0_ASPFM      VAAAAGRAVLGSSANLTGQGTKTVVEAIEEEIREAADIVVDYGRVRDGWPRASSTMVDFE  196
WP_010868714.1                      --ELSERPIAAPSANISGKPSPTSAHHVAEDFYGKIECIIDGGETRIGVESTVIDLTEWP  190
tr|A0A0J5Q2G4|A0A0J5Q2G4_ASPFM      --HVADRPLAAPSANASTKPSPTAAEHVYHDLEGRINLILDGGPCGVGVESTVVDGLSKP  234
                                        : * : . *** : : : * .. : .::    : ::* *    *   :     .

tr|A0A0J5PIE0|A0A0J5PIE0_ASPFM      AMRVVRVGACYEAIHDV-------------------VKRFAGLNWPDPSV----------  227
WP_010868714.1                      PVLLRPGGLPLEEIEKVIGEIRIHPAVYGKS---VDTAKAPGMKYRHYAPSAEVIVVEGP  247
tr|A0A0J5Q2G4|A0A0J5Q2G4_ASPFM      PAILRPGGVGIEELRTVPGWENVQIAYHDGNLDVKEVPRAPGMKYRHYSPKARVVLFCAG  294
                                       :   *   * :. *                  . :  *::: . :

tr|A0A0J5PIE0|A0A0J5PIE0_ASPFM      ------------------------------------------------------------  227
WP_010868714.1                      RDKV--R----RKIEEL------IAKFK--EEGKKVGVIG--------------------  273
tr|A0A0J5Q2G4|A0A0J5Q2G4_ASPFM      SSEEEIAKYVYKDLEDTAIRAHMIGVVRTRQWKRGLGLVSEEDIQKTLKPIPSLVDDLVC  354

tr|A0A0J5PIE0|A0A0J5PIE0_ASPFM      ------------------------------------------------------------  227
WP_010868714.1                      -----SGS------YDADEVFYLGDTVEEIARNLFKALRHMDRTGVDVILAEGVEEK--G  320
tr|A0A0J5Q2G4|A0A0J5Q2G4_ASPFM      FSVPVKGRINNREIARAAFDCHLGDNIESIARGLFSALRAMDEMEVDVIYVEGVSDSQGD  414

tr|A0A0J5PIE0|A0A0J5PIE0_ASPFM      --------------------          227
WP_010868714.1                      LGLAVMNRLRKASGYRIIKV          340
tr|A0A0J5Q2G4|A0A0J5Q2G4_ASPFM      LAAAVMNRLRKAAGTVLKL-          433
```

L'alignement effectué avec Clustal Omega, montre la présence des motifs conservés (P^{227}-G^{228}-M^{229} et H^{234}-Y^{235} et K^{56}-X-R^{58}/S^{143}-X-N^{145}) dans AfSua5 et dans AfTsaC indiquant que ces protéines sont fonctionnelles.

4. Effectuez maintenant un BLASTp avec la AfTsaC contre le genre *Aspergillus*. Que trouvez-vous ? Qu'est ce que signifie ce résultat ?

On trouve l'orthologue de AfTsaC chez de nombreuses espèces d'*Aspergillus*. Cela signifie que la présence de TsaC est caractéristique pour ce groupe d'organismes. Cela suggère également que, suite au HGT, le gène TsaC a été hérité verticalement au sein de la lignée conduisant aux *Aspergillus* modernes.

5. Faites maintenant un BLASTp avec AfTsaC mais en excluant les champignons (Fungi) de votre recherche. Copiez-collez les noms des organismes donnant les 10 meilleurs résultats. De quels organismes taxonomiquement parlant s'agit-il (aidez-vous de la base de données Taxonomy du NCBI) ? Que suggère ce résultat ?



La recherche dans la NCBI Taxonomy nous apprend que les 8 premières espèces appartiennent à la classe Actinobacteria et que les 2 dernières espèces appartiennent à la classe Betaproteobacteria. Cela suggère qu'en dehors de champignons, AfTsaC semble être le plus similaire à des orthologues bactériens. On peut donc émettre l'hypothèse qu'un gène ancestral tsaC a été transféré depuis les bactéries dans l'ancêtre commun de la lignée conduisant aux Aspergillus.

6. Pour commencer les analyses nous allons récupérer la séquence protéique qui correspond au meilleur résultat du BLASTp ainsi que les séquences de TsaC issues de bactéries *E. coli*, *Thermus aquaticus* et *Prochlorococcus marinus* :

```
>WP_037952691.1 hypothetical protein [Streptomyces sp. PRh5]
MGRHDINADAKRVFDTITAGGAVILPGDIGYGAGASSPEALQRLFVAKQRAPHKRHAMVGNYELHREVHV
LGSREQEIVDAITIDADLPLTVVAAYRADHPVVAAVEPETLAASTVGNTLALLVNGGRLQDEVVRLCHQA
GLPFLGSSANLTGTGTKFRVEDIQQPILDAVDLVIDYGLRKYHHYRRSSTIIDFSTMEVVRIGTCYELIS
DIMATQFGITLPADPGRDALPSGHLREQTQ

>tr|A0A163U4K7|A0A163U4K7_PROMR Threonylcarbamoyl-AMP synthase
OS=Prochlorococcus marinus str. MIT 1342 OX=1801627 GN=ywlC PE=3 SV=1
MAVLPSAVLDASALASRLHAGSAALLPTDTLPALAAAPVHAAQLWTIKQRSADKPLILMG
ATPEELLAHVLPLALEDAWSMARRYWPGALTLVVPASGVLVEALNPGAFTLGLRVPDCGM
LRNLLKQSGPLATTSANLSGSAPTFSADAAHTCFPGLPLLGPLPWPTPSGLASTVVAWQG
AGRWHELRRGAVVLEL

>sp|P45748|TSAC_ECOLI Threonylcarbamoyl-AMP synthase OS=Escherichia coli
(strain K12) OX=83333 GN=tsaC PE=1 SV=2
MNNNLQRDAIAAAIDVLNEERVIAYPTEAVFGVGCDPDSETAVMRLLELKQRPVDKGLIL
IAANYEQLKPYIDDTMLTDVQRETIFSRWPGPVTFVFPAPATTPRWLTGRFDSLAVRVTD
HPLVVALCQAYGKPLVSTSANLSGLPPCRTVDEVRAQFGAAFPVVPGETGGRLNPSEIRD
ALTGELFRQG

>tr|A0A0M9AGB3|A0A0M9AGB3_THEAQ Threonylcarbamoyl-AMP synthase OS=Thermus
aquaticus OX=271 GN=ywlC PE=3 SV=1
```

```
MTEVYQKELAQAAEVLKKGGLVAFPTDTVWGVLARMEDEAACRRIYALKGREEKKPLQVL
VAGLEDALRLSDLGPLEERFLRLAEAFWPGALTIVVPGRGIPPWISRDGSVGLRMPAHEA
LRELLRRVGGHAAATSLNRSGEPPVRTEAEARAFPVDFVFPGEATGLASSVVDLRTGEIL
REGAIPKEALLPYLHG
```

Nous allons compléter ce jeu de séquences par celles issues d'eucaryotes. Nous allons utiliser les TsaC issues des Metazoa (animaux) car ce groupe est phylogénétiquement proche des champignons et il contient des organismes qui codent exclusivement pour le TsaC. Voici les séquences de TsaC *d'Homo sapiens*, *Mus musculus,* et *Drosophila melanogaster*.

```
>sp|Q86U90|YRDC_HUMAN YrdC domain-containing protein, mitochondrial OS=Homo
sapiens OX=9606 GN=YRDC PE=1 SV=1
MSPARRCRGMRAAVAASVGLSEGPAGSRSGRLFRPPSPAPAAPGARLLRLPGSGAVQAAS
PERAGWTEALRAAVAELRAGAVVAVPTDTLYGLACAASCSAALRAVYRLKGRSEAKPLAV
CLGRVADVYRYCRVRVPEGLLKDLLPGPVTLVMERSEELNKDLNPFTPLVGIRIPDHAFM
QDLAQMFEGPLALTSANLSSQASSLNVEEFQDLWPQLSLVIDGGQIGDGQSPECRLGSTV
VDLSVPGKFGIIRPGCALESTTAILQQKYGLLPSHASYL


>AAP37032.1 ischemia/reperfusion inducible protein [Mus musculus]
MGLSDGPASSGRGCRLLLPPEPAPALPGARLLRLPESEPVEAASPERAGWTEALRAAVAELRAGAVVAVP
TDTLYGLACSASSSAALSCVYRLKGRSEAKPLAVCLGRVADVYRYCQVRVPRELLEDLFPGPVTLVMERS
EELNKDLNPFTRLVGIRIPDHAFMLDLAQMFGGPLALTSANLSSQASSLSVEEFQDLWPHLSLVIDGGPI
GDSQSPECRLGSTVVDLSVPGKFGIIRPGCALENTTSILQQKYGLLPSQGSCS

>tr|Q8SYJ9|Q8SYJ9_DROME RE55868p OS=Drosophila melanogaster OX=7227
GN=CG10438 PE=2 SV=1
MRRLQTSLYRLLRAHHTSSRMQHQASELRTPVCAVGDEAALQLARQCLLGGQVIALPTDT
VYGLACDANNETAIQQLYEIKGRDEHKPVAICVHNIDALRRFGQAAHLSDELLTRLLPGP
LTIVIERTSQLSNRFLNPSTSKIGIRIPDFNFMRDLCAVWQEKPLALTSANRSSAPSSLQ
VSEFRSLWPQLGAVFDAGRIGLTEERRLASTVIDLATPGYYEIVRAGVALKPTLSLMEEF
GIRELKMM
```

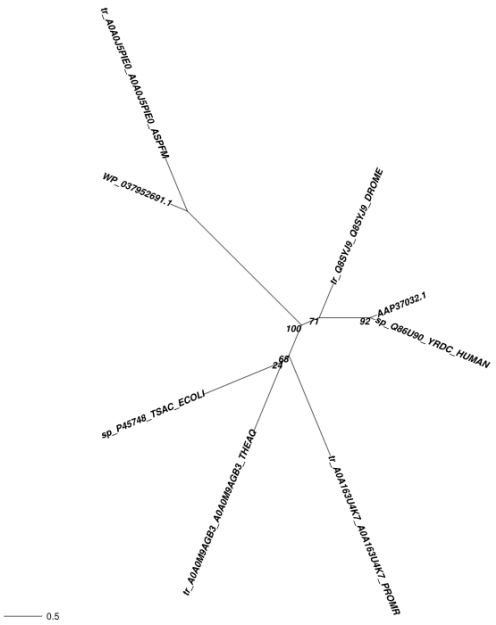
8. Alignez maintenant toutes les séquences y compris AfTsaC (donc huit séquences au total), puis construisez un arbre avec IQ-Tree. Comme précédemment, laissez le programme choisir le meilleur modèle d'évolution de séquences et demandez 1000 réplicas d'analyse ultrabootstrap. Copiez-collez l'alignement et l'arbre et commentez-les.

```
tr|A0A0J5PIE0|A0A0J5PIE0_ASPFM      ------------------------------------------------------------  0
WP_037952691.1                      ------------------------------------------------------------  0
sp|P45748|TSAC_ECOLI                ------------------------------------------------------------  0
tr|A0A0M9AGB3|A0A0M9AGB3_THEAQ      ------------------------------------------------------------  0
tr|A0A163U4K7|A0A163U4K7_PROMR       ------------------------------------------------------------MA  2
tr|Q8SYJ9|Q8SYJ9_DROME              ---------MRRLQTS------------LYRLLRAHHTSSRMQHQASELRTP-------  31
sp|Q86U90|YRDC_HUMAN                MSPARRCRGMRAAVAASVGLSEGPAGSR-SGRLFRPPSPAPA-APGARLLRLPGSGAVQA  58
AAP37032.1                          -----------------MGLSDGPASSGRGCRLLLPPEPAPA-LPGARLLRLPESEPVEA  42


tr|A0A0J5PIE0|A0A0J5PIE0_ASPFM      -MQTTIDIPTDAARVFSILAQGGIGIVPSSVGYGIIA--TEPPALQRIYTVKRRQPHKRH  57
WP_037952691.1                      --MGRHDINADAKRVFDTITAGGAVILPGDIGYGAGA--SSPEALQRLFVAKQRAPHKRH  56
sp|P45748|TSAC_ECOLI                --MNNNLQRDAIAAAIDVLNEERVIAYPTEAVFGVGCDPDSETAVMRLLELKQRPVDKGL  58
tr|A0A0M9AGB3|A0A0M9AGB3_THEAQ      ---MTEVYQKELAQAAEVLKKGGLVAFPTDTVWGVLARMEDEAACRRIYALKGREEKKPL  57
tr|A0A163U4K7|A0A163U4K7_PROMR      VLPSAVLDA---SALASRLHAGSAALLPTDTLPALAAAPV---HAAQLWTIKQRSADKPL  56
tr|Q8SYJ9|Q8SYJ9_DROME              --VCAVGDEAALQLARQCLLGGQVIALPTDTVYGLACDANNETAIQQLYEIKGRDEHKPV  89
sp|Q86U90|YRDC_HUMAN                ASPERAGWTEALRAAVAELRAGAVVAVPTDTLYGLACAASCSAALRAVYRLKGRSEAKPL  118
AAP37032.1                          ASPERAGWTEALRAAVAELRAGAVVAVPTDTLYGLACSASSSAALSCVYRLKGRSEAKPL  102
                                                  :       * .   .   .         :   * *   *


tr|A0A0J5PIE0|A0A0J5PIE0_ASPFM      AIIGS-YALH-REIHVLPPDKMALVRWLTVDLNL--PLGVIARYRRDHPLLARLDEETRA  113
WP_037952691.1                      AMVGN-YELH-REVHVLGSREQEIVDAITIDADL--PLTVVAAYRADHPVVAAVEPETLA  112
sp|P45748|TSAC_ECOLI                ILIAANYEQLKPYIDDTML---TDVQRETIFSRWPGPVTFVFPAPA--TT-------PRW  106
```

```
tr|A0A0M9AGB3|A0A0M9AGB3_THEAQ      QVLVAGLEDALRLSDLGPL---EERFLRLAEAFWPGALTIVVPGRG--I--------PPW  104
tr|A0A163U4K7|A0A163U4K7_PROMR      ILMGATPEELLAHVLPLAL----EDAWSMARRYWPGALTLVVPASG--VL-------VEA  103
tr|Q8SYJ9|Q8SYJ9_DROME             AICVHNIDALRRFGQAA------HLSDELLTRLLPGPLTIVIERTS--QL----S--NRF  135
sp|Q86U90|YRDC_HUMAN               AVCLGRVADVYRYCRV-------RVPEGLLKDLLPGPVTLVMERSE--EL-------NKD  162
AAP37032.1                         AVCLGRVADVYRYCQV-------RVPRELLEDLFPGPVTLVMERSE--EL-------NKD  146
                                    :                                 : .:

tr|A0A0J5PIE0|A0A0J5PIE0_ASPFM     ASSMDGTMAMLVNGGPFQEELVRVAAAA-GRAVLGSSANLTGQGTKTVVEAIEEEIREAA  172
WP_037952691.1                     ASTVGNTLALLVNGGRLQDEVVRLCHQA-GLPFLGSSANLTGTGTKFRVEDIQQPILDAV  171
sp|P45748|TSAC_ECOLI               LTGRFDSLAVRVTDH---PLVVALCQAY-GKPLVSTSANLSGLPPCRTVDEVRA-QFGAA  161
tr|A0A0M9AGB3|A0A0M9AGB3_THEAQ     IS-RDGSVGLRMPAH---EALRELLRRV-GGHAAATSLNRSGEPPVRTEAEARA-F-PVD  157
tr|A0A163U4K7|A0A163U4K7_PROMR     LNPGAFTLGLRVPDC---GMLRNLLK--QSGPLATTSANLSGSAPTFSADAAHT-CFPGL  157
tr|Q8SYJ9|Q8SYJ9_DROME            LNPSTSKIGIRIPDF---NFMRDLCAVWQEKPLALTSANRSSAPSSLQVSEFRS-LWPQL  191
sp|Q86U90|YRDC_HUMAN               LNPFTPLVGIRIPDH---AFMQDLAQMF-EGPLALTSANLSSQASSLNVEEFQD-LWPQL  217
AAP37032.1                         LNPFTRLVGIRIPDH---AFMLDLAQMF-GGPLALTSANLSSQASSLSVEEFQD-LWPHL  201
                                    .     :.: :       :  :          :* * :.         .

tr|A0A0J5PIE0|A0A0J5PIE0_ASPFM     DIVVDYGRVRDGW---PRASSTMVDFEAM---RVVRVGACYEAIHDVVKRFAGLNWPDPS  226
WP_037952691.1                     DLVIDYGLRKYHH---YRRSSTIIDFSTM---EVVRIGTCYELISDIMATQFGITLPADP  225
sp|P45748|TSAC_ECOLI               FPVVPGETGGRLN------PSEIRDALTG---ELFRQG----------------------  190
tr|A0A0M9AGB3|A0A0M9AGB3_THEAQ     F-VFPGEATGL--------ASSVVDLRTG---EILREGAIPKEALLPY--LHG-------  196
tr|A0A163U4K7|A0A163U4K7_PROMR     PLLGP---LP--WPTPSGLASTVVAWQGAGRWHELRRGAVVLEL----------------  196
tr|Q8SYJ9|Q8SYJ9_DROME            GAVFDAGRIGLT--EERRLASTVIDLATPGYYEIVRAGVALKPTLSLME-EFGIRELKMM  248
sp|Q86U90|YRDC_HUMAN               SLVIDGGQIGDGQSPECRLGSTVVDLSVPGKFGIIRPGCALESTTAILQQKYGLLPSHAS  277
AAP37032.1                         SLVIDGGPIGDSQSPECRLGSTVVDLSVPGKFGIIRPGCALENTTSILQQKYGLLPSQGS  261
                                     :               * :           .* *

tr|A0A0J5PIE0|A0A0J5PIE0_ASPFM     V--------------   227
WP_037952691.1                     GRDALPSGHLREQTQ   240
sp|P45748|TSAC_ECOLI               ---------------   190
tr|A0A0M9AGB3|A0A0M9AGB3_THEAQ     ---------------   196
tr|A0A163U4K7|A0A163U4K7_PROMR     ---------------   196
tr|Q8SYJ9|Q8SYJ9_DROME            ---------------   248
sp|Q86U90|YRDC_HUMAN               YL-------------   279
AAP37032.1                         CS-------------   263
```

Sur l'alignement on peut identifier le motif K^{56}-X-R^{58}/S^{143}-X-N^{145} (souligné en jaune) impliqué dans la fixation de l'ATP. Il s'agit donc probablement des TsaC fonctionnels.

Résultat d'analyse IQ-Tree

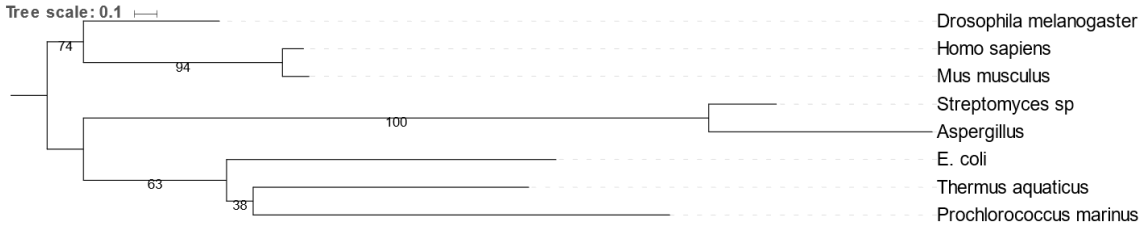


Modèle d'évolution de séquences utilisé : WAG+F+I+G4

```
(tr_A0A0J5PIE0_A0A0J5PIE0_ASPFM:0.7517048483,WP_037952691.1:0.2308037019,((
(sp_P45748_TSAC_ECOLI:1.0986000172,tr_A0A0M9AGB3_A0A0M9AGB3_THEAQ:0.9665203
323)24:0.1156960630,tr_A0A163U4K7_A0A163U4K7_PROMR:1.4094376756)68:0.428164
1237,(tr_Q8SYJ9_Q8SYJ9_DROME:0.4666681366,(sp_Q86U90_YRDC_HUMAN:0.070739394
```

3,AAP37032.1:0.0901858430)92:0.6619580960)71:0.2487546783)100:2.0983180858)
;

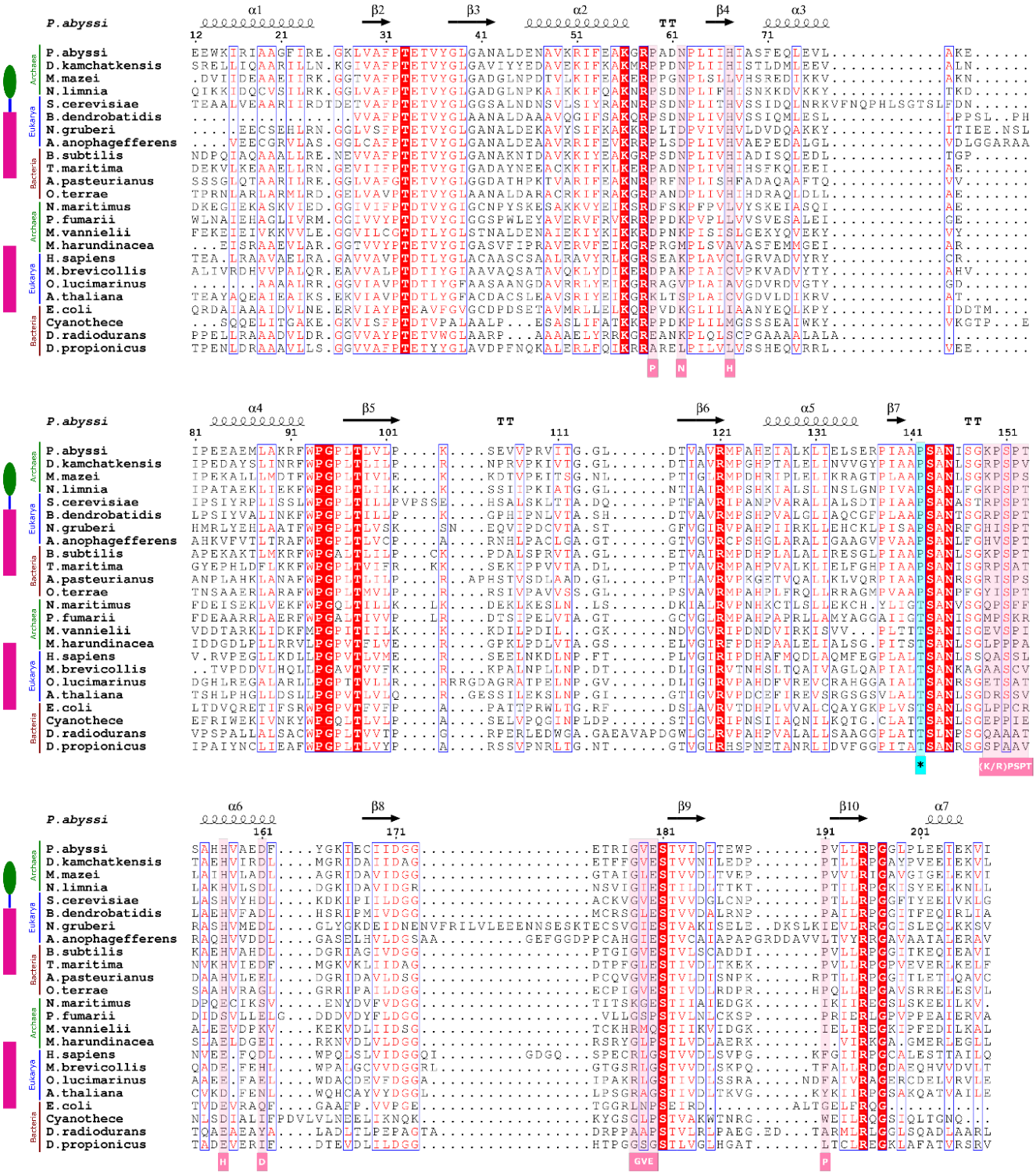
9. Editez votre arbre dans iTOL pour faire apparaitre les noms d'espèces. Quelle topologie votre arbre devrait-il avoir si AfTsaC était effectivement transféré aux *Aspergillus* à partir de Bactéries? Votre arbre correspond-il à cette prédiction ? Enracinez votre arbre, interprétez-le et concluez.



Si l'hypothèse de HGT est correcte on devrait trouver un arbre composé de deux clades : l'un comportant toutes les séquences d'eucaryotes (hormis celle d'Aspergillus) et l'autre comportant toutes les séquences de bactéries y compris celle d'Aspergillus. L'arbre ci-dessus correspond à cette topologie, cependant le clade d'eucaryotes et celui de bactéries sont faiblement soutenus (74 et 63% respectivement), de plus il n'y a aucun support pour la branche reliant le clade Aspergillus/Streptomyces avec le clade des bactéries. Le positionnement du clade Aspergillus/Streptomyces et donc incertain (on ne sait pas s'il est plus proche des bactéries ou des eucaryotes). L'analyse confirme bien cependant que les TsaC issues *d'Aspergillus* et *Streptomyces* sont les séquences les plus proches parmi celles que nous avons analysées. En conclusion, l'arbre semble bien indiquer un transfert horizontal de TsaC depuis les bactéries vers Aspergillus cependant le signal phylogénétique n'est pas assez robuste. Il faudrait faire d'autres arbres avec plus de séquences pour chaque groupe pour essayer d'obtenir un meilleur soutien pour les branches.

PARTIE III

Dans la première partie de ces travaux dirigés vous avez constaté que les séquences de Sua5 et TsaC se ségrégent en deux groupes vraisemblablement parce qu'il existe des résidus conservés spécifiques d'un ou de l'autre variant. En effet, l'analyse des séquences de TsaC et des domaines TsaC-like par alignement sur un échantillon taxonomique large montre l'existence de ces résidus (figure ci-dessous). Ainsi, on trouve 14 résidus conservés (indiqués en rectangles roses sous l'alignement) spécifiques des protéines Sua5. Un seul résidu est **conservé et spécifique des protéines TsaC**, il s'agit de la thréonine (Thr) adjacente au motif S^{143}-X-N^{145} (souligné en bleu). De manière intéressante, à la même position dans l'alignement on trouve un résidu Proline (Pro) conservé chez les protéines Sua5. Cela suggère que la présence du domaine supplémentaire chez les protéines Sua5 nécessite la présence de Pro à côté du motif S^{143}-X-N^{145}. L'inspection d'un grand nombre de séquences de la base de données UniProt confirme cette règle (données non-montrées). Mais, comme souvent en biologie, il y a des exceptions. Dans cette partie du TD, nous allons étudier l'une de ces exceptions : les Sua5 issues d'archées hyperthermophiles, les Archaeoglobales.


P.abyssi
α1 α2 α3 β2 β3 β4
12 21 31 41 51 61 71
P.abyssi EEWKIRIAAGFIRE.GKLVAFPTETVYGLGANALDENAVKRIFEAKGRPADNPLIIHIASFEQLEVL.....AKE.....
D.kamchatkensis SRELLIQAARILLN.KGIVAFPTETVYGLGAVIYYEDAVEKIFKAKMRPPDNPLIIHISTLDMLEEV.....ALN.....
M.mazei .DVIIDEAAEIIRK.GGTVAFPTETVYGLGADGLNPDTVLKIFEAKERPPGNPLSLLVHSREDIKKV.....AKN.....
N.limnia QIKKIDQCVSILRK.GGLVAFPTETVYGLGADGLNPKAIKKIFQAKNRPSDNPLIFHISNKKDVKKY.....TKD.....
S.cerevisiae TEAALVEAARIIRDTDETVAFPTETVYGLGGSALNDNSVLSIYRAKNRPSDNPLITHVSSIDQLNRKVFNQPHLSGTSLFDN.....
B.dendrobatidis .....VVAFPTETVYGLAANALDAAAVQGIFSAKQRPSDNPLIVHVSSIQMLESL.....LPPSL..PH
N.gruberi .....EECSEHLRN.GGLVSFPTETVYGLGANALDEKAVYSIFKAKKRPLTDPVIVHVLDVDQAKKY.....ITIEE.NSL
A.anophagefferens ....VEECGRVLAS.GGLCAFPTETVYGLGANALDEAAVRRIFEVKRRPLSDPLIVHVAEPEDALGL.....VDLGGARAA
B.subtilis NDPQIAQAAALLRE.NEVVAFPTETVYGLGANAKNTDAVKKIYEAKGRPSDNPLIVHIADISQLEDL.....TGP.....
T.maritima DEKVLKEAAELLRN.GEVIIFPTETVYGIGADAYNEEACKKIFKLKERPADNPLIVHIHSFKQLEEI.....AE.....
A.pasteurianus SSSGLQTAARILRE.GGLVAFGTETVYGLGGDATHPKTVARIFEAKNRPRFNPLISHFADAQAAFTQ.....VV.....
O.terrae TPRNLARLARMLRD.GELVAVPTETVYGLAANALDARACRKIFRAKGRPANDPLIVHIHDRAQLDLL.....AE.....
N.maritimus DKEGIEKASKVIED.GGIVIFPTDTVYGIGCNPYSKESAKKVYEIKSRDFSKPFPVLVYSKEIASQI.....AE.....
P.fumarii WLNAIEHAGLIVRM.GGIVVYPTDTVYGIGGSPWLEYAVERVFRVKRRPPDKPVPLLVVSVESALEI.....GE.....
M.vannielii FEKEIEIVKKVVLE.GGVILCGTDTLYGLSTNALDENAIEKIYKIKKRDPNKPISISLGEKYQVEKY.....VY.....
M.harundinacea ...EISRAAEVLAR.GGTVVYPTETVYGIGASVFIPRAVERVFEIKGRPRGMPLSVAVASFEMMGEI.....AR.....
H.sapiens TEA.LRAAVAELRA.GAVVAVPTDTLYGLACAASCSAALRAVYRLKGRSEAKPLAVCLGRVADVYRY.....CR.....
M.brevicollis ALIVRDHVVPALQR.EAVVALPTDTIYGIAAVAQSATAVQKLYDIKERDPAKPIAICVPKVADVYTY.....AHV.....
O.lucimarinus .....AAAALRR.GGVIAVPTDTIYGFAASAANGDAVRRLYEIKGRRAGVPLATAVGDVRDVGTY.....GD.....
A.thaliana TEAYAQEAIEAIKS.EKVIAVPTDTLYGFACDACSLEAVSRIYEIKGRKLTSPLAICVGDVLDIKRV.....AT.....
E.coli QRDAIAAAIDVLNE.ERVIAYPTEAVFGVGCDPDSETAVMRLLELKQRPVDKGLILIAANYEQLKPY.....IDDTM.....
Cyanothece ...SQQELITGAKE.GKVISFPTDTVPALAALP...ESASLIFATKKRPPDKPLILMGSSSEAIWKY.....VKGTP...E
D.radiodurans PPELLRAAADVLDR.GGVVAYPTETVWGLAARP...AAAAELYRRKGREANKPLQLSCPGAAAALALA.....
D.propionicus TPENLDRAAAVLLS.GGVVAFPTETYYGLAVDPFNQKALERLFQIKRRARELPILVLVSSHEQVRRL.....VEE.....
P N H
α4 α5 α6 α7 α8 β5 β6 β7 β8 β9 β10
81 91 101 111 121 131 141 151 161 171 181 191 201
P.abyssi IPEEAEMLAKRFWPGPLTLVLP....K....SEVVPRVITG.GL.....DTVAVRMPAHEIALKLIELSERPIAAPSANISGKPSPT
D.kamchatkensis IPEDAYSLINRFWPGPLTLILP....R....NPRVPKIVTG.GL.....DTIAVRMPGHPTALELINVVGYPIAAPSANLSGRPSPV
M.mazei IPEKALLLMDTFWPGPLTTVLE....K....KDTVPEITSG.NL.....PTIGTRMPDHRIPLELIKRAGTPLAAPSANLSGKPSPS
N.limnia IPATAEKLIEKFWPGPLTLILK....K....SSIIPKIATG.GL.....NTIAIRMPSHKIALSLINSLGNPIVAPSANLFGKPSPT
S.cerevisiae IPSIYRPLISSLWPGPLTILLPVPSSE....HSALSKLTTA.DQ.....PTFAVRIPANPVARALIALSDTPIAAPSANASTRPSPT
B.dendrobatidis LPSIYVALINKFWPGPLTILLP....K....GPHIPNLVTA.SH.....DTVAVRMPSHPVALGLIAQCGFPLAAPSANTSGRPSPT
N.gruberi HMRLYEHLAATFWPGPLTLVSK....SN...EQVIPDCVTA.ST.....GFVGIRVPAHPIIRKLLEHCKLPISAPSANRFGHISPT
A.anophagefferens AHKVFVTLTRAFWPGPLTLVCP....A....RNHLPACLGA.GT.....GTVGVRCPSHGLARALIGAAGVPVAAPSANLFGHVSPT
B.subtilis APEKAKTLMKRFWPGPLTILP....CK....PDALSPRVTA.GL.....ETVAIRMPDHPLALALIRESGLPIAAPSANLSGKPSPT
T.maritima GYEPHLDFLKKFWPGPLTVIFR...KK....SEKIPPVVTA.DL.....PTVAVRMPAHPVALKLIELFGHPIAAPSANISGRPSAT
A.pasteurianus ANPLAHKLANAFWPGPLTLILP....R..APHSTVSDLAAD.GL.....PTLAVRVPKGETVQALLKLVQRPIAAPSANRSGRISPS
O.terrae TNSAAERLARAFWPGPLTMVLP....R....RSIVPAVVSS.GL.....PSVAVRMPAHPLFRQLLRRAGMPVAAPSANPFGYISPT
N.maritimus FDEISEKLVEKFWPGQLTLLLK...LK....DEKLKESLN..LS.....DKIALRVPNHKCTLSLLEKCH.YLIGTSANVSGQPSFF
P.fumarii FDEAARRLAERFWPGALTVVP....LR....DTSIPELVTA.GT.....GCVGLRMPAHPAPRLLAMYAGGAIIGTSANISGMPSKR
M.vannielii VDDTARKLIDKFMPGPITLILK....K....KDILPDILL...GK.....NDVGVRIPDNDVIRKISVV...PLTTTSANISGEVSPI
M.harundinacea IDDGDLPLLRRVLPGPVTFLVE....R....GPKLPDLVTA.GS.....ELVGIRFPDHPAALELIASLG.PITSTSANLSGLPPPA
H.sapiens V.RVPEGLLKDLLPGPVTLVME....R....SEELNKDLNP.FT.....PLVGIRIPDHAFMQDLAQMFEGPLALTSANLSSQASSL
M.brevicollis ..TVPDDVLHQLLPGAVTVVFK....R....KPALNPLLNP.DT.....DLIGIRVTNHSLTQAIVAGLQAPIALTSANKAGAASCV
O.lucimarinus DGHLREGALARLLPGPTTVLLR....RRRGDAGRATPELNP.GV.....DLIGIRVPAHDFVREVCRAHGGAIALTSANRSGETSAT
A.thaliana TSHLPHGLLDSLLPGPVTLVLQ....R..GESSILEKSLNP.GI.....GTIGVRVPDCEFIREVSRGSGSVLALTSANLSGDRSSV
E.coli LTDVQRETIFSRWPGPVTFVFP....A....PAITPRWLTG.RF.....DSLAVRVTDHPLVVALCQAYGKPLVSTSANLSGLPPCR
Cyanothece EFRIWEKIVNKYWPGQLTLVLP....A....SELVPQGINPLDP.....STIGVRIPNSIIAQNILKQTG.CLATTSANLSGEPPIE
D.radiodurans VPSPALLALSACWPGPLTVVTP....G....RPERLEDWGA.GAEAVAPDGWLGLRVPAHPVALALLSAAGGPLATTSLNPSGQAAAAT
D.propionicus IPAIYNCLIEAFWPGPLTLVYP....A....RSSVPNRLTG.NT.....GTVGIRHSPNETANRLLDVFGGPITATSANRSGSPAAV
* (K/R)PSPT
P.abyssi SAHHVAEDF....YGKIECIIDGG.....ETRIGVESTVIDLTEWP.....PVLLRPGGLPLEELTEKVI
D.kamchatkensis TAEHVIRDL....MGRIDAIIDAG.....ETFFGVESTVVNILEDP.....PTLLRPGAYPVEELEKVL
M.mazei LATHVLADL....AGRIDAVIDGG.....GTAIGLESTVVDLTVEP.....PVVLRIGAVGIGELEKVI
N.limnia LAKHVLSDL....DGKIDAVIDGR.....NSVIGIESTILDLTTKT.....PTILRPGKISYEELKNLL
S.cerevisiae LASHVYHDL....KDKIPIILDGG.....ACKVGVESTVVDGLCNP.....PTLLRPGGFTYEEIVKLG
B.dendrobatidis LAEHVFADL....HSRIPMIVDGG.....MCRSGLESTVVDALRNP.....PAILRPGGITFEQIRLIA
N.gruberi RASHVMEDL....GLYGKDEIDNENVFRILVLEEENNSESKTECSVGIESTVAKISELE..DKSLKIEVLRRGGISLEQLKKSV
A.anophagefferens RAQHVVDDL....GASELHVLDGSAA.....GEFGGDPPCAHGIESTVCAIAPAPGRDDAVVLTVYRRGAVAATALERAV
B.subtilis KAEHVAHDL....DGRIAGIVDGG.....PTGIGVESTVLSCADDI.....PVLLRPGGITKEQIEAVI
T.maritima NVKHVIEDF....MGKVKLIIDAG.....DTPFGLESTIVDLTKEK.....PVLLRPGPVEVERLKELF
A.pasteurianus DAAHVLEEL....DGRIDAVLDSG.....PCQVGVESTVLDISNPK.....RPTLLRPGGITLETLQAVC
O.terrae SAAHVRAGL....GRRIPAILDGG.....ECPIGVESTIVDLRDPR.....HPQLLRPGAVSRREIESVL
N.maritimus DPQECIKSV....ENYDVFVDGG.....TITSKGESTIIAIEDGK.....IKIIREGGSLSKEEILKV.
P.fumarii DIDSVLLELG....DDDVDYFLDGG.....VLLGSPSTVLNLCKSP.....PRIERLGPVPPEAIERVA
M.vannielii ALEEVDPKV....KEKVDLIIDSG.....TCKHRMGSTIIKVIDGK.....IELIREGKIPFEDILKAL
M.harundinacea SLAELDGEI....RKNVDLVLDGG.....RSRYGLESTLVDLAERK.....VIRKGAGMERLEGLL
H.sapiens NVEEFQDLL....WPQLSLVIDGGQI.....GDGQGSPECRLGSTVVDLSVPG.....KFGIIREGCALESTIAILQ
M.brevicollis QADEFEHLL....WPALGCVVDGGRL.....GTGSRLGSTVVDLSKPQ.....TFALLRDGDAEQHVVDVLT
O.lucimarinus AAEEFAELL....WDACDEVFDGGA.....IPAKRLGSTIVDLSSRA.....NDFAIVRAGERCDELVRVLE
A.thaliana CVKDFENLL....WQHCAYVYDGGL.....LPSGRAGSTIVDLTKVG.....KYKIIREGSAKQATVAILE
E.coli TVDEVRAQF....GAAFPVVPGE.....TGGRLNESEIRD.....ALTGELFRQG.....
Cyanothece NLSDIALIFPDVLVLNELIKNQK.....KYGSGLESTVAKWTNRG.....WEILRQGSIQLTGNQ...
D.radiodurans TQAEAEAYA....LADLTLPEPAGTA.....DRPPAAESTVLRLPAEG.ED...TARMLRLGLSQADLAARL
D.propionicus TADEVERIF....DTEVDLILDGG.....HTPGGSGSTLVGLHGAT.....LTCLREGKLAFATVRSRV
H D GVE P

1. Pour commencer vous avez fait un BLASTp en utilisant Sua5 de *P. abyssi* et en restreignant la recherche uniquement aux Archaeoglobi. Vous avez récupéré les séquences de toutes les Sua5 d'espèces cultivées (*A. sulfaticallidus*, *G. acetivorans, G. ahangari, F. placidus, A. veneficus, A. profundus, A. fulgidus*). Vous allez inclure également la Sua5 de *P. abyssi* pour servir d'outgroup.

>WP_015590325.1 threonylcarbamoyl-AMP synthase [Archaeoglobus sulfaticallidus]

MKKTRIIKVNPENPEEDRIMAGAEIIKNGGLVAFPTETVYGLGANALDRKAVLGIYKAKERPADNPLIVHVCNKDMIEDI
AELNDVAEKLIKRFFPGPLTVVVKKKECIPQEVTAGLDTVAIRMPDNKVALKLIELSEVPISAPSANLAGKPSPTKAEHV
IEDLYGRIDAILDGGPTNIGLESTVVDTTVYPVELLRPGGLSVEELEKVVDIKIPELDEGIPRSPGMKYRHYAPSAELIV
IYGKREDVVSKIIKTAGQLIDKSNKKIGLVVSDSEPYKDIDVEIVEIGRSVEEVARNLFSALRELDRRGIDLIIAEGVEE
KGLGLAVMNRLKKASNYRIFRV

>WP_048093555.1 threonylcarbamoyl-AMP synthase [Geoglobus acetivorans]
MIRISSEKFSEEELSYPAELIKAGNLVAFPTETVYGLGANALEGKAVRKIFAAKGRPSDNPLIVHVHEHEQIYELAKPNR
VAEKLVEEFFPGPLTLVMRKKEVVPGETTGGLNTVAVRMPAHKVALKLIELSGVPIAAPSANKSGKPSPTRAEHVIEDFG
NEIDCIIDAGKTRIGLESTVVDTTTYPLEILRPGAITREMLEEFFEVKVIKKSDVARSPGMKYRHYSPEADVIVLVGDTA
RDEMRELAERLSRDGRKVGIAAMKADEFEGIATYNLGSTLREFAERLFDALRELDRVCDVIIVEGVEEKGLGLAIMNRLS
KAGRVYRV

>WP_048095299.1 threonylcarbamoyl-AMP synthase [Geoglobus ahangari]
MRVIRVAPERFRDEELEPAADIIRQGGLVAFPTETVYGLGADALNERVVRRIFEAKGRPSDNPLIVHVSSIEEVYRIAKP
NRVAEKLMEEFFPGPLTLVMEKREAVPEVTTGGLRTVAVRMPKHRVALKLIELSGTPIAAPSANKSGKPSPTRAEHVMED
FESEIECVIDAGRTEVGLESTVVDTTVYPIEILRPGAITREMLEELFEVRVAGKSDVARSPGMKYRHYSPEADTIVIVGE
KRREEIKRLAHRLAGEGRTVGVAAMNGEEFESVAPHVYDLGASLEEFAERLFDALRTLDRFCDVIVVEGVEQKGIGLAIM
NRLSKAGRVYRV

>WP_012965368.1 threonylcarbamoyl-AMP synthase [Ferroglobus placidus]
MKIRTRIIKVDPENPEDDKIGIAAEEIKKGNLVAFPTETVYGLGADALNEKAVRKIFEVKGRPADNPLIVHVSSVDQIYE
IAQPNEIAKKLIDAFFPGPLTLVMKNKKVPKITTAGLDTVAVRMPDHKVALKLIELSETPIAAPSANKSGKPSPTKAEHV
LEDLGNLVDVIIDGGETKIGLESTVVDTTVYPVEILRPGAITKEELEEYVDVKYAEDFSVAKSPGMKYKHYAPDAETIVL
VGKNFLKKATQIAEDFVKKGYRVGVAGLNLEEKRNDVIYKNFGRNLEEMAKNLFKVLRELDKECDVIIVQGVEEKGIGKA
IMNRLYKAGRVFRI

>WP_013683929.1 threonylcarbamoyl-AMP synthase [Archaeoglobus veneficus]
MIVIRVDPVNPERGKIARAAEIIKKGGLVAFPTETVYGLGGDALNENAVRRIFEAKERPPRNPLIVHVSSVEQVYRIAEV
NEVAEKLMGEFFPGPLALVLKKKDVVPDITTAGMKKVAVRMPDHKVALTLIELSETPIAAPSANISGRPSPTKAEHVIED
LAGRIDAILDAGEVKIGIESTVLDVTSKPAKILRLGAITPEMLVERGIEVEVIDTRPFRHYQTKAKLYVVNAENLAEFVG
SLREKGVKVGVARITAECDADRIVELGKSIEDVAKNLFSALRELDRSGVDIIVVEAVERKGLGKVIMSKLEEAGEIV

>WP_010878284.1 threonylcarbamoyl-AMP synthase [Archaeoglobus fulgidus]
MAEVLEPSREGIRKAVAVLRAGGIVAFPTETVYGMGCDATNEEALRRLYEIKGRSLNKPFIVGVWSRDYVGKIAEVDERA
EKLMEAFFPGPLTLVLKSKGVMPSLLSPKGKIAVRMPAHEVPLQLMEMLRKPIVVPSANLSGRPSLMRWEHVVEELGSRI
DAVVKGECKVGVESTIVDLTETPAKVLRVGAVSVESIKKHVEVVVEPRKETYSLSCQVYVFVGERALQRIKEFVDEAEKR
GRVVVIAREKITDETIVIGKSAEEYSANLFSAVREAESRRPDIIVIEGVENEAIMDRLRRLAGERVFRV

>WP_012939514.1 threonylcarbamoyl-AMP synthase [Archaeoglobus profundus]
MTKIYKVDPRNPDESVLREVAHMIKEGKLVAYPTETVYGLGTNALDENAVKRLFEVKGRPKKPVSVVVSDLDHVERIAEP
NETALKLMEKFFPGPITIIVKKKKVIPPIVTAGTDKIGIRLPDYKIPLKLAEFSGVPVTSTSANVSGKPSPTKPEHIMVD
FMGKIDAILDAGETPLKIESTVIDTTTEPPRVLRVGALPLNEIEKVVGEVEVLDYRAYKPKAKVVAVFKGDVEEVASRFE
GRVCIIKEVKDLMDVLRRCKDCDVIVFECIDLSEAVRARIMSIADVIYD

>WP_010868714.1 threonylcarbamoyl-AMP synthase [Pyrococcus abyssi]
MTIIINVRERIEEWKIRIAAGFIREGKLVAFPTETVYGLGANALDENAVKRIFEAKGRPADNPLIIHIAS
FEQLEVLAKEIPEEAEMLAKRFWPGPLTLVLPKSEVVPRVITGGLDTVAVRMPAHEIALKLIELSERPIA
APSANISGKPSPTSAHHVAEDFYGKIECIIDGGETRIGVESTVIDLTEWPPVLLRPGGLPLEEIEKVIGE
IRIHPAVYGKSVDTAKAPGMKYRHYAPSAEVIVVEGPRDKVRRKIEELIAKFKEEGKKVGVIGSGSYDAD
EVFYLGDTVEEIARNLFKALRHMDRTGVDVILAEGVEEKGLGLAVMNRLRKASGYRIIKV

2. Faites un alignement de ces séquences avec Clustal Omega et copiez-collez l'alignement ci-dessous. Identifiez et soulignez les motifs conservés (P^{227}-G^{228}-M^{229} et H^{234}-Y^{235} et K^{56}-X-R^{58}/S^{143}-X-N^{145}). Que remarquez-vous ?

```
WP_010878284.1      --------MAEVLEPSREGIRKAVAVLRAGGIVAFPTETVYGMGCDATNEEALRRLYEIK      52
WP_012939514.1      ---MTKIYKVDPRNPDESVLREVAHMIKEGKLVAYPTETVYGLGTNALDENAVKRLFEVK      57
WP_010868714.1      ----MTIIINVRERIEEWKIRIAAGFIREGKLVAFPTETVYGLGANALDENAVKRIFEAK      56
WP_013683929.1      ----MIVIRVDPVNPERGKIARAAEIIKKGGLVAFPTETVYGLGGDALNENAVRRIFEAK      56
WP_015590325.1      -MKKTRIIKVNPENPEEDRIMAGAEIIKNGGLVAFPTETVYGLGANALDRKAVLGIYKAK      59
WP_012965368.1      MKIRTRIIKVDPENPEDDKIGIAAEEIKKGNLVAFPTETVYGLGADALNEKAVRKIFEVK      60
WP_048093555.1      ------MIRISSEKFSEEELSYPAELIKAGNLVAFPTETVYGLGANALEGKAVRKIFAAK      54
WP_048095299.1      ----MRVIRVAPERFRDEELEPAADIIRQGGLVAFPTETVYGLGADALNERVVRRIFEAK      56
                                  .     :   .  :: * :**:*******:* :* : ..:  ::  *

WP_010878284.1      GRSLNKPFIVGVWSRDYVGKIAE-VDERAEKLMEAFFPGPLTLVLKSKGVMPSLLS-PKG      110
WP_012939514.1      GRPK-KPVSVVVSDLDHVERIAE-PNETALKLMEKFFPGPITIIVKKKKVIPPIVTAGTD      115
WP_010868714.1      GRPADNPLIIHIASFEQLEVLAKEIPEEAEMLAKRFWPGPLTLVLPKSEVVPRVITGGLD      116
WP_013683929.1      ERPPRNPLIVHVSSVEQVYRIAE-VNEVAEKLMGEFFPGPLALVLKKKDVVPDITTAGMK      115
WP_015590325.1      ERPADNPLIVHVCNKDMIEDIAE-LNDVAEKLIKRFFPGPLTVVVKKKECIPQEVTAGLD      118
WP_012965368.1      GRPADNPLIVHVSSVDQIYEIAQ-PNEIAKKLIDAFFPGPLTLVMKNK-KVPKITTAGLD      118
WP_048093555.1      GRPSDNPLIVHVHEHEQIYELAK-PNRVAEKLVEEFFPGPLTLVMRKKEVVPGETTGGLN      113
WP_048095299.1      GRPSDNPLIVHVSSIEEVYRIAK-PNRVAEKLMEEFFPGPLTLVMEKREAVPEVTTGGLR      115
                     *   :*. : : . : :  :*:     *  *   *:***:::::.   :*   :

WP_010878284.1      KIAVRMPAHEVPLQLMEMLRKPIVVPSANLSGRPSLMRWEHVVEELGSRIDAVVK-GECK      169
WP_012939514.1      KIGIRLPDYKIPLKLAEFSGVPVTSTSANVSGKPSPTKPEHIMVDFMGKIDAILDAGETP      175
WP_010868714.1      TVAVRMPAHEIALKLIELSERPIAAPSANISGKPSPTSAHHVAEDFYGKIECIIDGGETR      176
WP_013683929.1      KVAVRMPDHKVALTLIELSETPIAAPSANISGRPSPTKAEHVIEDLAGRIDAILDAGEVK      175
WP_015590325.1      TVAIRMPDNKVALKLIELSEVPISAPSANLAGKPSPTKAEHVIEDLYGRIDAILDGGPTN      178
WP_012965368.1      TVAVRMPDHKVALKLIELSETPIAAPSANKSGKPSPTKAEHVLEDLGNLVDVIIDGGETK      178
WP_048093555.1      TVAVRMPAHKVALKLIELSGVPIAAPSANKSGKPSPTRAEHVIEDFGNEIDCIIDAGKTR      173
WP_048095299.1      TVAVRMPKHRVALKLIELSGTPIAAPSANKSGKPSPTRAEHVMEDFESEIECVIDAGRTE      175
                    .:.:*:*  .: * * *:   *:   *** :*:**    .*:  :: . :: ::. *

WP_010878284.1      VGVESTIVDLTETPAKVLRVGAVSVESIKKH-VEVVVE---------------PRKETYS      213
WP_012939514.1      LKIESTVIDTTTEPPRVLRVGALPLNEIEKVVGEVEV---------------LDYRAYK      219
WP_010868714.1      IGVESTVIDLTEWPPVLLRPGGLPLEEIEKVIGEIRIHPAVYGKSVDTAKAPGMKYRHYA      236
WP_013683929.1      IGIESTVLDVTSKPAKILRLGAITPEMLVERGIEVE-------------VIDTRPFRHYQ      222
WP_015590325.1      IGLESTVVDTTVYPVELLRPGGLSVEELEKV-VDIKI----PELDEGIPRSPGMKYRHYA      233
WP_012965368.1      IGLESTVVDTTVYPVEILRPGAITKEELEEY-VDVKY----AE-DFSVAKSPGMKYKHYA      232
WP_048093555.1      IGLESTVVDTTTYPLEILRPGAITREMLEEF-FEVKV----IK-KSDVARSPGMKYRHYS      227
WP_048095299.1      VGLESTVVDTTVYPIEILRPGAITREMLEEL-FEVRV----AG-KSDVARSPGMKYRHYS      229
                    : :***::* *  *  :** *.:  : : :   ::                    . *

WP_010878284.1      LSCQVYVFVGERALQR---IKEFVDEAEKRGRVVVIAR------EKITDETIVIGKSAEE      264
WP_012939514.1      PKAKVVAVFKGDVEEV-------ASRFEGRVC-------------------------IIK      247
WP_010868714.1      PSAEVIVVEGPRDKVRR-KIEELIAKFKEEGKKVGVIGSGSY----DADEVFYLGDTVEE      291
WP_013683929.1      TKAKLYVVNAENLAE-------FVGSLREKGVKVGVARITAEC---DADRIVELGKSIED      272
WP_015590325.1      PSAELIVIYGKREDVVSKIIKTAGQLIDKSNKKIGLVVSDSEPYKDIDVEIVEIGRSVEE      293
WP_012965368.1      PDAETIVLVGKNFLKK---ATQIAEDFVKKGYRVGVAGLNLEEKRND-VIYKNFGRNLEE      288
WP_048093555.1      PEADVIVLVGDTARDE---MRELAERLSRDGRKVGIAAMKADEFEGIA--TYNLGSTLRE      282
WP_048095299.1      PEADTIVIVGEKRREE---IKRLAHRLAGEGRTVGVAAMNGEEFESVAPHVYDLGASLEE      286
                     ...  ..                                                .

WP_010878284.1      YSANLFSAVREAESRRPDIIVIEGVENE----AIMDRLRRLAGERVFRV       309
WP_012939514.1      EVKDLMDVLRRC--KDCDVIVFECIDLSE---AVRARIMSIAD-VIYD-       289
WP_010868714.1      IARNLFKALRHMDRTGVDVILAEGVEEKGLGLAVMNRLRKASGYRIIKV       340
WP_013683929.1      VAKNLFSALRELDRSGVDIIVVEAVERKGLGKVIMSKLEEAGEIV----       317
WP_015590325.1      VARNLFSALRELDRRGIDLIIAEGVEEKGLGLAVMNRLKKASNYRIFRV       342
WP_012965368.1      MAKNLFKVLRELDKE-CDVIIVQGVEEKGIGKAIMNRLYKAGR--VFRI       334
WP_048093555.1      FAERLFDALRELDRV-CDVIIVEGVEEKGLGLAIMNRLSKAGR--VYRV       328
WP_048095299.1      FAERLFDALRTLDRF-CDVIVVEGVEQKGIGLAIMNRLSKAGR--VYRV       332
                        *:..:*       *:*: : :: .    .:  ::   .
```

On remarque que le motif de fixation d'ATP K^{56}-X-R^{58}/S^{143}-X-N^{145} est conservé chez toutes les séquences, cependant le motif P^{227}-G^{228}-M^{229} a été perdu chez 3 espèces (un indel), 2 de ces 3 espèces ont également perdu (substitution) le résidu His dans le motif H^{234}-Y^{235}.

3. Sachant que les protéines Sua5/TsaC sont essentielles pour la survie des cellules et que chez ces Archaeaoglobi chaque espèce code pour un gène *sua5*, que suggère votre analyse de conservation de motifs importants pour la fonction des protéines Sua5 ?

Elle suggère que, chez 3 espèces *A. profundus*, *A.fulgidus* et *A. veneficus*, le linker (qui relie le domaine catalytique au domaine SUA5) a perdu sa fonction mais que la protéine est fonctionnelle tout de même. Autrement dit ces trois protéines n'ont plus besoin de linker pour fonctionner alors que le linker est essentiel pour les Sua5 classiques.

4. Regardez de nouveau l'alignement. Comparez la taille du linker et domaine SUA5 entre les Sua5 des trois espèces que vous avez identifiées comme porteuses de mutations dans le linker et la Sua5 *d'A. sulfaticallidus*. Pour information, chez *P. abyssi* le linker débute au résidu His^{212} de Sua5. Il comporte environ 20 résidus. Que remarquez-vous ?

On remarque que le linker et de domaine SUA5 chez *A. profundus, A.fulgidus* et *A. veneficus* comportent des délétions et que le nombre de résidus perdus est le plus grand chez *A. profundus*.

5. Enfin, trouvez dans l'alignement le résidu signature (adjacent au motif S^{143}-X-N^{145}) et soulignez-le en rouge ; ce résidu est toujours une Pro (P) pour les protéines Sua5 et toujours une Thr (T) pour les protéines TsaC. Que remarquez-vous ? À la vue de l'ensemble de vos observations quelle hypothèse pouvez-vous émettre ?

On remarque que le résidu signature est une Pro comme attendu pour six Sua5, une exception remarquable est le Sua5 de *A. profundus* qui code pour une Thr au lieu d'une Pro. De plus, c'est chez le même organisme que le linker et le domaine SUA5 portent le plus de délétions. A partir de ces données on peut proposer que le Sua5 *d'A. profundus* est une forme intermédiaire entre la variante longue Sua5 et la variante courte TsaC. Il semblerait que dans ce groupe d'organismes les Sua5 sont en train de perdre leur domaine SUA5 pour devenir la forme courte TsaC ; la Sua5 de *A. profundus* est le cas le plus avancé de ce processus.

Nous allons maintenant tenter de reconstituer le processus évolutif conduisant aux Sua5 plus courtes en utilisant l'analyse cladistique. Pour cela vous allez identifier parmi les résidus spécifiques et conservés des Sua5 ceux qui sont des synapomorphies.

6. Question préliminaire : dans l'analyse cladistique qu'est qu'un caractère plésiomorphe ? Apomorphe ? Synapomorphe ?

Selon l'approche cladistique de classification un caractère peut présenter soit un état ancestral (plésiomorphe) ou alors un état dérivé, évolué (apomorphe). Quand l'état apomorphe est partagé par deux ou plusieurs espèces cela s'appelle une synapomorphie.

7. Parmi ces trois catégories de caractères lequel peut être utilisé pour inférer la phylogénie ?

Les synapomorphies : Le partage par au moins deux taxons certifie que la dérive du caractère n'est pas simplement contingente, mais est apparue chez l'ancêtre commun des espèces qui partagent la synapomorphie.

8. Le tableau ci-dessous répertorie quelques résidus spécifiques (pas tous pour simplifier l'exercice) et conservés pour les protéines Sua5. Complétez-le en vous servant de votre alignement des Sua5 d'Archaeoglobi. Quand le résidu est présent mettez un X dans la case, quand il est absent coloriez la case en bleu, quand le résidu a été substitué coloriez la case en rouge. La numérotation des résidus correspond à celle de la Sua5 de *P. abyssi*.

	P59	N62	H67	K/R149	P152	T153	H157	D161	G178	P228	G229	M230	H234	Y235
P.abyssi	X	X	X	X	X	X	X	X	X	X	X	X	X	X
A. sulfaticallidus	X	X	X	X	X	X	X	X	X	X	X	X	X	X
A. veneficus	X	X	X	X	X	X	X	X	X				X	X
A. fulgidus														
A. profundus														

Correction:

	P59	N62	H67	K/R149	P152	T153	H157	D161	G178	P228	G229	M230	H234	Y235
P.abyssi	X	X	X	X	X	X	X	X	X	X	X	X	X	X
A. sulfaticallidus	X	X	X	X	X	X	X	X	X	X	X	X	X	X
A. veneficus	X	X	X	X	X	X	X	X	X				X	X
A. fulgidus				X			X	E	X					X
A. profundus	X			X	X	X	X	X						X

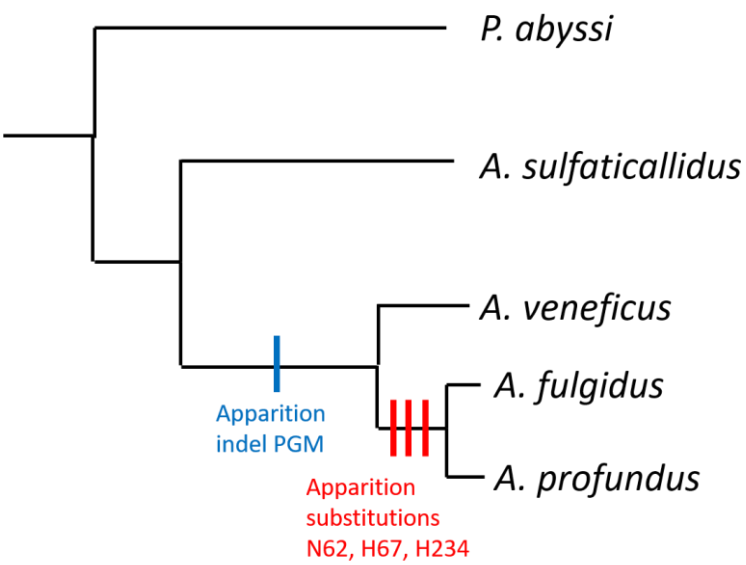
9. Faites maintenant une liste de caractères synapomorphes et des espèces chez lesquelles on les trouve. Pour cela complétez le texte ci-dessous :

Délétion PGM est présente chez : A. veneficus, A. fulgidus et A. profundus

Substitution N62 et H67 est présente chez : A. fulgidus et A. profundus

Substitution H234 est présente chez : A. fulgidus et A. profundus

10. A partir de cette liste dessinez maintenant un cladogramme décrivant l'évolution des protéines Sua5 issues des 5 espèces étudiées. Placez sur les branches l'apparition des synapomorphies chez l'ancêtre d'un clade. Vous pouvez faire votre dessin avec le logiciel Power Point puis le copier-coller ci-dessous.



11. A partir de l'analyse de l'ensemble des données au cours de ce TD proposez un scénario pour l'histoire évolutive de la famille Sua5/TsaC permettant d'expliquer l'existence de deux variantes et leur distribution chaotique (qui ne suit pas l'arbre des espèces) au sein du vivant.

Les données suggèrent que la forme ancestrale est Sua5 et qu'elle était donc présente chez LUCA, l'ancêtre des toutes les cellules modernes. La forme courte TsaC a pu apparaitre au cours de l'évolution par perte du domaine SUA5 au moins une fois. Ce phénomène semble s'opérer actuellement chez certaines espèces d'Archaeoglobi. Les gènes *tsaC* et *sua5* semblent pouvoir se propager au sein du vivant par HGT (exemple d'Aspergillus). Les deux phénomènes (perte du domaine SUA5 et HGT) combinés peuvent expliquer la distribution actuelle de Sua5 et TsaC au sein du vivant.

12. Ces TDs sont intitulés « L'évolution des protéines universelles Sua5/TsaC : quand la simplicité est l'ultime sophistication ? » Pouvez-vous maintenant expliquer pourquoi les termes simplicité et sophistication ont été utilisés ?

Simplification : car les données suggèrent que les protéines Sua5 peuvent perdre un domaine, devenir donc des polypeptides plus simples, contentant un seul domaine au lieu de deux.
Sophistication : on peut spéculer sur les « raisons » pour lesquelles l'évolution conduit vers une perte du domaine SUA5 ; peut-être que les TsaC sont des versions « améliorées » plus performantes ? Cela reste à vérifier par des expériences biochimiques, structurales et *in vivo*.