**Pràctica alineament múltiple**

**Objectiu:** Amplificar el gen *rpfF* d'una nova espècie bacteriana de la família *Xanthomonadaceae* que se sospita pertany al gènere *Xylella sp.*

**Metodologia:** Construir un alineament múltiple amb les seqüències de les proteïnes RpfF amb activitat enoyl-CoA hydratase *rpfF*[gen] **en diverses espècies de la família** ***Xanthomonadaceae***. Trobar regions conservades i dissenyar primers mitjançant traducció reversa.

1. Obtenir del NCBI **6 seqüències de proteïnes en format FASTA** dels microorganismes següents:

*Stenotrophomonas maltophilia* D457

*Stenotrophomonas acidaminiphila*

*Lysobacter enzymogenes*

*Xylella fastidiosa* Temecula1

*Xanthomonas campestris pv. raphani* 756C

*Xanthomonas translucens pv. poae*

1. Preneu les 6 proteïnes i les introduïu en un **únic** fitxer FASTA. Editar els descriptors de cada seqüència que hagueu introduït en el arxiu fasta. Els primers caràcters d'aquest nom han de ser diferents, i s'han d'evitar els espais i claudàtors. Per exemple:

>L\_enzymogenes

>S\_acidaminiphila

>S\_maltophilia

>X\_campestris

>X\_translucens

>X\_fastidiosa

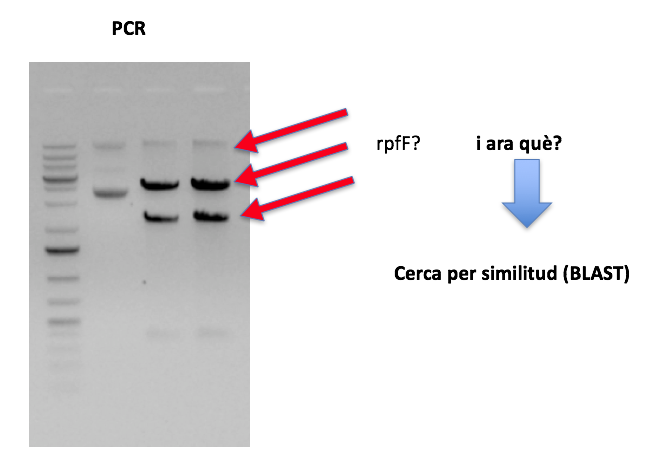
1. Executar ClustalW a: <https://www.genome.jp/tools-bin/clustalw> ). Paràmetres predefinits. **Comentar paràmetres**: Output format. Pairwaise alignment FAST vs SLOW (plana 5 pdf de teoria alineaments múltiples).Penalitzacions per gap "open" i gap "extension" (planes 23 i 24 ppt teoria alineaments per parelles)
2. **Comentar** resultats: Matriu de distàncies: seqüències més properes i més allunyades; **posicions** a l'alineament conservades o semi-conservades, presència de *gaps*, identificar **motius** conservats.
3. Identifica aquest **motiu** al alineament: Qr code

   Description automatically generated
4. **Guardar** localment l’arxiu clustalw.aln i canviar-li el nom a **multiple\_RffF.aln**. (obrir amb la "llibreta" (AKA notepad.exe). **NO** obrir cap arxiu amb cap altre programari).
5. Visualitzar l'arbre guia:
   1. descarrega't l'arxiu clustalw.dnd i canviar-li el nom a **multiple\_RffF.dnd**. i obre'l amb la "llibreta" (AKA notepad.exe). Això que estàs veient és un arbre en format "**newick**". (també ho tens a la plana de resultats del clustalw)
   2. ves a <http://etetoolkit.org/treeview/> Neteja el quadre de l'esquerra, enganxa el contingut de multiple\_RffF.dnd i apreta "View tree"
   3. Compara l'arbre que estàs visualitzant amb les puntuacions de la matriu de distàncies i amb l'ordre de les seqüències dins l'alineament múltiple.
6. Visualitzar l'alineament en format clustalw (el què et dona el programa) amb el jalviewJS ( <https://www.jalview.org/jalview-js/JalviewJS/> )
   1. Obre el link (si et dona aquest error: "Error connecting to server: https://chemapps.stolaf.edu/jmol/... Apreta OK i segueix)
   2. File->"Input alignment"->"From textbox". Enganxa l'alinement múltiple en format clustal. Prem "new window" i després "Close"
   3. Si no funciona el "paste". File->"Input alignment"->"From File". Puja el fitxer multiple\_RffF.aln descarregat anteriorment.
   4. Colour->"Percentage Identity" i/o Colour->"BLOSUM62 Score" (tu mateix)
   5. Ves a la posició 80 de l'alineament, passa el cursor del mouse per sobre dels AA d'aquesta posició i observa quina posició són dins de cada seqüència.
7. A l’alineament en format clustal identificar la (les) possible(s) seqüències consens de **regions conservades** (6 aminoàcids, 5 d'ells 100% conservats). Identifica la més propera al N terminal i la més propera al C terminal. Guardar-la(s) en format FASTA.

ELS SEGUENTS PASSOS NO ELS FAREM.

Però caldrien per a obtenir els primers per a fer la PCR.

1. Fer la traducció reversa amb el [Reverse Translate](http://www.bioinformatics.org/sms2/rev_trans.html) del SMS adaptant-la a la taula d’usos de codons a la espècie *Xylella fastidiosa*. Tenim dos opcions:
   1. Agafar la taula de l'espècie representativa *Xylella fastidiosa 9a5c* que obtindràs de [Codon Usage Database](https://www.kazusa.or.jp/codon/). (**Format**: Codi 11 Eubacterial. **Botó**: A style like CodonFrequency output in GCG Wisconsin PackageTM)
   2. Construir una taula d'usos de codons amb el [Codon Usage](http://www.bioinformatics.org/sms2/codon_usage.html) del SMS i un fragment (d’1 a 100.000) del genoma de *Xylella fastidiosa Temecula1* (NC\_004556). Per construir aquest taula veure el tutorial "Construcció Taula d’ús codons" al final d'aquest document.
2. Anotar les seqüències “més degenerades” i “més probables” per ambdós primers.
3. Al laboratori, fer la PCR amb els primers degenerats y seqüenciar els fragments de DNA obtinguts.

****

**Resultats**

**Cerca al NCBI protein:**

(Stenotrophomonas maltophilia D457[PORG] OR Stenotrophomonas acidaminiphila[PORG] OR Lysobacter enzymogenes[PORG] OR Xylella fastidiosa Temecula1[PORG] OR Xanthomonas campestris pv. raphani 756C[PORG] OR Xanthomonas translucens pv. poae[PORG]) AND rpff[gene]

**Alineament múltiple amb ClustalW**

S\_acidaminiphila MNAIELLPVGTPATSYAPPYSTIRVHADAAGDVHWLHMHADAGPGVRPCFRKPLMTEMHA

S\_maltophilia MSTIEKLPS------SGSPFATIRTEDSADGAAHWLFMHADAATGIRPCCRKDMLDEMWS

L\_enzymogenes MSTIEKLHR-------VRAYPTLRIESAPDGLAHWLYMHEDVGLGVRPCFRTPLMEDMWA

X\_translucens MSMIDTLPR-------SRSYSTIRVDSDGAKDSHWLYMHAGRQPGTHPCFQSQMLDDVLA

X\_fastidiosa MSAVHPIPH-------PICESSIRIIEETHRNVYWIYMHAHLARTTGAAYFSLKLIDDIM

X\_campestris MSAVQPFIR-------TNIGSTLRIIEEPQRDVYWIHMHADLAINPGRACFSTRLVDDIT

\*. :. : .::\* :\*:.\*\* . . : :

S\_acidaminiphila YLDGITGDPQARPE-RLRHLVLASAADAFNLGGDLDVFSALIRARDGDGLLRYARSCVDG

S\_maltophilia YMAAITRSPAERHNGTLRHFVLASDAVAYNLGGDLDLFTRLIREGNRDLLLNYAQRCVEG

L\_enzymogenes FLSSISLREGQAPS-ELRHVVLASSAPAFNLGGDLELFSRLIRSQDRERLLSYAQRCVDG

X\_translucens ALSALTSQELQQQKLSLKHLVIASDANVFNLGGDLALFSDLIRRQDRNQLLNYAYRCVTG

X\_fastidiosa NYQSVLRQRLKEQTVQLPFVVLASDSNVFNLGGDLQLFCDLIRRKEREALLDYACRCVRG

X\_campestris GYQTNLGQRLNTAGVLAPHVVLASDSDVFNLGGDLALFCQLIREGDRARLLDYAQRCVRG

..\*:\*\* : .:\*\*\*\*\*\* :\* \*\*\* : \*\* \*\* \*\* \*

S\_acidaminiphila VYRFHAGLDGSFRTIALLQGDALGGGLEMALSCHTIVAEEGVGMGLPEVLFGLFPGMGAY

S\_maltophilia VHHLHTGFGGDVRSIALIQGDALGGGLEMALACHTIVAEEGSGLGLPEVLFGLFPGMGAY

L\_enzymogenes VHHIHGGLGGDVHTIALVQGDALGGGLELALSCHTIVAEEGVDMGLPEVVFDLFPGMGAY

X\_translucens VHQLYRGLAGDVRTIALVQGDALGGGLEIALACHTIVAEEGVTMGLPEVMFGLFPGMGAY

X\_fastidiosa AYAFHAGLNANVHSIALLQGNALGGGFEAALCCHTIVAEEGVMMGFPEVLFDLFPGMGAY

X\_campestris VHAFHVGLGARAHSIALVQGNALGGGFEAALSCHTIIAEEGVMMGLPEVLFDLFPGMGAY

.: :: \*: . ::\*\*\*:\*\*:\*\*\*\*\*:\* \*\*.\*\*\*\*:\*\*\*\* :\*:\*\*\*:\*.\*\*\*\*\*\*\*\*

S\_acidaminiphila NFLSRRVSPQLAERIILEGRVYSAEEMHAMGVVDILVPRGQGEQAVNDLIRQQQRIPQSY

S\_maltophilia SFLCRRVSPHLAEKIILDGRVYSAEEMHAMGVVDVLVKKGEGRAAVEELIRQQQRTPQSY

L\_enzymogenes SYLCKRVSPRVAEKLIMDGALLSSEEMHRLGVVDVLVPRGQGEAAVADLIRRQQRSPHSQ

X\_translucens SFLCKRMPPQLAEKIILEGNLYSSEQMHAMGIVDVLVPRNQGAAAVSELIRNQRRNPHSY

X\_fastidiosa SFMRQRISPKLAERLILEGNLYSSEELLAIGLIDKVVPRGKGIEAVEQIIRDSKRRQYTW

X\_campestris SFMCQRISAHLAQKIMLEGNLYSAEQLLGMGLVDRVVPRGQGVAAVEQVIRESKRTPHAW

.:: :\*:..::\*::::::\* : \*:\*:: :\*::\* :\* :.:\* \*\* ::\*\* .:\* :

S\_acidaminiphila LAMNAVRGISNRVDYQELMDITRVWVDAALALDEKSLRTMDRLVKAQTRRSGQVLS-

S\_maltophilia LAMNAARAIAQPVSYDELLEITKVWVDSALALGDRSLRTMDRLIKAQTRRASLDAA-

L\_enzymogenes LAMNKMRQIAQPTSYQELMAITELWVDTALALPDKALRTMDRIVRAQERRAVAA---

X\_translucens LAMNAARNLAQPVSLQELKAITEVWVDSALALEEKSLRTMNRLVRAQTRRLQDVAA-

X\_fastidiosa AAMQEVKKIAHEVSLEEMIRITELWVDSALKLSNKSLRTMERLIRAQQTHKNTALKN

X\_campestris AAMQQVREMTTAVPLEEMMRITEIWVDTAMQLGEKSLRTMDRLVRAQSRRSGLDAG-

\*\*: : :: . :\*: \*\*.:\*\*\*:\*: \* :::\*\*\*\*:\*:::\*\* :

Regions conservades per al disseny de primers

**Arbre guia de ClustalW**

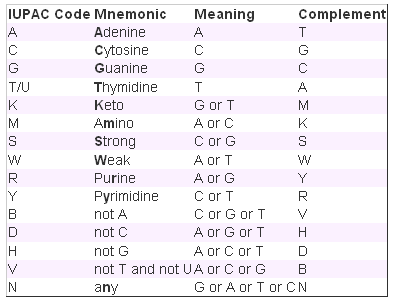
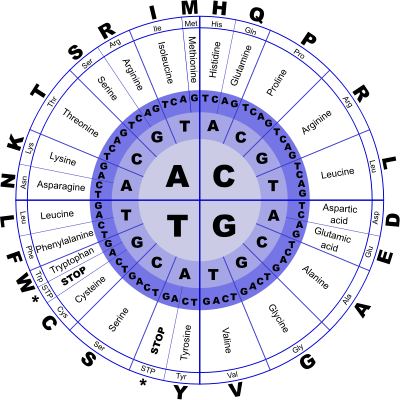
**A diagram of a dna

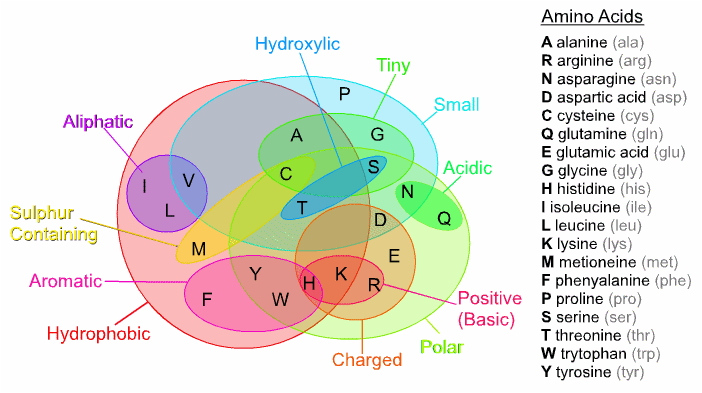
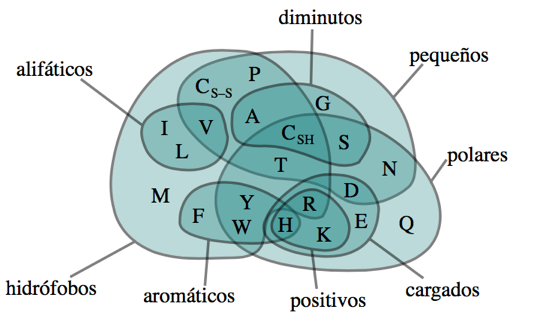
Description automatically generated with medium confidence**

**A screenshot of a graph

Description automatically generated**

**Taules complementaries**





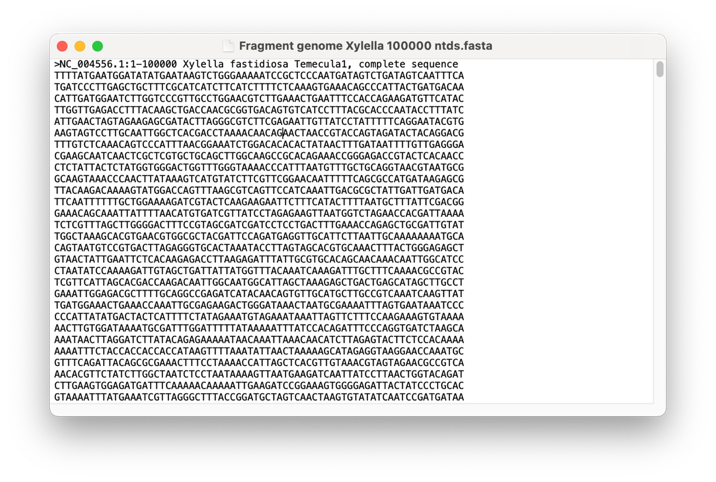
**Construcció Taula d’ús codons**

Aquesta activitat es farà usant la [Sequence Manipulation Suite](https://www.bioinformatics.org/sms2/index.html)

1. Descarregar un fragment de 100.000 bps del genoma NC\_004556 del NCBI en format FASTA

Graphical user interface, text, application, email

Description automatically generated



1. Fer una cerca d’ORFs amb el [ORF Finder](https://www.bioinformatics.org/sms2/orf_find.html) amb les condicions següents:

Graphical user interface, text, application

Description automatically generated

1. Guardar la predicció en un fitxer txt

Graphical user interface, text

Description automatically generated

1. Editar el fitxer anterior eliminant les seqüències d’aminoàcids predites. Deixar **sols** les seqüencies de nucleòtids en els seu format FASTA i guardar novament el fitxer en format txt.

Text

Description automatically generated

1. Combinar les seqüències FASTA de l’arxiu anterior amb [Combine Fasta](https://www.bioinformatics.org/sms2/combine_fasta.html) del SMS i guardar el resultat en un fitxer txt.

Graphical user interface, text, application

Description automatically generated

1. Comprovar que el nou fitxer conté una seqüència única en format FASTA.

Graphical user interface, text

Description automatically generated

1. Fer una predicció d’ús de codons amb [Codon Usage](https://www.bioinformatics.org/sms2/codon_usage.html) del SMS

Graphical user interface, text, application

Description automatically generated

1. Guardar el resultat en un fitxer txt

Table

Description automatically generated