Oral: BQ4CY010 Programmation 3 et projet tuteuré

Université Paris Cité

ClustalLite : Alignement multiple heuristique par la méthode Clustal

Stéphanie Gnanalingam

M2 Bio-informatique parcours biologie informatique

02/09/2024 - 13/09/2024

Développement

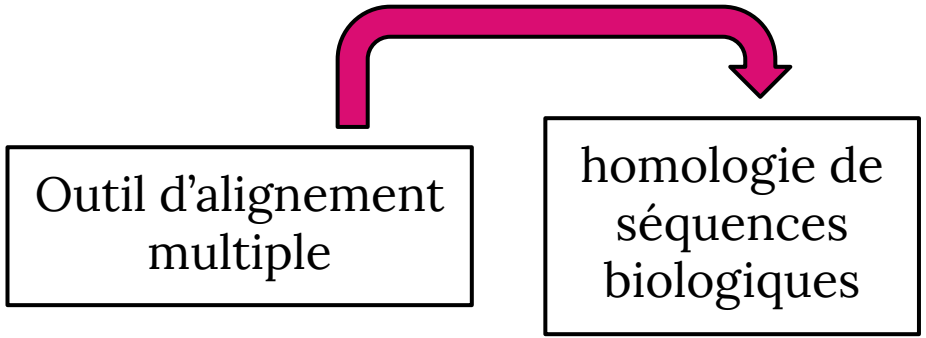
- https://github.com/gnanalin/ClustalLite/tree/main

- CONDA (v.24.7.1)  env_clustallite

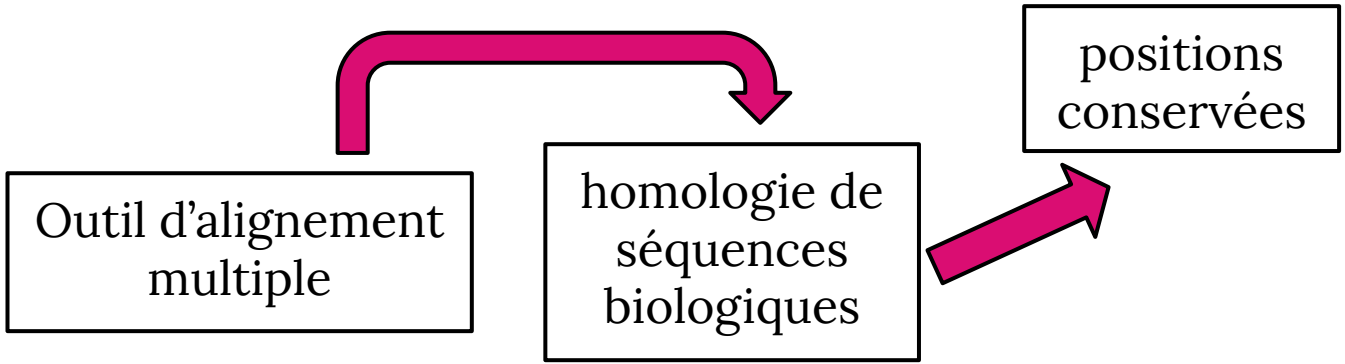
Applications

Outil d’alignement multiple

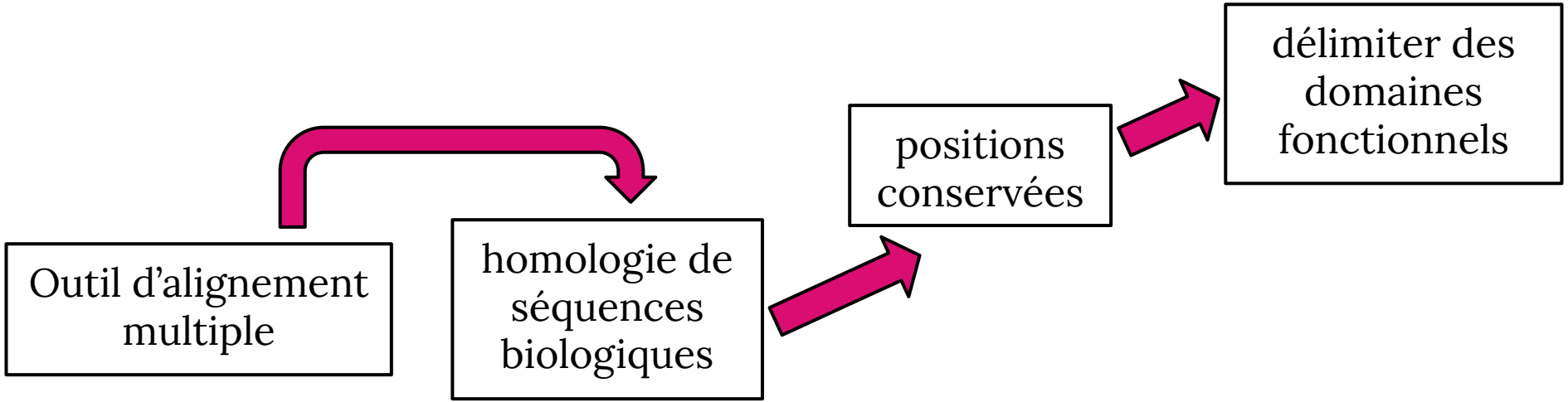
Applications



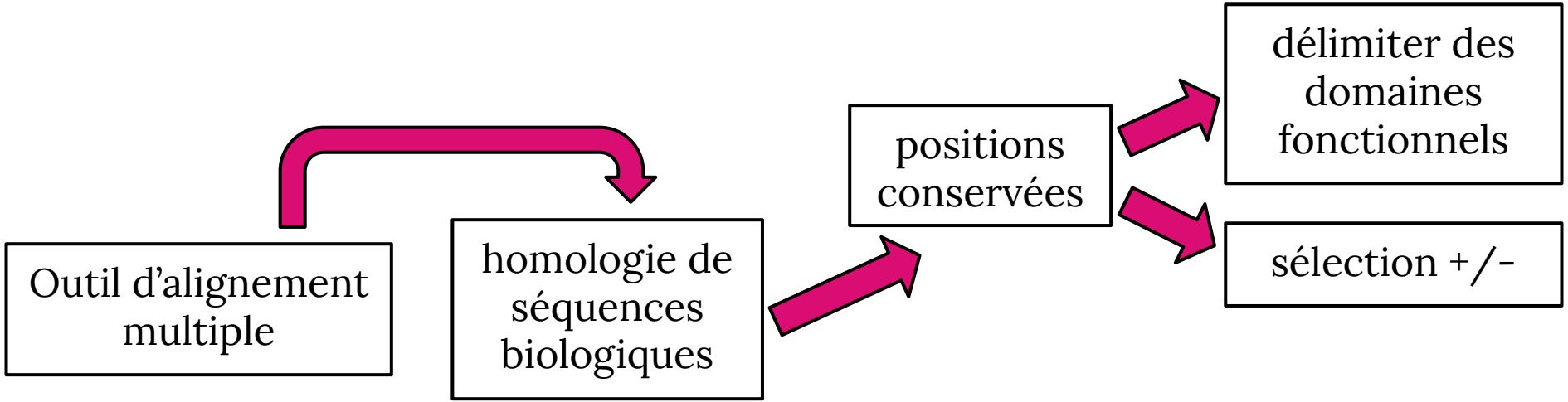
Applications



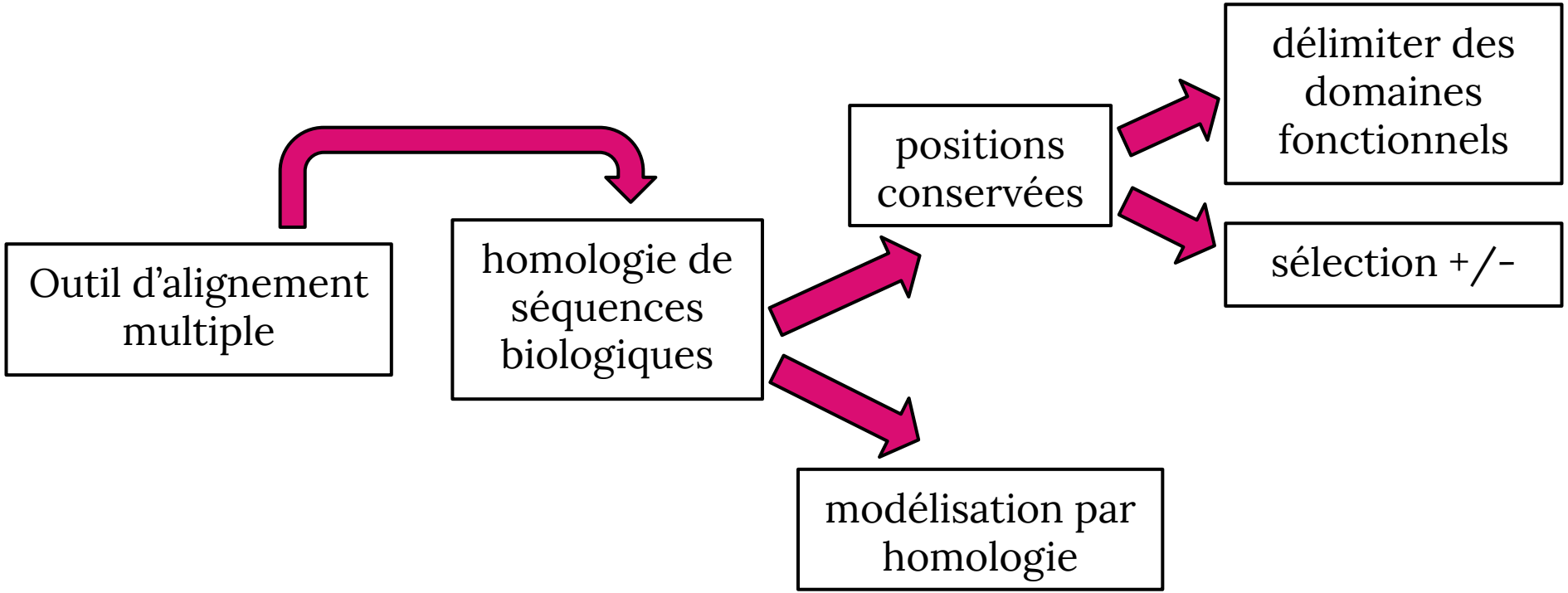
Applications



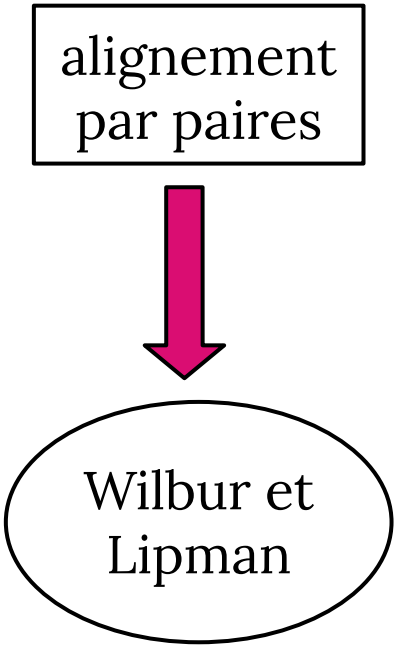
Applications



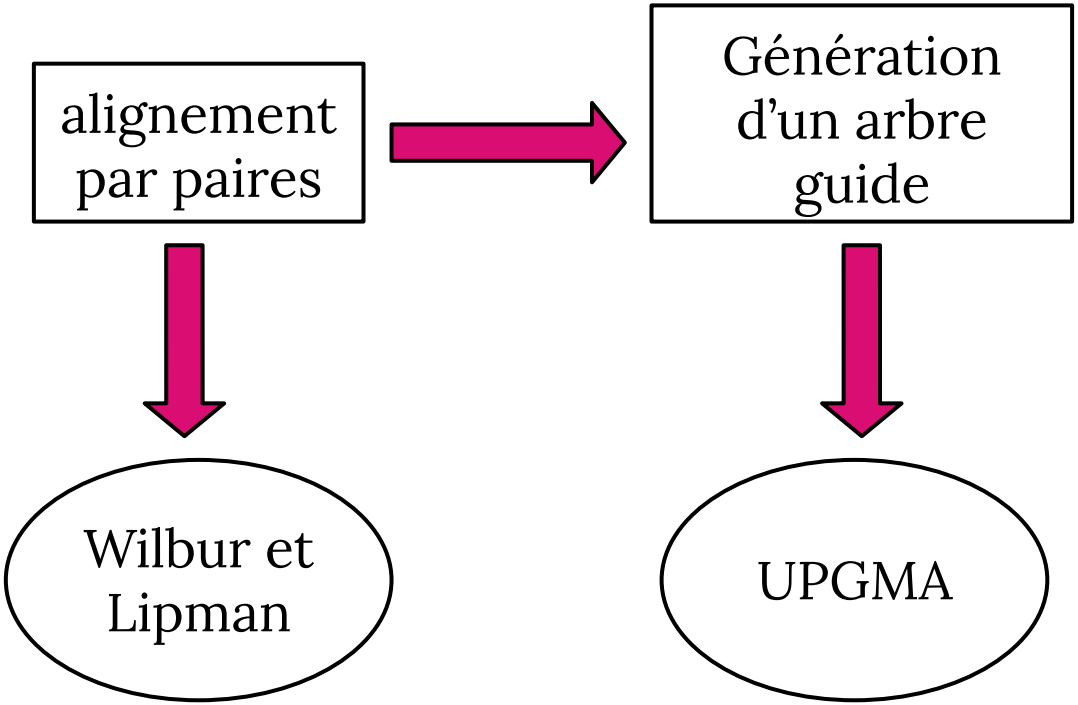
Applications



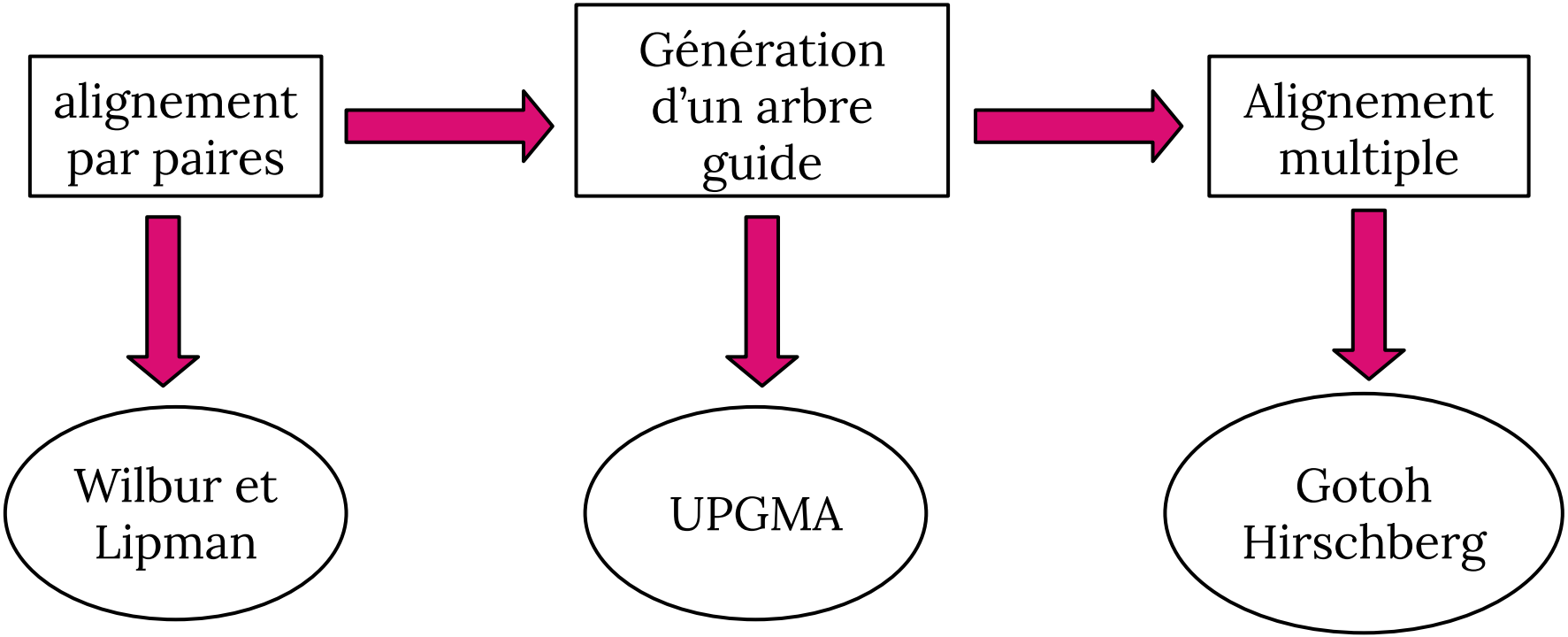
Étapes



Étapes


alignement par paires
Génération d'un arbre guide
Wilbur et Lipman
UPGMA

Étapes

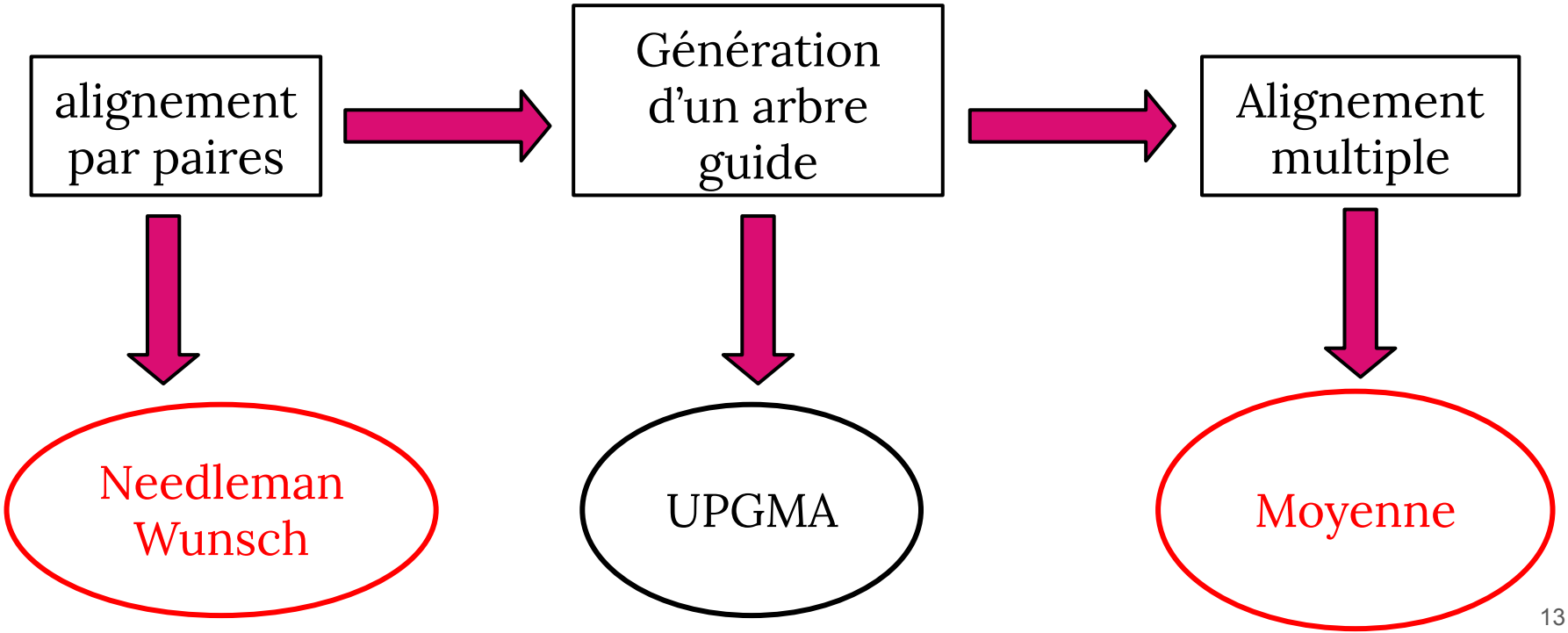

alignement par paires
Génération d'un arbre guide
Alignement multiple
Wilbur et Lipman
UPGMA
Gotoh Hirschberg

Objectif

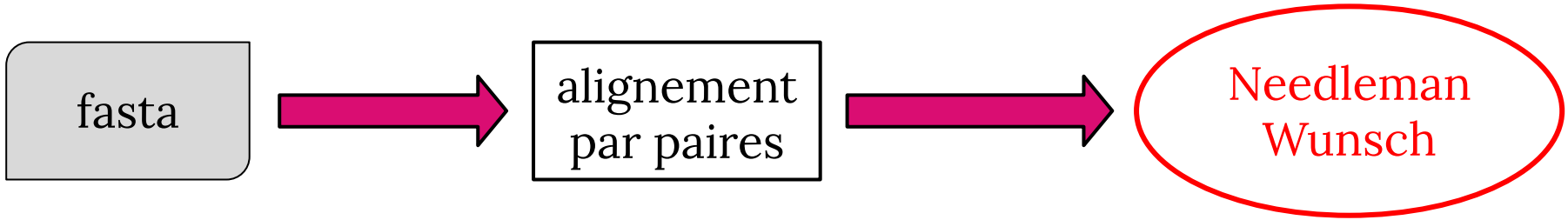
- ***Réimplémenter une version simplifiée de Clustal***

Objectif

- ***Réimplémenter une version simplifiée de Clustal***



Les algorithmes



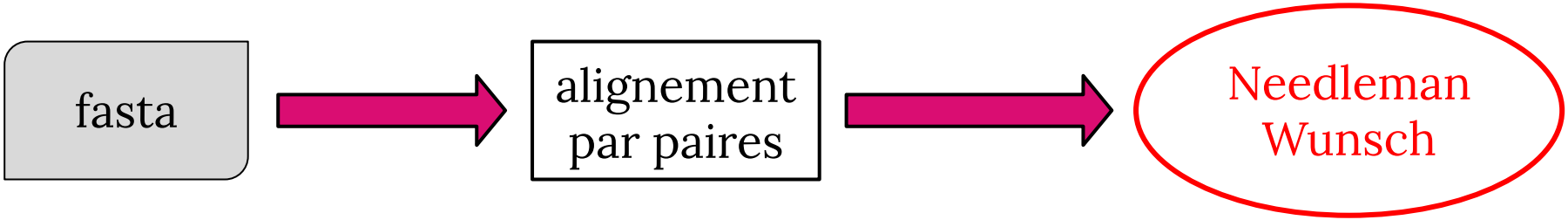
numpy (v2.1.1)

F(I – 1, J – 1) + S(I, J)	F(I – 1, J) – gap
F(I, J – 1) – gap	F(I, J)

- gap = 8
- S(I, J) : BLOSUM62

pandas (v2.2.2)

Les algorithmes



F(I – 1, J – 1) + S(I, J)	F(I – 1, J) – gap
F(I, J – 1) – gap	F(I, J)

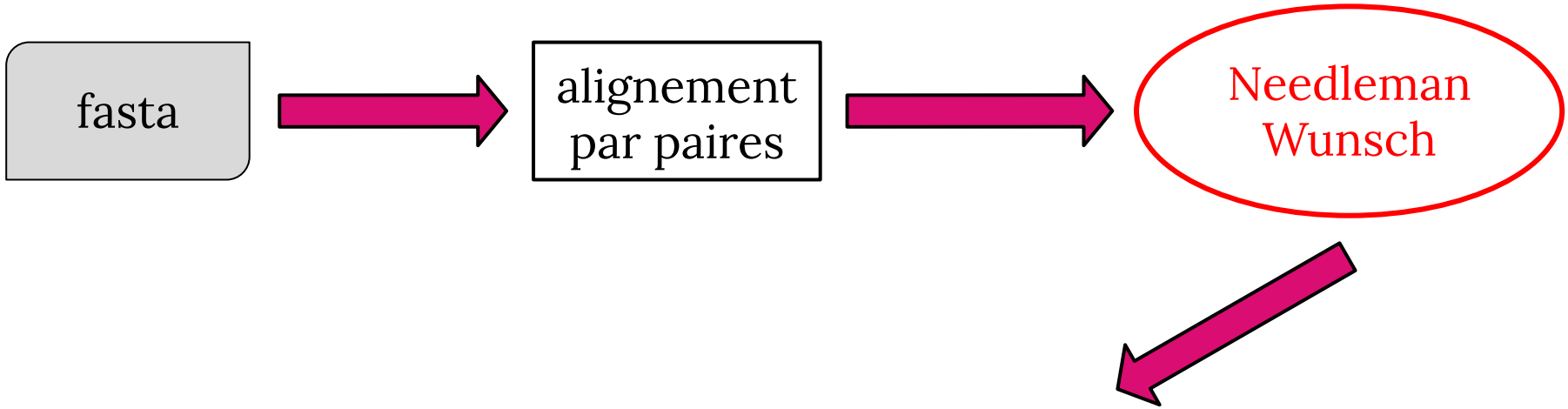
1	1
2	0

0 : diagonale

1: gauche

2: haut

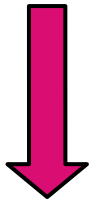
Les algorithmes



	Seq1	Seq2	Seq3
Seq1	-10	12	89
Seq2	-10	-10	15
Seq3	-10	-10	-10

Les algorithmes

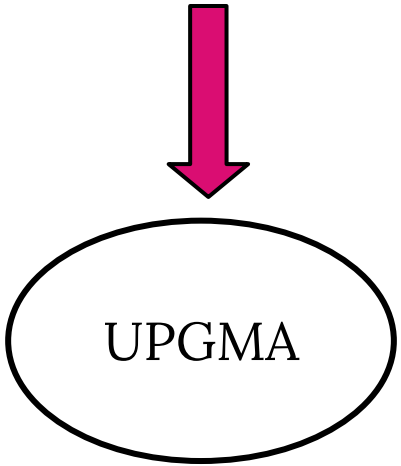
	Seq1	Seq2	Seq3
Seq1	-10	12	89
Seq2	-10	-10	15
Seq3	-10	-10	-10



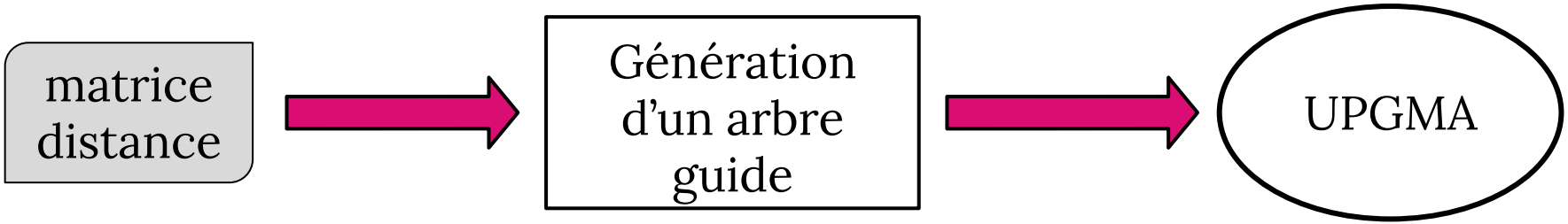
$$matrice\ distances = 1 - ((matrice\ scores - minimum) / maximum - minimum)$$

Les algorithmes

	Seq1	Seq2	Seq3
Seq1	-10	1	0
Seq2	-10	-10	0.96
Seq3	-10	-10	-10



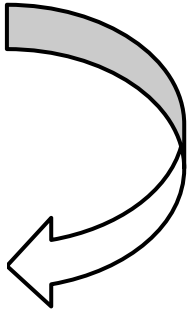
Les algorithmes



	Seq1	Seq2	Seq3
Seq1	-10	1	0
Seq2	-10	-10	0.96
Seq3	-10	-10	-10

Les algorithmes

	Seq1, Seq3	Seq2
Seq1, Seq3	-10	0.98
Seq2	-10	-10



$$distance\,((i,\ j),\ k) = \frac{n_i * d_{ki} + n_k + d_{kj}}{n_i + n_j}$$

Les algorithmes

	Seq1, Seq3	Seq2
Seq1, Seq3	-10	0.98
Seq2	-10	-10

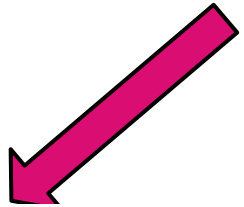
$$distance\ ((i,\ j),\ k) = \frac{n_i * d_{ki} + n_k + d_{kj}}{n_i + n_j}$$

Arbre guide : ((Seq1, Seq3), Seq2)

Les algorithmes

Arbre guide : ((Seq1, Seq3), Seq2)

Alignement multiple



$$diagonale(F(i,\ j)) \ = \ F(i\ -\ 1,\ j\ -\ 1)\ +\ moyenne(cluster1(i),\ cluster2(j))$$

Les algorithmes

Arbre guide : ((Seq1, Seq3), Seq2) → Alignement multiple

$$diagonale(F(i, j)) = F(i-1, j-1) + moyenne(cluster1(i), cluster2(j))$$

Seq1 : **A**WVED
Seq2 : **-**TALD

Seq3 : **R**EDAL

→ diagonale = (S(A, R)-gap)/2

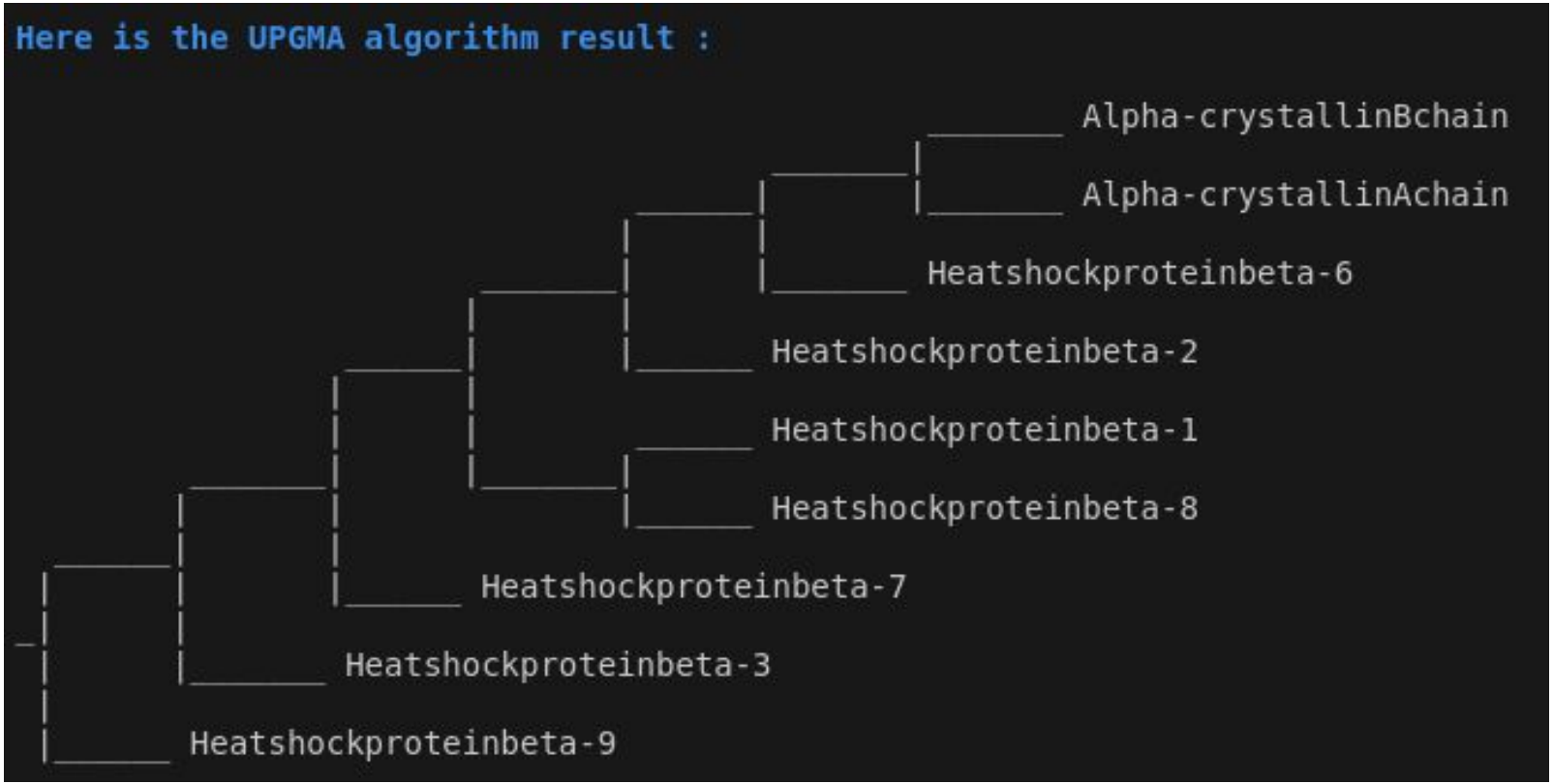
Résultats

```
The file exists and is going to be parsed...

Here is the file's content :

Alpha-crystallinBchain: MDIAIHHPWIRRPFFPFHSPSRLFDQFFGEHLLESDLFPTSTSLSPFYLRPPSFLRAPSWFDTGLSEMRLEKDRFSVNLDVKHFSPEELKVKVLGDVIEVH
GKHEERQDEHGFISREFHRKYRIPADVDPLTITSSLSSDGVLTVNGPRKQVSGPERTIPITREEKPAVTAAPKK

Alpha-crystallinAchain: MDVTIQHPWFKRTLGPFYPSRLFDQFFGEGLFEYDLLPFLSSTISPYYRQSLFRTVLDSGISEVRSDRDKFVIFLDVKHFSPEDLTVKVQDDFVEIHGKHN
ERQDDHGYISREFHRRYRLPSNVDQSALSCSLSADGMLTFCGPKIQTGLDATHAERAIPVSREEKPTSAPSS

Heatshockproteinbeta-6: MEIPVPVQPSWLRRASAPLPGLSAPGRLFDQRFGEGLLEAELAALCPTTLAPYYLRAPSVALPVAQVPTDPGHFSVLLDVKHFSPEEIAVKVVGEHVEVHA
RHEERPDEHGFVAREFHRRYRLPPGVDPAAVTSALSPEGVLSIQAAPASAQAPPPAAAK

Heatshockproteinbeta-1: MTERRVPFSLLRGPSWDPFRDWYPHSRLFDQAFGLPRLPEEWSQWLGGSSWPGYVRPLPPAAIESPAVAAPAYSRALSRQLSSGVSEIRHTADRWRVSLDV
NHFAPDELTVKTKDGVVEITGKHEERQDEHGYISRCFTRKYTLPPGVDPTQVSSSLSPEGTLTVEAPMPKLATQSNEITIPVTFESRAQLGGPEAAKSDETAAK

Heatshockproteinbeta-2: MSGRSVPHAHPATAEYEFANPSRLGEQRFGEGLLPEEILTPTLYHGYYVRPRAAPAGEGSRAGASELRLSEGKFQAFLDVSHFTPDEVTVRTVDNLLEVSA
RHPQRLDRHGFVSREFCRTYVLPADVDPWRVRAALSHDGILNLEAPRGGRHLDTEVNEVYISLLPAPPDPEEEEEAAIVEP

Heatshockproteinbeta-8: MADGQMPFSCHYPSRLRRDPFRDSPLSSRLLDDGFGMDPFPDDLTASWPDWALPRLSSAWPGTLRSGMVPRGPTATARFGVPAEGRTPPPFPGEPWKVCVN
VHSFKPEELMVKTKDGYVEVSGKHEEKQQEGGIVSKNFTKKIQLPAEVDPVTVFASLSPEGLLIIEAPQVPPYSTFGESSFNNELPQDSQEVTCT

Heatshockproteinbeta-3: MAKIILRHLIEIPVRYQEEFEARGLEDCRLDHALYALPGPTIVDLRKTRAAQSPPVDSAAETPPREGKSHFQILLDVVQFLPEDIIIQTFEGWLLIKAQHG
TRMDEHGFISRSFTRQYKLPDGVEIKDLSAVLCHDGILVVEVKDPVGTK

Heatshockproteinbeta-7: MSHRTSSTFRAERSFHSSSSSSSSSTSSSASRALPAQDPPMEKALSMFSDDFGSFMRPHSEPLAFPARPGGAGNIKTLGDAYEFAVDVRDFSPEDIIVTTS
NNHIEVRAEKLAADGTVMNTFAHKCQLPEDVDPTSVTSALREDGSLTIRARRHPHTEHVQQTFRTEIKI

Heatshockproteinbeta-9: MQRVGNTFSNESRVASRCPSVGLAERNRVATMPVRLLRDSPAAQEDNDHARDGFQMKLDAHGFAPEELVVQVDGQWLMVTGQQQLDVRDPERVSYRMSQKV
HRKMLPSNLSPTAMTCCLTPSGQLWVRGQCVALALPEAQTGPSPRLGSLGSKASNLTR
```

Résultats

```
Here are the Needleman-Wunsch alignements :

Alpha-crystallinBchain: MDIAIHHPWIRRPFFPFHSPSRLFDQFFGEHLLESDLFP-TSTSLSPFYLRPPSFLRAPS
Alpha-crystallinAchain: MDVTIQHPWFKRTLGPFY-PSRLFDQFFGEGLFEYDLLPFLSSTISPYY-R-QSLFR--T

Alpha-crystallinBchain: WFDTGLSEMRLEKDRFSVNLDVKHFSPEELKVKVLGDVIEVHGKHEERQDEHGFISREFH
Alpha-crystallinAchain: VLDSGISEVRSDRDKFVIFLDVKHFSPEDLTVKVQDDFVEIHGKHNERQDDHGYISREFH

Alpha-crystallinBchain: RKYRIPADVDPLTITSSLSSDGVLTVNGPRKQ--VSG--PERTIPITREEKPAVTAAPKK
Alpha-crystallinAchain: RRYRLPSNVDQSALSCSLSADGMLTFCGPKIQTGLDATHAERAIPVSREEKP--TSAPSS

 Alignement score : 466

Alpha-crystallinBchain: MDIAIH-HP-WIRR-PF-FP-FHSPSRLFDQFFGEHLLESDLFPTSTSLSPFYLRPPSFL
Heatshockproteinbeta-6: MEIPVPVQPSWLRRASAPLPGLSAPGRLFDQRFGEGLLEAEL---A-ALCPTTL-APYYL

Alpha-crystallinBchain: RAPSWFDTGLSEMRLEKDRFSVNLDVKHFSPEELKVKVLGDVIEVHGKHEERQDEHGFIS
Heatshockproteinbeta-6: RAPS-VALPVAQVPTDPGHFSVLLDVKHFSPEEIAVKVVGEHVEVHARHEERPDEHGFVA

Alpha-crystallinBchain: REFHRKYRIPADVDPLTITSSLSSDGVLTVNGPRKQVSGPERTIPITREEKPAVTAAPKK
Heatshockproteinbeta-6: REFHRRYRLPPGVDPAAVTSALSPEGVLSI-----Q-AAP-AS---AQAPPPA--AA--K

 Alignement score : 254
```

Résultats


Here is the UPGMA algorithm result :
Alpha-crystallinBchain
Alpha-crystallinAchain
Heatshockproteinbeta-6
Heatshockproteinbeta-2
Heatshockproteinbeta-1
Heatshockproteinbeta-8
Heatshockproteinbeta-7
Heatshockproteinbeta-3
Heatshockproteinbeta-9

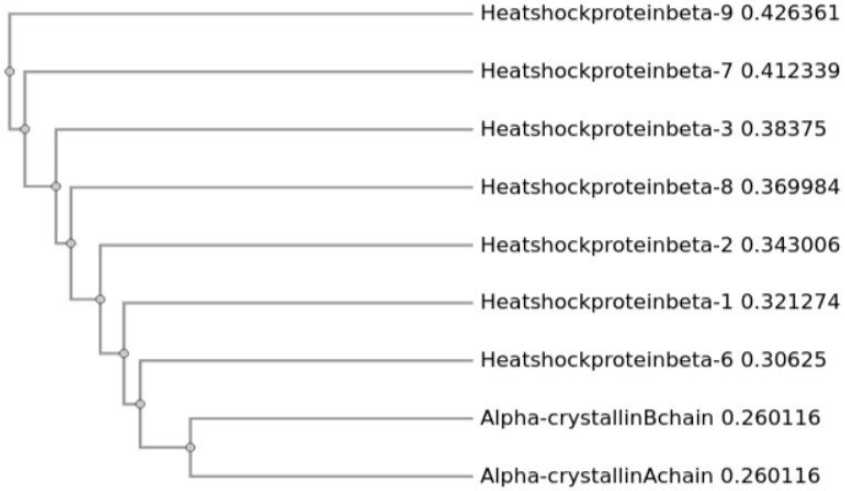
Résultats

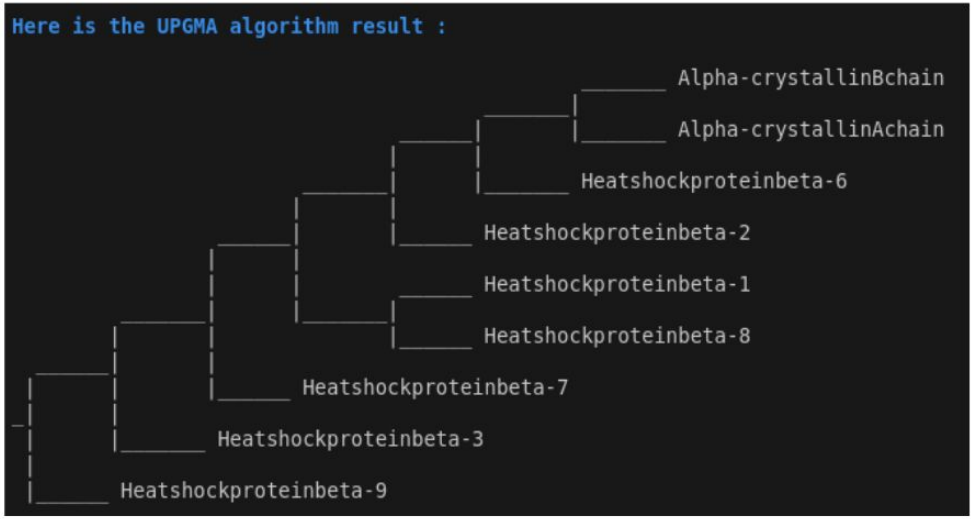
```
Here is the multiple alignment result :

Alpha-crystallinBchain: M-DIAIH-HP-WIRR-PFFPF--HS-PSRLFDQFFGEHLLESD-LFP-TSTSLSPFYLRP--P----S-F
Alpha-crystallinAchain: M-DVTIQ-HP-WFKR-TLGPF--Y--PSRLFDQFFGEGLFEYD-LLPFLSSTISPYY-R---Q----S-L
Heatshockproteinbeta-6: M-EIPVPVQPSWLRR-ASAPLPGLSAPGRLFDQRFGEGLLEAE-LAALCPTTLAPYYLRA--P----S--
Heatshockproteinbeta-2: M-S-GRS-VP-HAHP-ATAEYE-FANPSRLGEQRFGEGLLPEE-IL--TPTLYHGYYVR---P----R-A
Heatshockproteinbeta-1: MTERRVPFSL-LRGP-SWDPFRDWYPHSRLFDQAFGLPRLPEE-WSQWLGGSSWPGYVRPLPPAAIESPA
Heatshockproteinbeta-8: MADGQMPFSCHYPSRLRRDPFRDSPLSSRLLDDGFGMDPFPDDLTASW-PDWALP-RLSSAWPGTLRSGM
Heatshockproteinbeta-7: M-SHRTS-ST-F--R-AERSFHSSSSSSSSSTSSSASRALPAQ-DPP-MEKALSMFSDDF--G----S-F
Heatshockproteinbeta-3: M-A-KI--I--LRHL-IEIPVR-Y--QEEFEARGLEDCRLDHA-LYA-LPGPTIV-DLRK--T----R-A
Heatshockproteinbeta-9: M-Q-RVG-NT-FSNE-SR--V-----ASRC-PS-VG--LAERN--R--VAT-M-P--VR----------L


Alpha-crystallinBchain: LRAP--S-W-FD--T-GLSEMRLEKDRFSVNLDVKHFSPEELKVKVLGDVIEVHGKHE-E-RQDEH-GF-
Alpha-crystallinAchain: FR----T-V-LD--S-GISEVRSDRDKFVIFLDVKHFSPEDLTVKVQDDFVEIHGKHN-E-RQDDH-GY-
Heatshockproteinbeta-6: V-----A---L---P--VAQVPTDPGHFSVLLDVKHFSPEEIAVKVVGEHVEVHARHE-E-RPDEH-GF-
Heatshockproteinbeta-2: APAG--E-G-SR--A-GASELRLSEGKFQAFLDVSHFTPDEVTVRTVDNLLEVSARHP-Q-RLDRH-GF-
Heatshockproteinbeta-1: VAAPAYSRALSRQLSSGVSEIRHTADRWRVSLDVNHFAPDELTVKTKDGVVEITGKHE-E-RQDEH-GY-
Heatshockproteinbeta-8: VPRGPTATARFGVPAEGRTPPPFPGEPWKVCVNVHSFKPEELMVKTKDGYVEVSGKHE-E-KQQEG-GI-
Heatshockproteinbeta-7: MRPHSEPLA-FPARPGGAGNIKTLGDAYEFAVDVRDFSPEDIIVTTSNNHIEVRA----E-KLAAD-GT-
Heatshockproteinbeta-3: AQSP--P-V-DS--A-AETPPREGKSHFQILLDVVQFLPEDIIIQTFEGWLLIKAQHG-T-RMDEH-GF-
Heatshockproteinbeta-9: LR-D--S----P--A-AQEDNDHARDGFQMKLDAHGFAPEELVVQVDGQWLMVTGQQQLDVRDPERVSYR


Alpha-crystallinBchain: ISREFHRKYRIPADVDPLTITSSLSSDGVLTVNGPRKQ--VSG--PE-R-T-IPITR-EE-KPAV-TAA-
Alpha-crystallinAchain: ISREFHRRYRLPSNVDQSALSCSLSADGMLTFCGPKIQTGLDATHAE-R-A-IPVSR-EE-KP---TSA-
Heatshockproteinbeta-6: VAREFHRRYRLPPGVDPAAVTSALSPEGVLSI-----Q----A-------A--PASA-QA-PP---PAA-
Heatshockproteinbeta-2: VSREFCRTYVLPADVDPWRVRAALSHDGILNLEAPRGGRHLDTEVNEVYISLLPAPP-DP-EEEE-EAAI
Heatshockproteinbeta-1: ISRCFTRKYTLPPGVDPTQVSSSLSPEGTLTVEAP-MP-KLATQSNEITIPVTFESRAQLGGPEAAKSDE
Heatshockproteinbeta-8: VSKNFTKKIQLPAEVDPVTVFASLSPEGLLIIEAPQVP-PYST-FGE-S---SF-NN-EL--PQ--DSQE
Heatshockproteinbeta-7: VMNTFAHKCQLPEDVDPTSVTSALREDGSLTIRARRHP-H--T---E-H---V---Q-QT--F---RTE-
Heatshockproteinbeta-3: ISRSFTRQYKLPDGVEIKDLSAVLCHDGILVV--------------E-----V---K-D---P----VG-
Heatshockproteinbeta-9: MSQKVHRKM-LPSNLSPTAMTCCLTPSGQLWVRGQCVA--LAL--PEAQ-T-GPSPRLGS-LGS--KASN


Alpha-crystallinBchain: PK-K
Alpha-crystallinAchain: PS-S
Heatshockproteinbeta-6: -A-K
Heatshockproteinbeta-2: VE-P
Heatshockproteinbeta-1: TAAK
Heatshockproteinbeta-8: VTCT
Heatshockproteinbeta-7: IK-I
Heatshockproteinbeta-3: -T-K
Heatshockproteinbeta-9: LT-R
```

Résultats

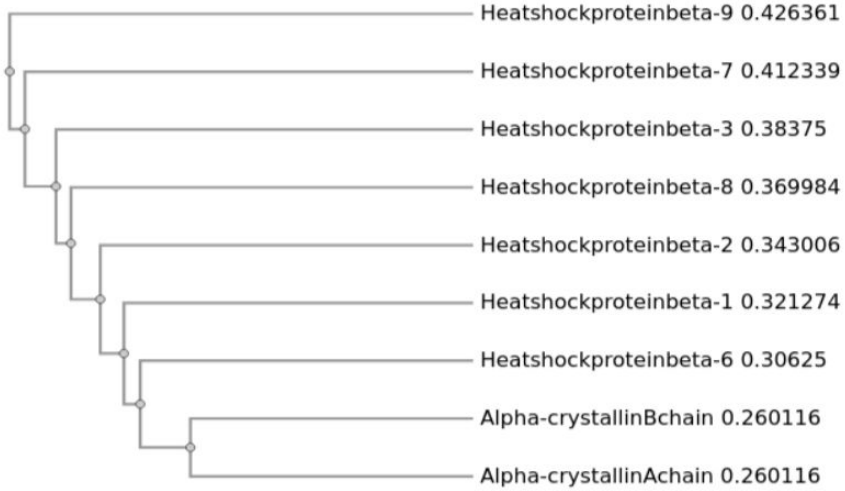
9 séquences : COMPLEXITÉ EN TEMPS DE $O(n^2)$	
175 acides aminés	14 secondes
355 acides aminés	37 secondes
860 acides aminés	2 minutes 19 secondes

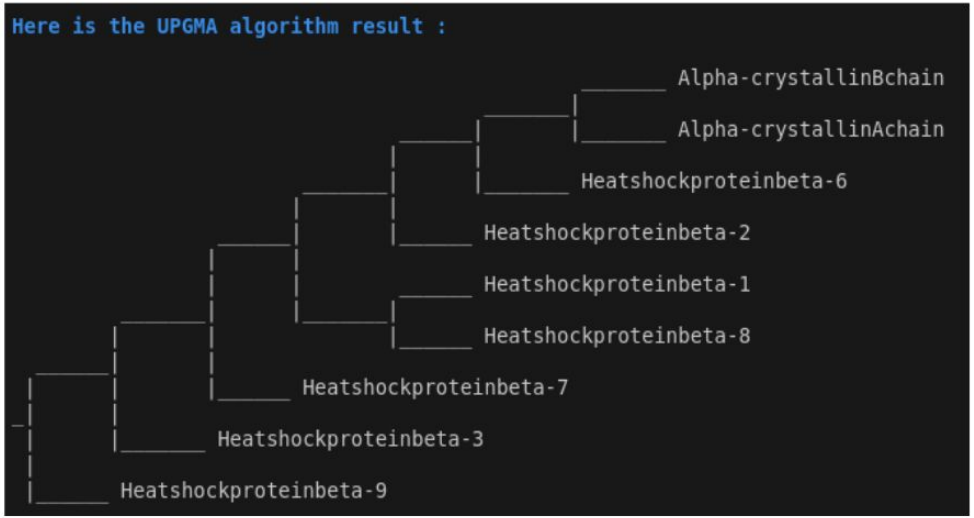
Résultats : Comparaison





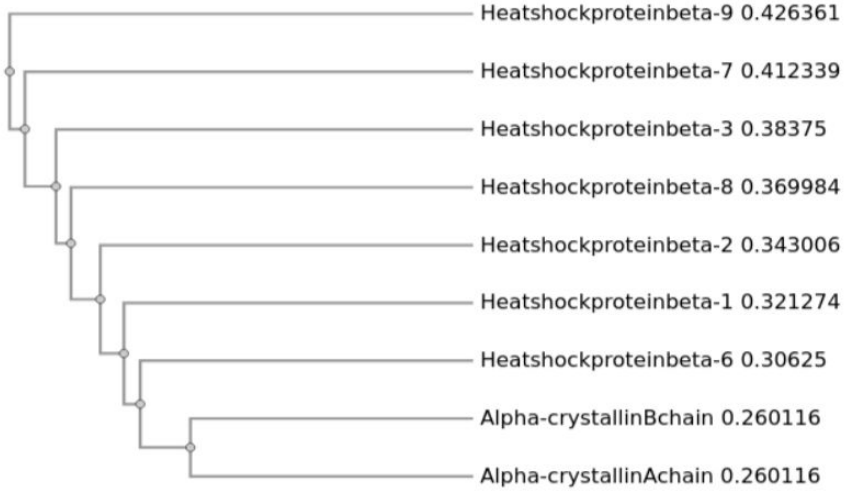
Résultats : Comparaison

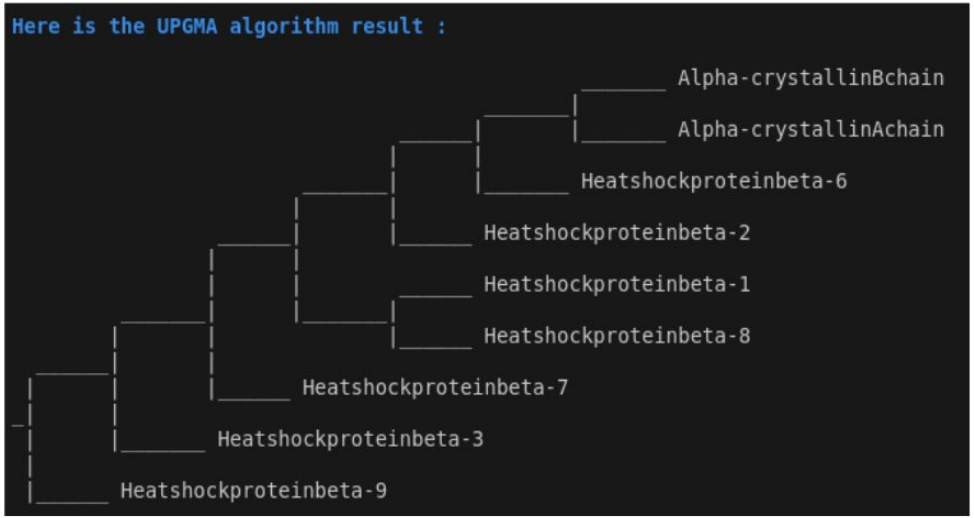




- Needleman-Wunsch VS Wilbur et Lipman

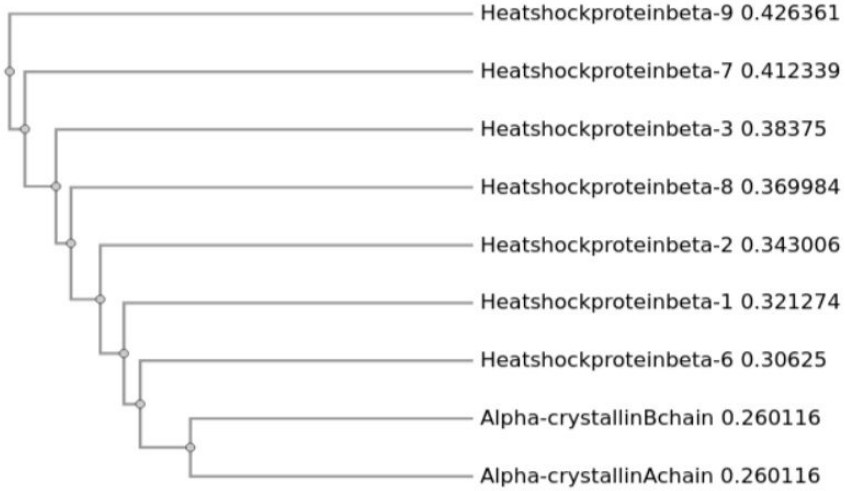
Résultats : Comparaison

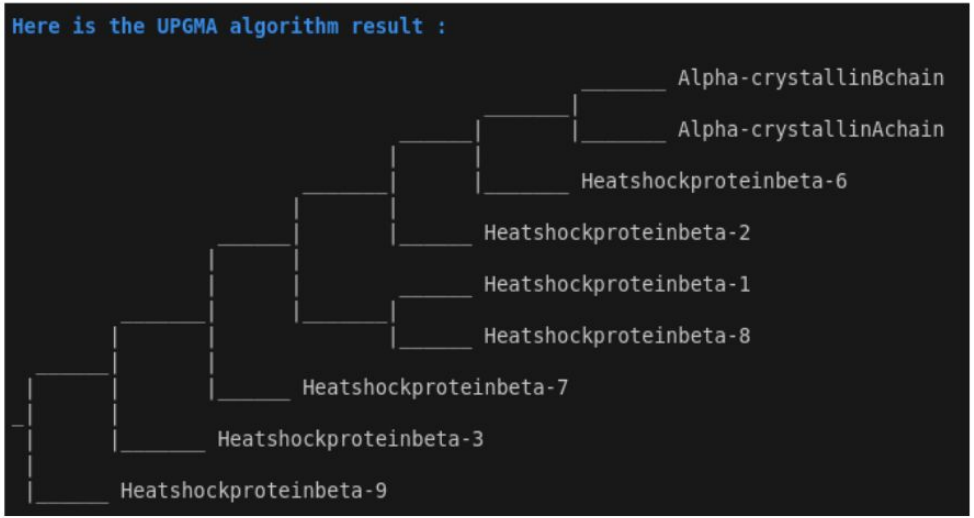




- Needleman-Wunsch VS Wilbur et Lipman
- pénalité de gap constante

Résultats : Comparaison





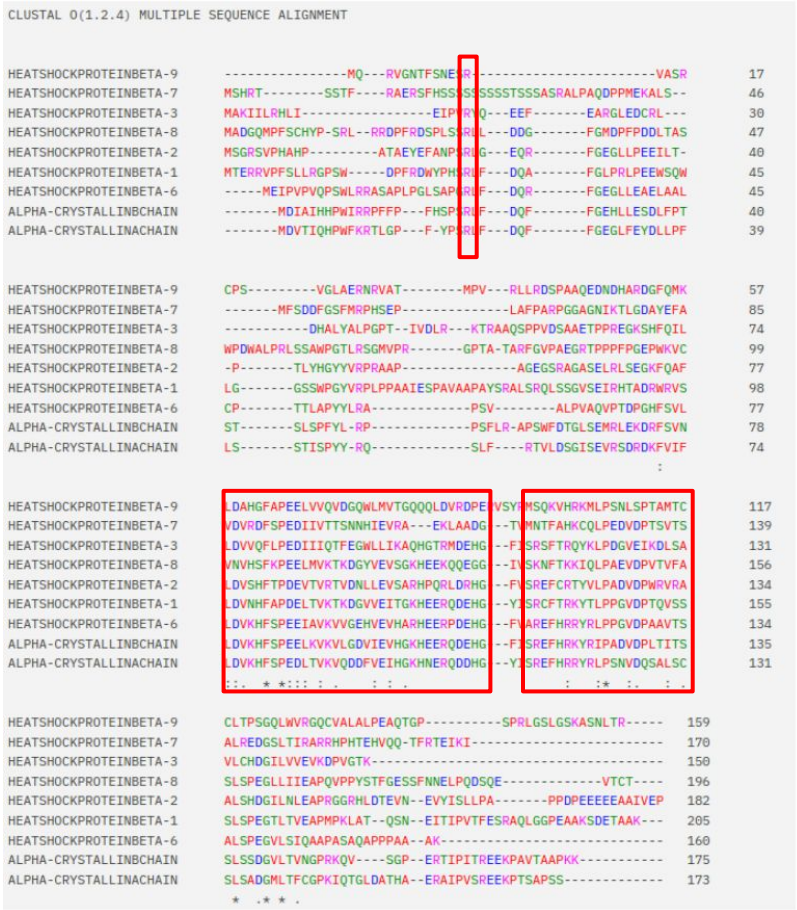
- Needleman-Wunsch VS Wilbur et Lipman
- pénalité de gap constante
- plusieurs méthodes de clusterisation (mBed, k-means, UPGMA)

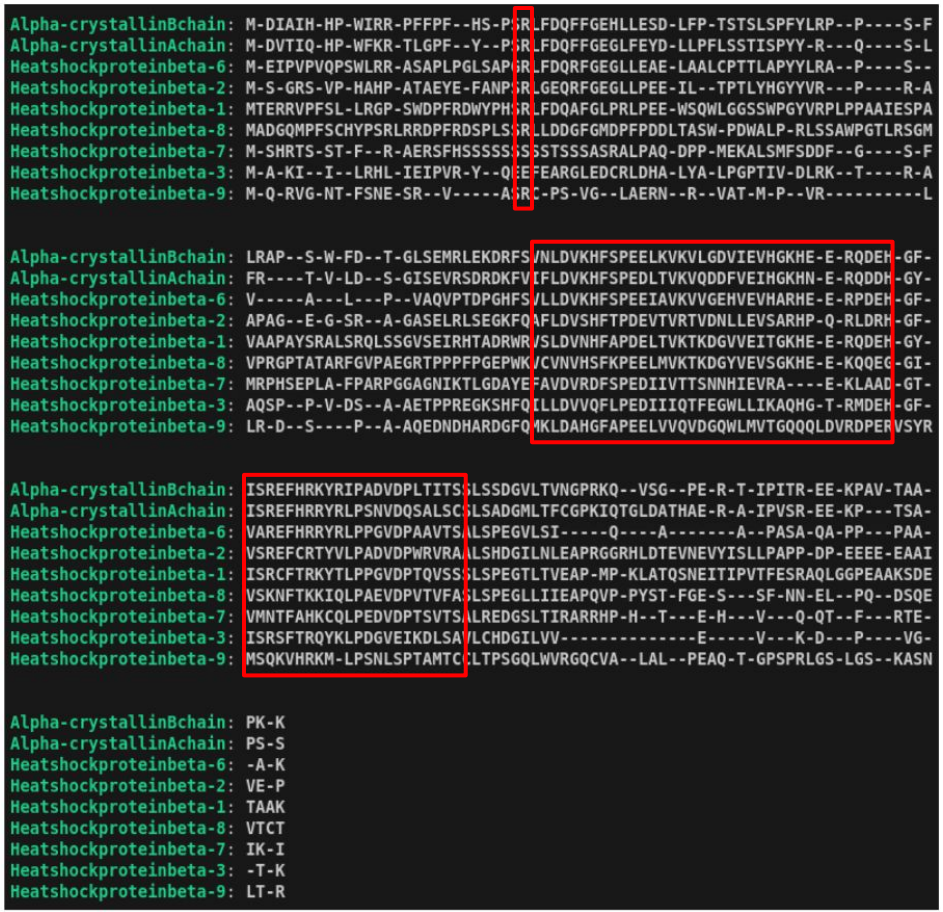
Résultats : Comparaison

```
CLUSTAL O(1.2.4) MULTIPLE SEQUENCE ALIGNMENT


HEATSHOCKPROTEINBETA-9      ----------------MQ---RVGNTFSNESR-----------------------VASR       17
HEATSHOCKPROTEINBETA-7      MSHRT--------SSTF----RAERSFHSSSSSSSSSTSSSASRALPAQDPPMEKALS--       46
HEATSHOCKPROTEINBETA-3      MAKIILRHLI-----------------EIPVRYQ---EEF-------EARGLEDCRL---       30
HEATSHOCKPROTEINBETA-8      MADGQMPFSCHYP-SRL--RRDPFRDSPLSSRLL---DDG-------FGMDPFPDDLTAS       47
HEATSHOCKPROTEINBETA-2      MSGRSVPHAHP---------ATAEYEFANPSRLG---EQR-------FGEGLLPEEILT-       40
HEATSHOCKPROTEINBETA-1      MTERRVPFSLLRGPSW-----DPFRDWYPHSRLF---DQA-------FGLPRLPEEWSQW       45
HEATSHOCKPROTEINBETA-6      -----MEIPVPVQPSWLRRASAPLPGLSAPGRLF---DQR-------FGEGLLEAELAAL       45
ALPHA-CRYSTALLINBCHAIN      -------MDIAIHHPWIRRPFFP---FHSPSRLF---DQF-------FGEHLLESDLFPT       40
ALPHA-CRYSTALLINACHAIN      -------MDVTIQHPWFKRTLGP---F-YPSRLF---DQF-------FGEGLFEYDLLPF       39


HEATSHOCKPROTEINBETA-9      CPS---------VGLAERNRVAT--------MPV---RLLRDSPAAQEDNDHARDGFQMK       57
HEATSHOCKPROTEINBETA-7      -------MFSDDFGSFMRPHSEP--------------LAFPARPGGAGNIKTLGDAYEFA       85
HEATSHOCKPROTEINBETA-3      ----------DHALYALPGPT--IVDLR---KTRAAQSPPVDSAAETPPREGKSHFQIL       74
HEATSHOCKPROTEINBETA-8      WPDWALPRLSSAWPGTLRSGMVPR-------GPTA-TARFGVPAEGRTPPPFPGEPWKVC       99
HEATSHOCKPROTEINBETA-2      -P-------TLYHGYYVRPRAAP---------------AGEGSRAGASELRLSEGKFQAF       77
HEATSHOCKPROTEINBETA-1      LG-------GSSWPGYVRPLPPAAIESPAVAAPAYSRALSRQLSSGVSEIRHTADRWRVS       98
HEATSHOCKPROTEINBETA-6      CP-------TTLAPYYLRA-------------PSV--------ALPVAQVPTDPGHFSVL       77
ALPHA-CRYSTALLINBCHAIN      ST-------SLSPFYL-RP-------------PSFLR-APSWFDTGLSEMRLEKDRFSVN       78
ALPHA-CRYSTALLINACHAIN      LS-------STISPYY-RQ-------------SLF----RTVLDSGISEVRSDRDKFVIF       74
                                                                                    :


HEATSHOCKPROTEINBETA-9      LDAHGFAPEELVVQVDGQWLMVTGQQQLDVRDPERVSYRMSQKVHRKMLPSNLSPTAMTC      117
HEATSHOCKPROTEINBETA-7      VDVRDFSPEDIIVTTSNNHIEVRA---EKLAADG---TVMNTFAHKCQLPEDVDPTSVTS      139
HEATSHOCKPROTEINBETA-3      LDVVQFLPEDIIIQTFEGWLLIKAQHGTRMDEHG---FISRSFTRQYKLPDGVEIKDLSA      131
HEATSHOCKPROTEINBETA-8      VNVHSFKPEELMVKTKDGYVEVSGKHEEKQQEGG---IVSKNFTKKIQLPAEVDPVTVFA      156
HEATSHOCKPROTEINBETA-2      LDVSHFTPDEVTVRTVDNLLEVSARHPQRLDRHG---FVSREFCRTYVLPADVDPWRVRA      134
HEATSHOCKPROTEINBETA-1      LDVNHFAPDELTVKTKDGVVEITGKHEERQDEHG---YISRCFTRKYTLPPGVDPTQVSS      155
HEATSHOCKPROTEINBETA-6      LDVKHFSPEEIAVKVVGEHVEVHARHEERPDEHG---FVAREFHRRYRLPPGVDPAAVTS      134
ALPHA-CRYSTALLINBCHAIN      LDVKHFSPEELKVKVLGDVIEVHGKHEERQDEHG---FISREFHRKYRIPADVDPLTITS      135
ALPHA-CRYSTALLINACHAIN      LDVKHFSPEDLTVKVQDDFVEIHGKHNERQDDHG---YISREFHRRYRLPSNVDQSALSC      131
                            ::.  * *::: : .    : : .                  :   :*  :.   : .


HEATSHOCKPROTEINBETA-9      CLTPSGQLWVRGQCVALALPEAQTGP----------SPRLGSLGSKASNLTR-----   159
HEATSHOCKPROTEINBETA-7      ALREDGSLTIRARRHPHTEHVQQ-TFRTEIKI-------------------------   170
HEATSHOCKPROTEINBETA-3      VLCHDGILVVEVKDPVGTK---------------------------------------   150
HEATSHOCKPROTEINBETA-8      SLSPEGLLIIEAPQVPPYSTFGESSFNNELPQDSQE------------VTCT----   196
HEATSHOCKPROTEINBETA-2      ALSHDGILNLEAPRGGRHLDTEVN--EVYISLLPA-------PPDPEEEEEAAIVEP   182
HEATSHOCKPROTEINBETA-1      SLSPEGTLTVEAPMPKLAT--QSN--EITIPVTFESRAQLGGPEAAKSDETAAK---   205
HEATSHOCKPROTEINBETA-6      ALSPEGVLSIQAAPASAQAPPPAA--AK-----------------------------   160
ALPHA-CRYSTALLINBCHAIN      SLSSDGVLTVNGPRKQV----SGP--ERTIPITREEKPAVTAAPKK-----------   175
ALPHA-CRYSTALLINACHAIN      SLSADGMLTFCGPKIQTGLDATHA--ERAIPVSREEKPTSAPSS------------   173
                             *  .* * .
```

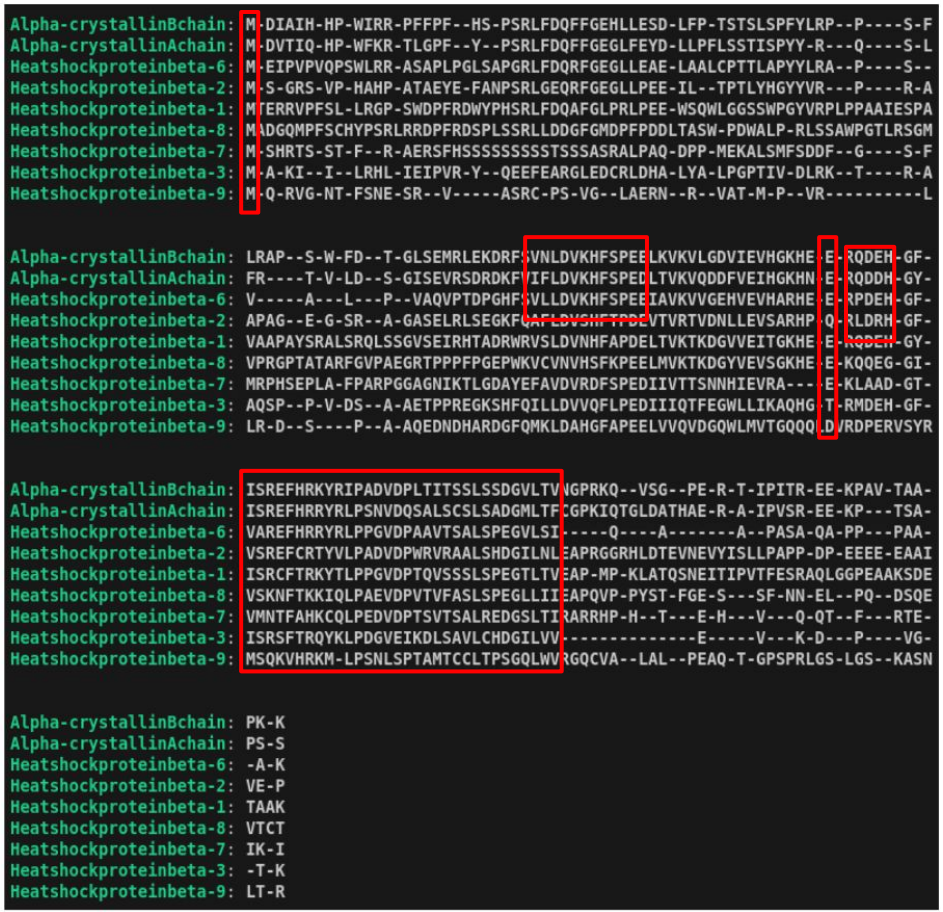
```
Alpha-crystallinBchain: M-DIAIH-HP-WIRR-PFFPF--HS-PSRLFDQFFGEHLLESD-LFP-TSTSLSPFYLRP--P----S-F
Alpha-crystallinAchain: M-DVTIQ-HP-WFKR-TLGPF--Y--PSRLFDQFFGEGLFEYD-LLPFLSSTISPYY-R---Q----S-L
Heatshockproteinbeta-6: M-EIPVPVQPSWLRR-ASAPLPGLSAPGRLFDQRFGEGLLEAE-LAALCPTTLAPYYLRA--P----S--
Heatshockproteinbeta-2: M-S-GRS-VP-HAHP-ATAEYE-FANPSRLGEQRFGEGLLPEE-IL--TPTLYHGYYVR---P----R-A
Heatshockproteinbeta-1: MTERRVPFSL-LRGP-SWDPFRDWYPHSRLFDQAFGLPRLPEE-WSQWLGGSSWPGYVRPLPPAAIESPA
Heatshockproteinbeta-8: MADGQMPFSCHYPSRLRRDPFRDSPLSSRLLDDGFGMDPFPDDLTASW-PDWALP-RLSSAWPGTLRSGM
Heatshockproteinbeta-7: M-SHRTS-ST-F--R-AERSFHSSSSSSSSSTSSSASRALPAQ-DPP-MEKALSMFSDDF--G----S-F
Heatshockproteinbeta-3: M-A-KI--I--LRHL-IEIPVR-Y--QEEFEARGLEDCRLDHA-LYA-LPGPTIV-DLRK--T----R-A
Heatshockproteinbeta-9: M-Q-RVG-NT-FSNE-SR--V-----ASRC-PS-VG--LAERN--R--VAT-M-P--VR----------L


Alpha-crystallinBchain: LRAP--S-W-FD--T-GLSEMRLEKDRFSVNLDVKHFSPEELKVKVLGDVIEVHGKHE-E-RQDEH-GF-
Alpha-crystallinAchain: FR----T-V-LD--S-GISEVRSDRDKFVIFLDVKHFSPEDLTVKVQDDFVEIHGKHN-E-RQDDH-GY-
Heatshockproteinbeta-6: V-----A---L---P--VAQVPTDPGHFSVLLDVKHFSPEEIAVKVVGEHVEVHARHE-E-RPDEH-GF-
Heatshockproteinbeta-2: APAG--E-G-SR--A-GASELRLSEGKFQAFLDVSHFTPDEVTVRTVDNLLEVSARHP-Q-RLDRH-GF-
Heatshockproteinbeta-1: VAAPAYSRALSRQLSSGVSEIRHTADRWRVSLDVNHFAPDELTVKTKDGVVEITGKHE-E-RQDEH-GY-
Heatshockproteinbeta-8: VPRGPTATARFGVPAEGRTPPPFPGEPWKVCVNVHSFKPEELMVKTKDGYVEVSGKHE-E-KQQEG-GI-
Heatshockproteinbeta-7: MRPHSEPLA-FPARPGGAGNIKTLGDAYEFAVDVRDFSPEDIIVTTSNNHIEVRA----E-KLAAD-GT-
Heatshockproteinbeta-3: AQSP--P-V-DS--A-AETPPREGKSHFQILLDVVQFLPEDIIIQTFEGWLLIKAQHG-T-RMDEH-GF-
Heatshockproteinbeta-9: LR-D--S----P--A-AQEDNDHARDGFQMKLDAHGFAPEELVVQVDGQWLMVTGQQQLDVRDPERVSYR


Alpha-crystallinBchain: ISREFHRKYRIPADVDPLTITSSLSSDGVLTVNGPRKQ--VSG--PE-R-T-IPITR-EE-KPAV-TAA-
Alpha-crystallinAchain: ISREFHRRYRLPSNVDQSALSCSLSADGMLTFCGPKIQTGLDATHAE-R-A-IPVSR-EE-KP---TSA-
Heatshockproteinbeta-6: VAREFHRRYRLPPGVDPAAVTSALSPEGVLSI-----Q----A-------A--PASA-QA-PP---PAA-
Heatshockproteinbeta-2: VSREFCRTYVLPADVDPWRVRAALSHDGILNLEAPRGGRHLDTEVNEVYISLLPAPP-DP-EEEE-EAAI
Heatshockproteinbeta-1: ISRCFTRKYTLPPGVDPTQVSSSLSPEGTLTVEAP-MP-KLATQSNEITIPVTFESRAQLGGPEAAKSDE
Heatshockproteinbeta-8: VSKNFTKKIQLPAEVDPVTVFASLSPEGLLIIEAPQVP-PYST-FGE-S---SF-NN-EL--PQ--DSQE
Heatshockproteinbeta-7: VMNTFAHKCQLPEDVDPTSVTSALREDGSLTIRARRHP-H--T---E-H---V---Q-QT--F---RTE-
Heatshockproteinbeta-3: ISRSFTRQYKLPDGVEIKDLSAVLCHDGILVV--------------E-----V---K-D---P----VG-
Heatshockproteinbeta-9: MSQKVHRKM-LPSNLSPTAMTCCLTPSGQLWVRGQCVA--LAL--PEAQ-T-GPSPRLGS-LGS--KASN


Alpha-crystallinBchain: PK-K
Alpha-crystallinAchain: PS-S
Heatshockproteinbeta-6: -A-K
Heatshockproteinbeta-2: VE-P
Heatshockproteinbeta-1: TAAK
Heatshockproteinbeta-8: VTCT
Heatshockproteinbeta-7: IK-I
Heatshockproteinbeta-3: -T-K
Heatshockproteinbeta-9: LT-R
```

Résultats : Comparaison

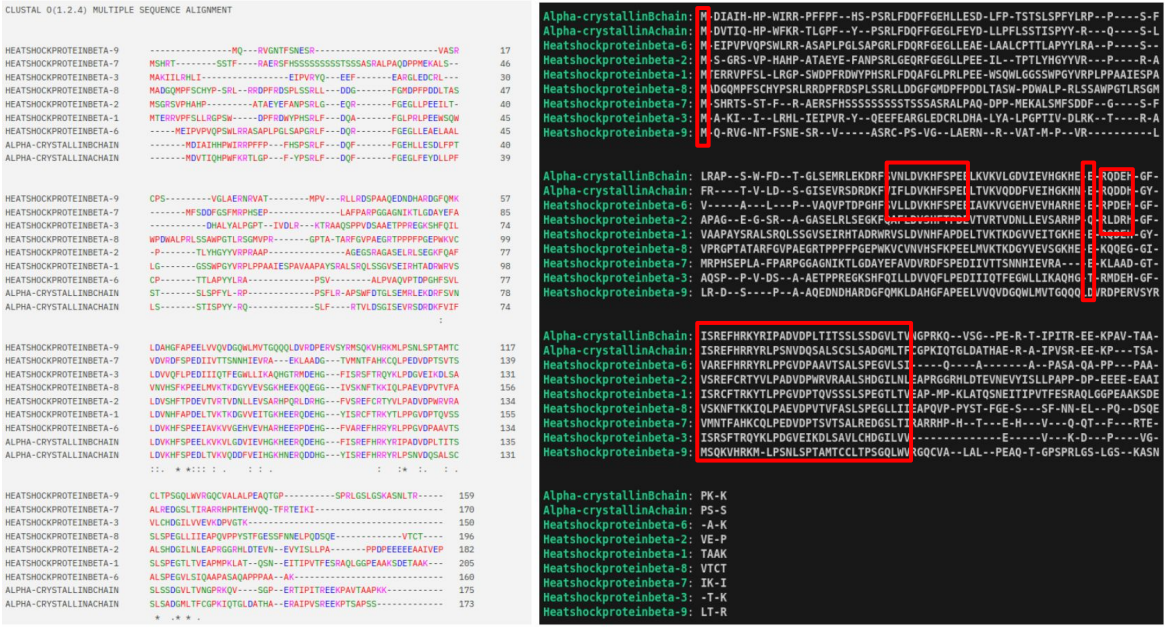




Résultats : Comparaison

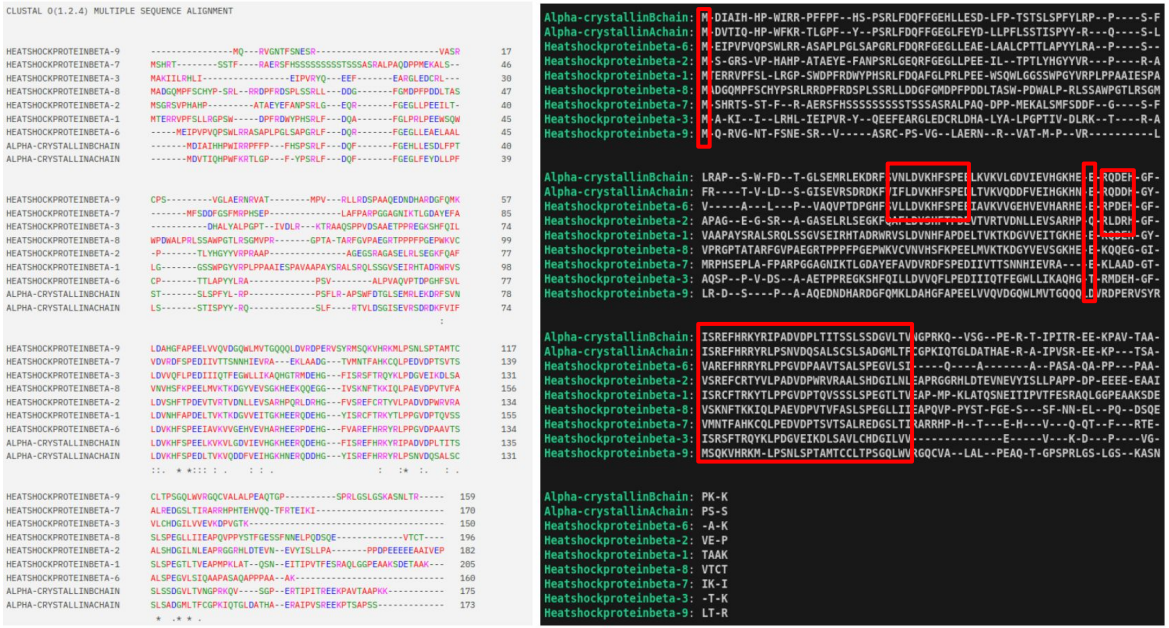


Résultats : Comparaison



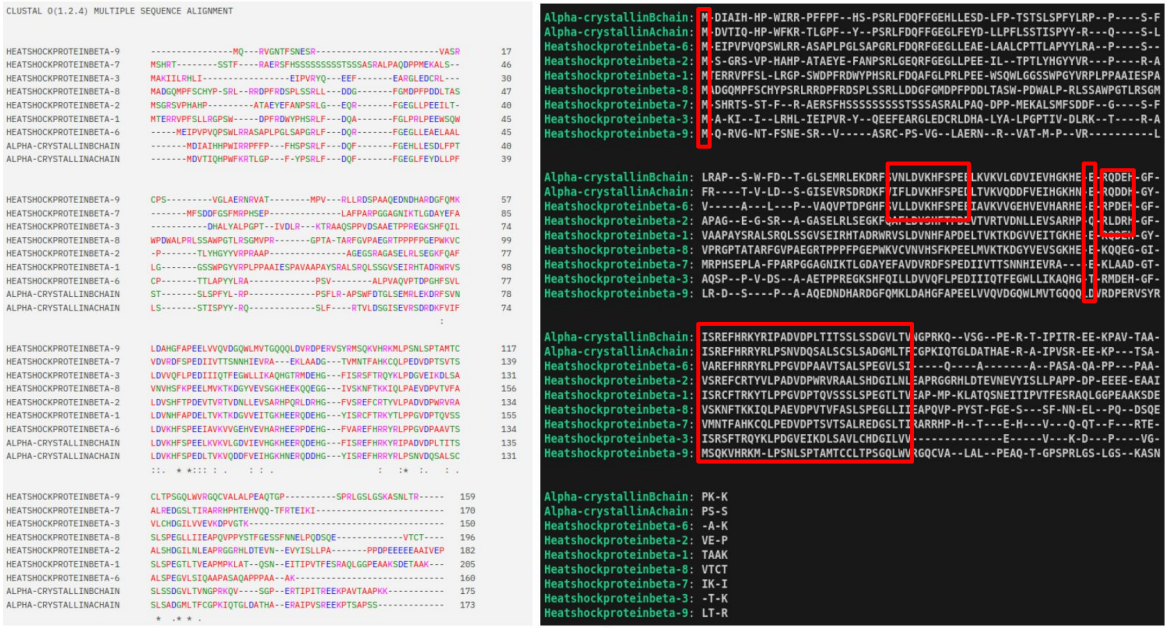
- Arbre guide différent

Résultats : Comparaison



- Arbre guide différent
- Modèle de Markov caché

Résultats : Comparaison



- Arbre guide différent
- Modèle de Markov caché
- pénalité de gap constante

Conclusion

- Trois étapes implémentées

Conclusion

- Trois étapes implémentées
- Complexité temporelle

Conclusion

- Trois étapes implémentées

- Complexité temporelle

- Des résultats satisfaisants