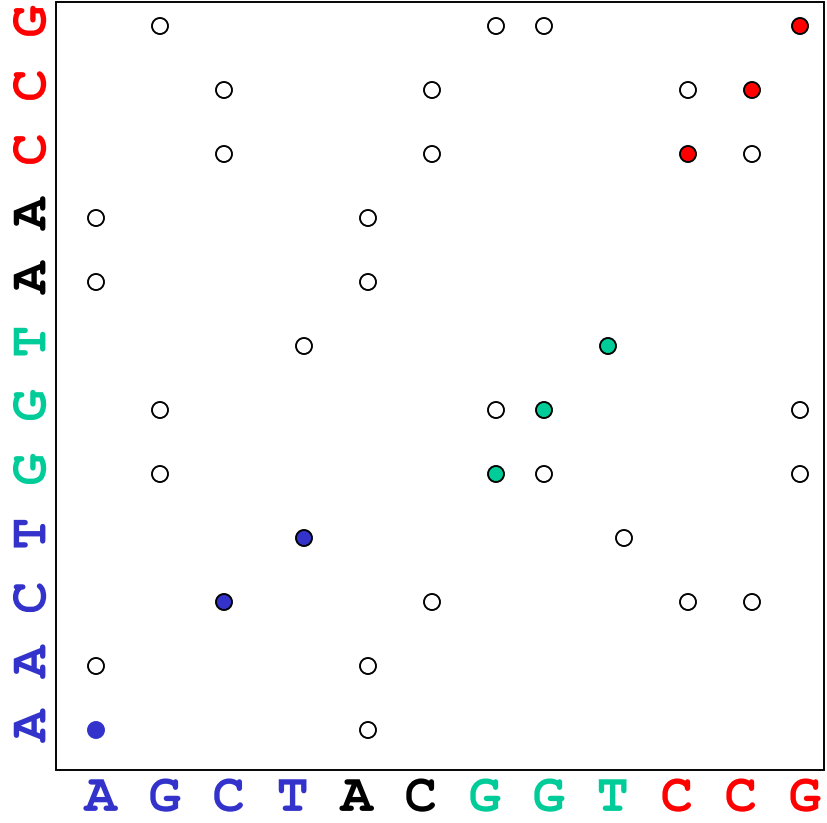


Recherche par similarité dans les banques/bases de données
La suite Blast (Basic Local Alignment Search Tool)



Algorithme de Blast (version 1)

(Altschul *et al.* (1990) *J. Mol. Biol.*, 215, 403-410)

- ne recherche que les diagonales
- > établir un score pour chaque paire de segments comparée (diagonale)
(séquences ADN : $Se(\text{identité}) = 5$ et $Se(\text{mismatch}) = -4$
séquences protéiques -> matrice de substitution)
- > MSP : Maximal Segment Pair correspond à la paire de segments (diagonale) qui possède le score le plus élevé comparé au score de toutes les autres paires de segments.
- > HSP : High Scoring segment Pair définit une paire de segments comme localement maximale, c'est-à-dire que l'on ne pourra pas augmenter son score soit en la rallongeant, soit en la raccourcissant.
 - > problème : trop de paires de segments, on ne peut pas toutes les retenir.
 - > définition d'un seuil S_{HSP} : Blast ne retient comme HSP que les paires de segments dont le score est supérieur à S_{HSP} .

Algorithme de Blast (version 1)

(Altschul *et al.* (1990) *J. Mol. Biol.*, 215, 403-410)

-> Définition de la valeur du seuil S_{HSP} :

Résultats en statistique de Karling *et al.* permettent d'estimer le score le plus élevé que peut avoir une MSP entre deux séquences dont la similarité ne serait due qu'au hasard

-> ce score est utilisé comme seuil S_{HSP}

-> Sélection des HSP

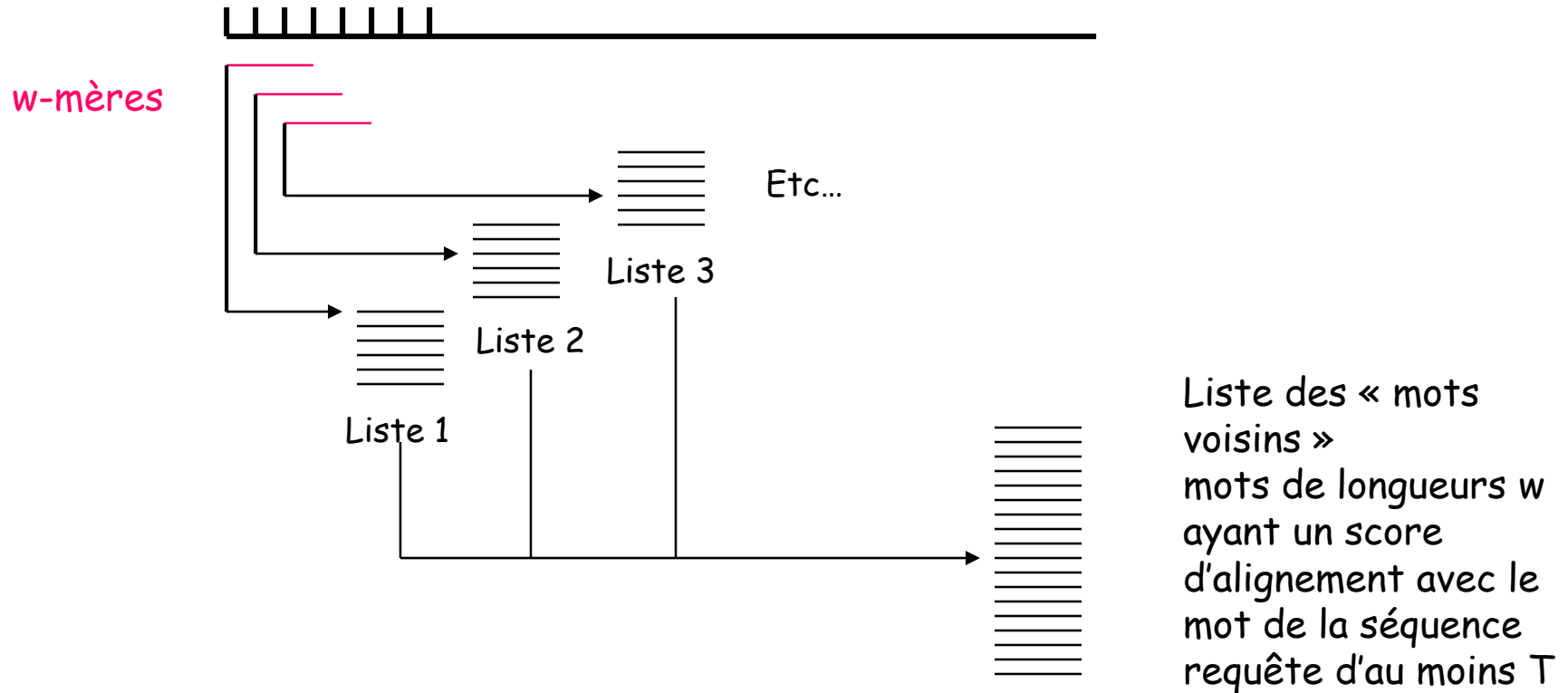
- Paire de segments dont le score est $<$ à S_{HSP} , similarité non significative due au hasard. Cette paire n'est pas retenue.
- Paire de segments dont le score est \geq à S_{HSP} , similarité due à une histoire évolutive commune -> séquence apparentée -> retenue comme HSP

Algorithme de Blast (version 1)

(Altschul *et al.* (1990) *J. Mol. Biol.*, 215, 403-410)

1ère étape : Établissement d'une liste de « mots voisins »

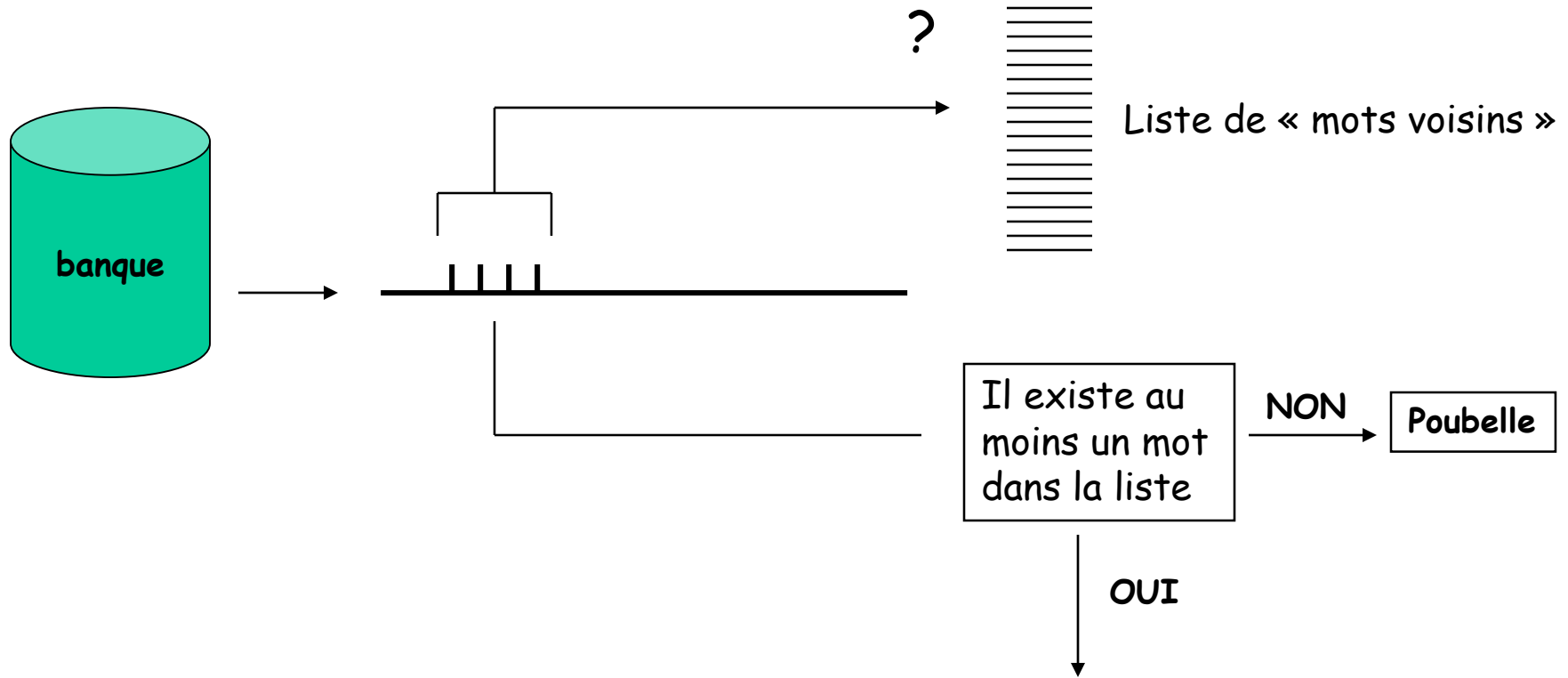
Séquence requête Q



w-mères : mot de taille *w* (*w* = 3 pour les protéines et *w* = 11 pour les séquences d'acides nucléiques)

Algorithme de Blast (version 1)

2ème étape : Recherche de la présence de ces mots dans les séquences de la banque



Extension de l'alignement : 3ème étape

Algorithme de Blast (version 1)

3ème étape : si un « hit » extension de l'alignement

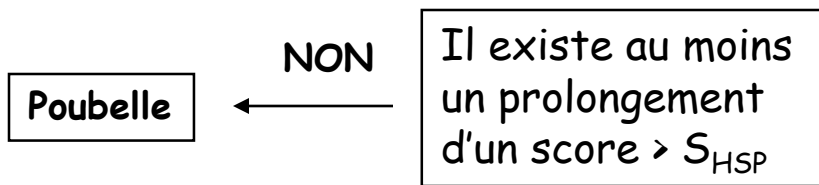
Séquence requête Q



Séquence banque B



Essayer d'étendre l'alignement dans les deux sens (arrêt de l'extension quand le score obtenu décroît au minimum d'une valeur X fixée (drop-off score))



Meilleur score obtenu par prolongement d'un hit

Score séquence B de la banque

4ème étape : Calcul de la p-value

Statistique Poisson

Classement des séquences de la banque

Normalisation du score

Le score normalisé S' pour un HSP est donné par :

$$S' = \frac{\lambda S - \ln K}{\ln 2}$$

Avec :

S score du HSP

λ et K paramètres calculés à partir du système de score et de la composition des séquences

S' est exprimé en une unité appelée *bits*

La p-value

La p-value est la probabilité qu'il existe au moins un HSP, obtenu lors de la comparaison de deux séquences aléatoires (de même longueur et composition que les séquences d'intérêt), dont le score soit supérieur ou égal à celui du HSP issu de la comparaison des séquences d'intérêt.

La p-value

On compare une séquence de longueur n avec une banque de données de longueur m , avec une matrice de comparaison donnée.

On s'intéresse à la distribution du nombre $N(S)$ d'alignements dont le score dépasse la valeur S . Cette distribution suit approximativement une loi de Poisson. On a donc l'expression suivante pour la probabilité qu'il existe un alignement fortuit de score $\geq S$:

$$p(\text{score} \geq S) = 1 - e^{-E(S)}$$

$E(S)$ étant l'espérance mathématique de $N(S)$ et qui s'exprime par :

$$E(S) = Knme^{-\lambda S} = E\text{-value} = \text{nombre attendu d'alignements ayant un score au moins égal à } S$$

Avec λ et K paramètres calculés à partir du système de score et de la composition des séquences

Dans le cas des alignements locaux sans gap, la théorie de Karlin et Altschul fournit donc des formules analytiques dont les paramètres dépendent de ceux de la recherche.

Query= Bsub.RbsA
(493 letters)

Database: Coli
4289 sequences; 1,358,990 total letters

Searching.....done

Sequences producing significant alignments:	Score (bits)	E Value	N
EcolA01.RBSA "ATP-binding component of D-ribose high-affini...	270	e-141	3
EcolA01.MGLA "ATP-binding component of methyl-galactoside t...	164	e-124	4
EcolA01.YJWC "putative ATP-binding component of a transport...	145	e-121	4
EcolA01.XYLG "putative ATP-binding protein of xylose transp...	157	e-120	6

>EcolA01.RBSA "ATP-binding component of D-ribose high-affinity transport system"
Length = 501

Score = 270 bits (587), Expect(3) = e-141
Identities = 115/258 (44%), Positives = 167/258 (64%)

Query: **3** IEMKDIHKTFGKNQVLSGVSFQLMPGEVHALMGENGAGKSTLMNLTGLHKADKQISIN 62
 +++K I K F + LSG + + PG V AL+GENGAGKST+M +LTG++ D G +
 Sbjct: **5** LQLKGIDKAPPGVKALSGAALNVYPGRVMALVGENGAGKSTMMKVLGTGIYTRDAGTLLWL 64

Query: 63 GNETYFSNPKEAEQHGIAFIHQELNIWPEMTVLENLFIGKEISSKLGVLQTRKMKALAKE 122
 G ET F+ PK +++ GI IHQELN+ P++T+ EN+F+G+E ++ G + + M A A +
 Sbjct: 65 GKETTFTGPKSSQEAGIGIIHQELNLIPQLTIAENIFLGRFVNRFKIDWKTMYAEADK 124

Query: 123 QFDKLSVLSLSDQEGECSVGGQQMIEIAKALMTNAEVIIMDEPTAALTEREISKLFEVI 182
 KL++ D+ G+ S+G QOM+EIAK L ++VIIMDEPT ALT+ E LF VI
 Sbjct: 125 LLAKLNLRFKSDKLVGDLSIGDQOMVEIAKVLFSFESKVIIMDEPTDALTDTEESLFRVI 184

Query: 183 TALKKNGVSIYIISHRMEEIFAICDRITIMRDGKTVDTTNISETDFDEVVKMVGRELTE 242
 LK G IVYISHRM+EIF ICD +T+ RDG+ + ++ D +++ MVGR+L +
 Sbjct: 185 RELKSQGRGIVYIISHRMKEIFEICDDVTVFRDQGFIAEREVASLTEDSLIEMMVGRKLED 244

Query: 243 RYPKRTPSLGDKVFEVKN **260**
 +YP + GD +V N
 Sbjct: 245 QYPHLDKAPGDIRLKVDN **262**

Score = 103 bits (220), Expect(3) = e-141
Identities = 43/94 (45%), Positives = 66/94 (69%)

Query: **269** DVSFYVRSGEIVGVSGLMGAGRTEMMRALFGVDRLDTGEIWIAGKKTAIKNPQEAVKKG 328
 DVSF +R GEI+GVSGLMGAGRTE+M+ L+G +G + + G + ++PQ+ + G+
 Sbjct: **270** DVSFTRLRKGIEILGVSGLMGAGRTEMLMKVLYGALPRTSGYVTLDGHEVVTRSPQDGLANGI 329

Query: 329 GFITENRKDEGLLLDTSIRENIALPNLSSFSFKG **362**
 +I+E+RK +GL+L S++EN++L L FS G
 Sbjct: 330 VYISEDRKRDGLVGLMSVKENMSLTALRYFSRAG **363**

Résultat d'une recherche avec la version 1 de BlastP

Score = 165 bits (356), Expect(3) = e-141
Identities = 74/130 (56%), Positives = 92/130 (69%)

Query: **362** GLIDHKREAEFVDLLIKRLTIKTASPETHARHLSGGNQKQVVIKAWIGIGPKVLILDEPT 421
 G + H E + V I+ +K T S E LSGGNQKQV IA+ + PKVLILDEPT
 Sbjct: **364** GSLKHADEQQAVSDFIRLFNVKTPSMEQAIGLLSGGNQKQVAIARGLMTRPKVLILDEPT 423

Query: 422 RGVVDVGAKREIYTLMNELTERGVVAIIMVSSSELPEILGMSDRIIVVHEGRISGEIHAREAT 481
 RGVVDVGAK+EIY L+N+ G++II+VSSE+PE+LGMSDRIIV+HEG +SGE +AT
 Sbjct: 424 RGVVDVGAKKEIYQLINQFKADGLSIIIVSSEMPEVLGMSDRIIVMHEGHLSGEFTREQAT 483

Query: 482 QERIMTLATG **491**
 QE +M A G
 Sbjct: 484 QEVLMAAAVG **493**

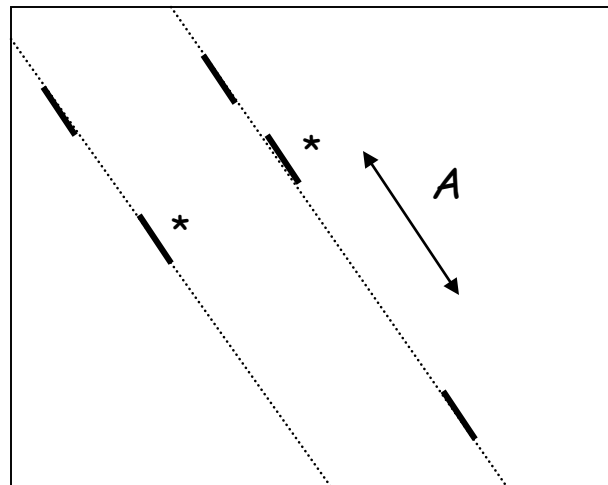
Gapped Blast ou Blast2

(Altschul *et al.* (1997) *Nucleic Acids Res.*, 25, 3389-3402)

Les deux premières étapes identiques à la version 1 de l'algorithme de Blast

Différences :

- étape d'extension des hits (90% du temps d'exécution du Blast)
Rq : une HSP contient souvent plusieurs hits sur la même diagonale et à une distance peu éloignée les uns des autres.
Sélection des hits pour l'étape d'extension : présence de deux hits sur la même diagonale séparés par une distance inférieure ou égale à A .



Les hits marqués d'une * subiront une extension sans gaps analogue à la 3ème étape du Blast version 1

Gapped Blast ou Blast2

Différences :

- Les HSP sélectionnés qui auront un score supérieur ou égal à S_{HSP} servent ensuite de points d'ancrage à une recherche d'alignement local optimal par programmation dynamique. La matrice de programmation dynamique est explorée dans les deux directions à partir d'un résidu aligné (graine).

Choix de la graine : on choisit le long d'une HSP le segment de 11 résidus ayant le meilleur score et le résidu central est utilisé comme graine.

La recherche du chemin optimal est limité aux cellules de la matrice telles que le score de l'alignement ne devienne pas inférieur de plus de Xg au score maximal atteint jusque là (modification de l'algorithme de Smith et Waterman)

Significativité des alignements : La E-value

Dans le cas des alignements locaux avec gaps, il n'existe pas de théorie décrivant la distribution attendue des scores.

La E-value (Expect value), le nombre attendu d'alignement qui par chance aurait un score $\geq S$, est évaluée en regardant les scores des alignements générés par comparaison de séquences aléatoires ayant même longueur et même composition que la séquence requête.

La suite Blast

Un ensemble de programmes :

programme	séquence requête	Banque
BlastN	nucléique	nucléique
BlastP	protéique	protéique
BlastX	nucléique (séquence traduite dans les 6 cadres de lecture)	protéique
tblastN	protéique	nucléique (séquences de la banque traduites dans les 6 cadres de lecture)
tblastX	nucléique (séquence traduite dans les 6 cadres de lecture)	nucléique (séquences de la banque traduites dans les 6 cadres de lecture)

La suite Blast

Possibilité d'appliquer des filtres et masques (paramètres de l'algorithme) :

- masquer les séquences de faible complexité (proposé pour l'ensemble des programmes de la suite Blast)
- dans le cas d'une recherche avec une séquence d'acides nucléiques contre une banque de séquences nucléiques (BlastN), masquer les séquences répétées (ex: les séquences Alu chez les primates).

Une première analyse compare la séquence sonde à une banque de séquences d'éléments répétés. Les zones de la séquence sonde s'alignant avec les séquences d'éléments répétés sont masquées pour la recherche de similarité dans la banque.

Page d'accueil de la suite Blast (site NCBI)

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

**N
E
W
S**

A new version IgBLAST (1.15.0) is here.

We've released a new version of IgBLAST with four new improvements.

Wed, 22 Jan 2020 16:00:00 EST

[More BLAST news...](#)

Web BLAST



Nucleotide BLAST
nucleotide ► nucleotide



blastx
translated nucleotide ► protein



tblastn
protein ► translated nucleotide



Protein BLAST
protein ► protein

BLAST Genomes

Human Mouse Rat Microbes

Exemple d'interface du programme BlastP (site NCBI)

The image shows a screenshot of the NCBI Protein BLAST search interface. The browser address bar shows the URL: https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&. The page title is "Protein BLAST: search protein d". The interface is titled "Standard Protein BLAST".

The main section is "Enter Query Sequence". It contains a text input field for "Enter accession number(s), gi(s), or FASTA sequence(s)", a "Query subrange" section with "From" and "To" fields, and an "Or, upload file" section with a "Parcourir..." button. Below this is a "Job Title" field and a checkbox for "Align two or more sequences".

The "Choose Search Set" section includes a "Database" dropdown menu set to "Non-redundant protein sequences (nr)", an "Organism" field with an "exclude" button, and an "Exclude" section with checkboxes for "Models (XM/XP)", "Non-redundant RefSeq proteins (WP)", and "Uncultured/environmental sample sequences".

The "Program Selection" section has a radio button for "blastp (protein-protein BLAST)" selected, along with other options like "Quik BLASTP", "PSI-BLAST", "PHI-BLAST", and "DELTA-BLAST".

At the bottom, there is a "BLAST" button and a checkbox for "Show results in a new window".

Two blue annotations with arrows point to specific parts of the interface:

- An arrow points from the text "Votre séquence sonde" to the "Enter accession number(s), gi(s), or FASTA sequence(s)" input field.
- An arrow points from the text "Choix de la banque" to the "Database" dropdown menu.

Les paramètres « cachés »

Algorithm parameters

General Parameters

Max target sequences

100

Nombre de séquences cibles

Select the maximum number of aligned sequences to display

Short queries

Automatically adjust parameters for short input sequences

Expect threshold

10

Seuil E-value

Word size

3

Taille des mots pour construire la liste des mots voisins

Max matches in a query range

0

Scoring Parameters

Matrix

BLOSUM62

Choix de la matrice de substitution

Gap Costs

Existence: 11 Extension: 1

Pondération des gaps : ouverture et extension

Compositional adjustments

Conditional compositional score matrix adjustment

Filters and Masking

Filter

Low complexity regions

Mask

Mask for lookup table only

Mask lower case letters

Exemple de recherche BlastP (site NCBI)

BLAST » blastp suite

Standard Protein BLAST

blastn blastp **blastx** tblastn tblastx

BLASTP programs search protein databases using a protein query [more...](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) Query subrange

```
>Eco1A01.YRAP
MKALSPFIAVLISALLLGGCVAAAIVGTAAVGTAKAATDFRSVGTQVDDGTLVFRVNSALSK
DEQIKKKEARINVTAYQGRVLLVQGSFNAEL SARAKIANGVDGANEVYNEIRQQPIGLG
EASNDTWITKTVRAGQLLSDLVKSNNVKVTENGEVFLMGLVTEREAKAADIASRVSGV
KRYVIAFTTIIK
```

From To

Or, upload file Aucun fichier sélectionné.

Job Title Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database

Organism Exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude Models (XM/XP) Non-redundant RefSeq proteins (WP) Uncultured/environmental sample sequences

Program Selection

Algorithm

Quick BLASTP (Accelerated protein-protein BLAST)

blastp (protein-protein BLAST)

PSI-BLAST (Position-Specific Iterated BLAST)

PHI-BLAST (Pattern Hit Initiated BLAST)

DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm

BLAST Search database nr using Blastp (protein-protein BLAST)

Show results in a new window

Algorithm parameters **Note: Parameter values that differ from the default are highlighted in yellow and marked with * sign**

General Parameters

Max target sequences Select the maximum number of aligned sequences to display

Short queries Automatically adjust parameters for short input sequences

Expect threshold

Word size

Max matches in a query range

Scoring Parameters

Matrix

Banque nr : non redondante. Toutes les protéines non redondantes correspondant à la traduction des CDS de Genbank + PDB + SwissProt + PIR-NBRF

Recherche restreinte aux alphaproteobacteria

Nombre de séquences cibles : 500 (nombre de hits)

Page de résultat BlastP (site NCBI)

BLAST® » blastp suite » results for RID-38G2CGA901R

Home Recent Results Saved Strategies Help

[< Edit Search](#) [Save Search](#) [Search Summary](#) ▼

[How to read this report?](#) [BLAST Help Videos](#) [Back to Traditional Results Page](#)

i Your search is limited to records that include: Alphabacteria (taxid:28211)

Job Title EcoIA01.YRAP

RID [38G2CGA901R](#) *Search expires on 02-01 21:57 pm* [Download All](#) ▼

Program BLASTP [?](#) [Citation](#) ▼

Database nr [See details](#) ▼

Query ID Icl|Query_4796

Description EcoIA01.YRAP

Molecule type amino acid

Query Length 191

Other reports [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#) [?](#)

Filter Results

Organism *only top 20 will appear* exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity to

E value to

Query Coverage to

[Filter](#) [Reset](#)

Descriptions [Graphic Summary](#) [Alignments](#) [Taxonomy](#)

Sequences producing significant alignments

[Download](#) ▼ [Manage Columns](#) ▼ Show [?](#)

select all 371 sequences selected

[GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#)

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	osmotically-inducible protein OsmY [Thalassospira xiamenensis]	159	159	93%	5e-47	48.80%	PTC01376.1
<input checked="" type="checkbox"/>	osmotically-inducible protein OsmY [Rickettsiales bacterium]	137	137	96%	1e-38	39.67%	MBB20570.1
<input checked="" type="checkbox"/>	BON domain-containing protein [Rickettsiales bacterium]	123	123	99%	7e-33	39.69%	MBB86425.1
<input checked="" type="checkbox"/>	MULTISPECIES: BON domain-containing protein [Proteobacteria]	107	107	99%	1e-26	36.32%	WP_124735979.1
<input checked="" type="checkbox"/>	phospholipid-binding protein [Rhodospirillaceae bacterium]	105	105	94%	9e-26	34.62%	MBD21055.1
<input checked="" type="checkbox"/>	BON domain-containing protein [Alphaproteobacteria bacterium]	103	103	99%	2e-25	33.88%	NBR40331.1
<input checked="" type="checkbox"/>	transport-associated protein [Rickettsiales bacterium]	101	101	87%	2e-24	37.13%	MBV35553.1
<input checked="" type="checkbox"/>	Uncharacterized protein in gshII_5region [Candidatus Jidaibacter acanthamoeba]	100	100	98%	1e-23	33.51%	KIE04947.1
<input checked="" type="checkbox"/>	hypothetical protein BTY25_04340 [Alphaproteobacteria bacterium 16-39-46]	100	100	90%	3e-23	33.91%	OYZ37433.1
<input checked="" type="checkbox"/>	BON domain-containing protein [Candidatus Jidaibacter acanthamoeba]	98.8	98.8	89%	4e-23	35.47%	WP_039456942.1
<input checked="" type="checkbox"/>	BON domain-containing protein [endosymbiont of Acanthamoeba sp. UWCS]	98.8	98.8	89%	4e-23	34.88%	WP_038537893.1
<input checked="" type="checkbox"/>	hypothetical protein COV36_02350 [Alphaproteobacteria bacterium CG11_big_fil_rev_8_21_14_0_20_44_7]	96.7	96.7	92%	2e-22	34.46%	PIR33644.1

Page de résultat : résumé graphique

Descriptions

Graphic Summary

Alignments

Taxonomy

[hover to see the title](#) [click to show alignments](#) Show Conserved Domains

Alignment Scores

■ < 40

■ 40 - 50

■ 50 - 80

■ 80 - 200

■ >= 200

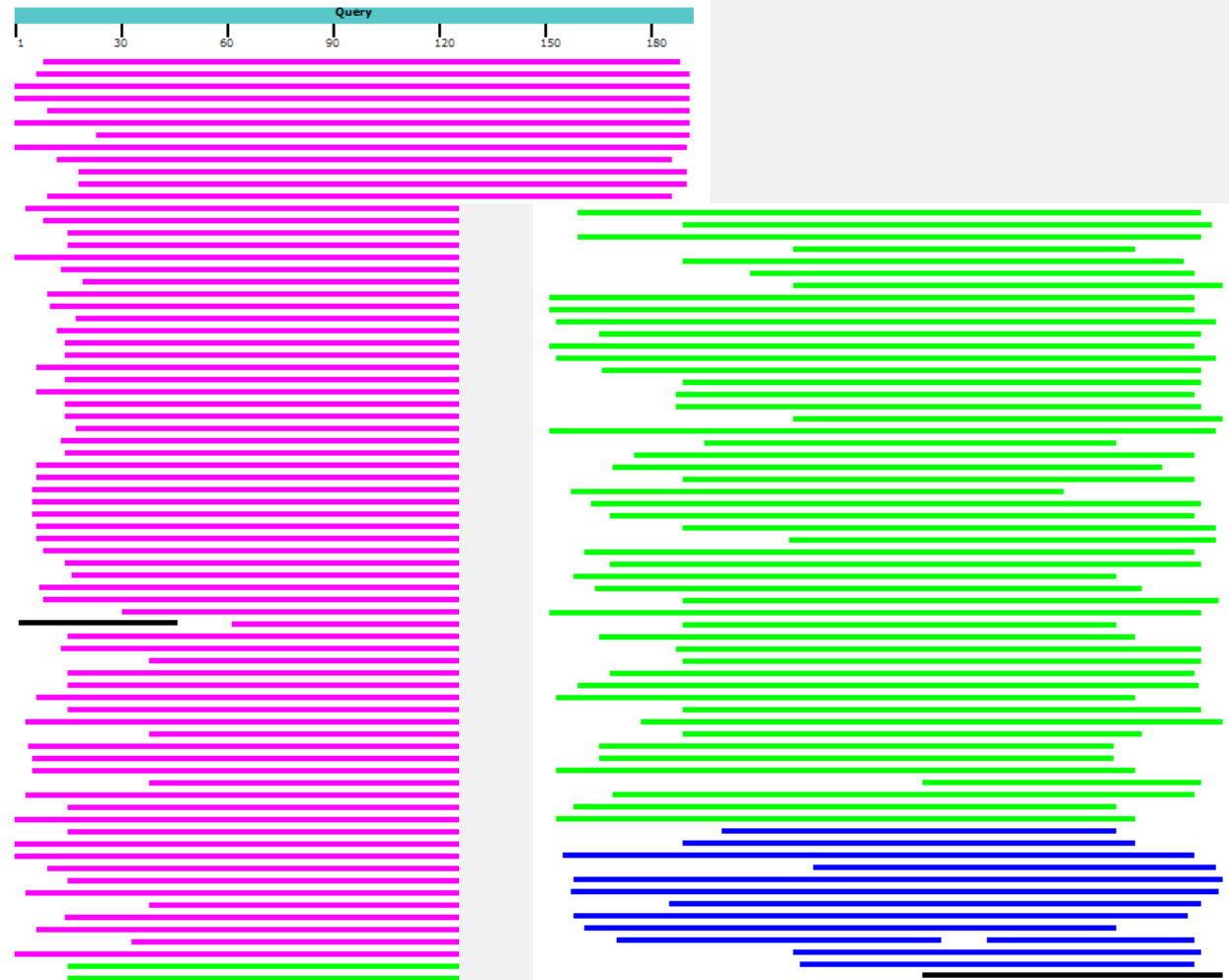


371 sequences selected [?](#)

Putative conserved domains have been detected, click on the image below for detailed results.

[CD search result summary](#)

Distribution of the top 371 Blast Hits on 371 subject sequences



Page de résultat : alignement deux à deux

Descriptions Graphic Summary **Alignments** Taxonomy

Alignment view Pairwise ?

Download ▼

371 sequences selected ?

[Download](#) ▼ [GenPept](#) [Graphics](#)

[Next](#) [Previous](#) [Descriptions](#)

osmotically-inducible protein OsmY [Thalassospira xiamenensis]

Sequence ID: [PTC01376.1](#) Length: 196 Number of Matches: 1

Range 1: 17 to 193 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
159 bits(403)	5e-47	Compositional matrix adjust.	87/179(49%)	116/179(64%)	2/179(1%)
Query 9	V LISALLLQGCVAAAVVGTAAVGTKAATDPRSVGTQVDDGTLLEVRVNSALS KDEQIKKEA	68			
	+LIS L+LQGC A A+VG AVG KAATD RS+GTQ+DD T+E R A+S+ A				
Sbjct 17	LLISVLM LQGC-AVALVGAGAVGKAATDRRSLGTQLDDQTIEYRAQRAISEGSAFDN-A	74			
Query 69	RINVTAYQ GKVLLVVGQSPNAEL SARAKQIAMGV D GANEVYNEIRQGQPIGLGEASNDTWI	128			
	R+ AY +VLLVGQ+P+ E+ +A+ G +V+NE+R +G + DTWI				
Sbjct 75	RVLAVAYNDRVLLVVGQAPSDEIRRQAE LAVREAVGERQVHNEVRIRSRVGF TTKTEDTWI	134			
Query 129	TTKVRSQLLTS DLVKSSNVKVT TENG E VFLMGLVTEREAKAAADIASRVSGVKRVTTAF	187			
	T+KV++ LL+ + V S +KV TENG E VFLMG+V E + A DIA +GV RV TAF				
Sbjct 135	TSKVK TALLS DERVDG SQIKVITENG E VFLMGI VDAEEGR IATDIARNTAGVVRVVTAF	193			

[Download](#) ▼ [GenPept](#) [Graphics](#)

[Next](#) [Previous](#) [Descriptions](#)

osmotically-inducible protein OsmY [Rickettsiales bacterium]

Sequence ID: [MBB20570.1](#) Length: 192 Number of Matches: 1

Range 1: 10 to 192 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
137 bits(346)	1e-38	Compositional matrix adjust.	73/184(40%)	113/184(61%)	1/184(0%)
Query 7	I AVLISALLLQGCVAAAVVGTAAVGTKAATDPRSVGTQVDDGTLLEVRVNSALS KDEQIKK	66			
	I L++ LLQGC A V AA D RS+GTQ+DD T+ R++ ALS ++ I +				
Sbjct 10	IGALLALSLLQGC AALVVGAGVGA-ASAADRSLGTQLDDKTVSSRISIALS DNKAIDE	68			
Query 67	E ARINVTAYQ GKVLLVVGQSPNAEL SARAKQIAMGV D GANEVYNEIRQGQPIGLGEASNDT	126			
	+A I+V + G VLLVGQ+PN L +A+ A V +++N+IR G P+ ++D				
Sbjct 69	Q ANIHVHVFN GSVLLVVGQAPNEALINQAQVAATSVKNIEKLNQIRL GSPVSA STTTHDV	128			
Query 127	W ITTKVRSQLLTS DLVKSSNVKVT TENG E VFLMGLVTEREAKAAADIASRVSGVKRVTTA	186			
	W+T+K+++ L+ + ++ V EN EVFLMGLV+ +EA+ AA +A + GVK+V A				
Sbjct 129	W L TSKIKANLVADKRIDGFH I HVA VENSEVFLMGLVSTKEA EVAATVARNIDGVKQVIKA	188			
Query 187	FTFI 190				
	F ++				
Sbjct 189	FEYL 192				


Related Information

[Identical Proteins](#) - Identical proteins to MAD17062.1

Page de résultat : distribution taxonomique des hits

Descriptions Graphic Summary Alignments **Taxonomy**

Reports Lineage Organism **Taxonomy**

371 sequences selected 

Taxonomy	Number of hits	Number of Organisms	Description
<input checked="" type="checkbox"/> Bacteria	508	246	
<input checked="" type="checkbox"/> Proteobacteria	1	246	Proteobacteria hits
<input checked="" type="checkbox"/> Alphaproteobacteria	507	245	
<input checked="" type="checkbox"/> Rhodospirillales	236	94	
<input checked="" type="checkbox"/> Rhodospirillaceae	222	88	
<input checked="" type="checkbox"/> Thalassospira	12	7	
<input checked="" type="checkbox"/> Thalassospira xiamenensis	1	1	Thalassospira xiamenensis hits
<input checked="" type="checkbox"/> Thalassospira marina	3	1	Thalassospira marina hits
<input checked="" type="checkbox"/> unclassified Thalassospira	4	2	
<input checked="" type="checkbox"/> Thalassospira sp. MCCC 1A01428	2	1	Thalassospira sp. MCCC 1A01428 hits
<input checked="" type="checkbox"/> Thalassospira sp. TSL5-1	2	1	Thalassospira sp. TSL5-1 hits
<input checked="" type="checkbox"/> Thalassospira australica	1	1	Thalassospira australica hits
<input checked="" type="checkbox"/> Thalassospira profundimaris	1	1	Thalassospira profundimaris hits
<input checked="" type="checkbox"/> Thalassospira mesophila	2	1	Thalassospira mesophila hits
<input checked="" type="checkbox"/> unclassified Rhodospirillaceae	52	5	
<input checked="" type="checkbox"/> Rhodospirillaceae bacterium	48	1	Rhodospirillaceae bacterium hits
<input checked="" type="checkbox"/> Rhodospirillaceae bacterium LM-1	1	1	Rhodospirillaceae bacterium LM-1 hits
<input checked="" type="checkbox"/> Rhodospirillaceae bacterium TMED23	1	1	Rhodospirillaceae bacterium TMED23 hits
<input checked="" type="checkbox"/> Rhodospirillaceae bacterium TMED256	1	1	Rhodospirillaceae bacterium TMED256 hits
<input checked="" type="checkbox"/> Rhodospirillaceae bacterium TMED63	1	1	Rhodospirillaceae bacterium TMED63 hits
<input checked="" type="checkbox"/> Azospirillum	1	28	Azospirillum hits
<input checked="" type="checkbox"/> unclassified Azospirillum	1	16	unclassified Azospirillum hits
<input checked="" type="checkbox"/> Azospirillum sp. TSH58	3	1	Azospirillum sp. TSH58 hits
<input checked="" type="checkbox"/> Azospirillum sp. TSO5	2	1	Azospirillum sp. TSO5 hits
<input checked="" type="checkbox"/> Azospirillum sp. TSH7	1	1	Azospirillum sp. TSH7 hits
<input checked="" type="checkbox"/> Azospirillum sp. TSH20	1	1	Azospirillum sp. TSH20 hits
<input checked="" type="checkbox"/> Azospirillum sp. TSA2s	1	1	Azospirillum sp. TSA2s hits
<input checked="" type="checkbox"/> Azospirillum sp. Sh1	2	1	Azospirillum sp. Sh1 hits
<input checked="" type="checkbox"/> Azospirillum sp. TSA6c	2	1	Azospirillum sp. TSA6c hits
<input checked="" type="checkbox"/> Azospirillum sp. M2T2B2	2	1	Azospirillum sp. M2T2B2 hits
<input checked="" type="checkbox"/> Azospirillum sp. TSH100	3	1	Azospirillum sp. TSH100 hits
<input checked="" type="checkbox"/> Azospirillum sp. TSH64	2	1	Azospirillum sp. TSH64 hits
<input checked="" type="checkbox"/> Azospirillum sp. B506	1	1	Azospirillum sp. B506 hits
<input checked="" type="checkbox"/> Azospirillum sp. B21	2	1	Azospirillum sp. B21 hits
<input checked="" type="checkbox"/> Azospirillum sp. TSO35-2	2	1	Azospirillum sp. TSO35-2 hits
<input checked="" type="checkbox"/> Azospirillum sp. K2W22B-5	2	1	Azospirillum sp. K2W22B-5 hits

Page de résultat : application d'un filtre

i Your search is limited to records that include: Alphabacteria (taxid:28211)
Your results are filtered to match records that include: Rickettsiales (taxid:766)

Job Title	EcolA01.YRAP
RID	38G2CGA901R Search expires on 02-01 21:57 pm Download All ▼
Program	BLASTP ? Citation ▼
Database	nr See details ▼
Query ID	lcl Query_4796
Description	EcolA01.YRAP
Molecule type	amino acid
Query Length	191
Other reports	Distance tree of results Multiple alignment MSA viewer ?

Filter Results

Organism *only top 20 will appear* exclude

Rickettsiales (taxid:766)

[+ Add organism](#)

Percent Identity

 to

E value

 to

Query Coverage

 to

Filter

Reset

[Descriptions](#) [Graphic Summary](#) [Alignments](#) [Taxonomy](#)

Sequences producing significant alignments

[Download](#) ▼ [Manage Columns](#) ▼ Show [?](#)

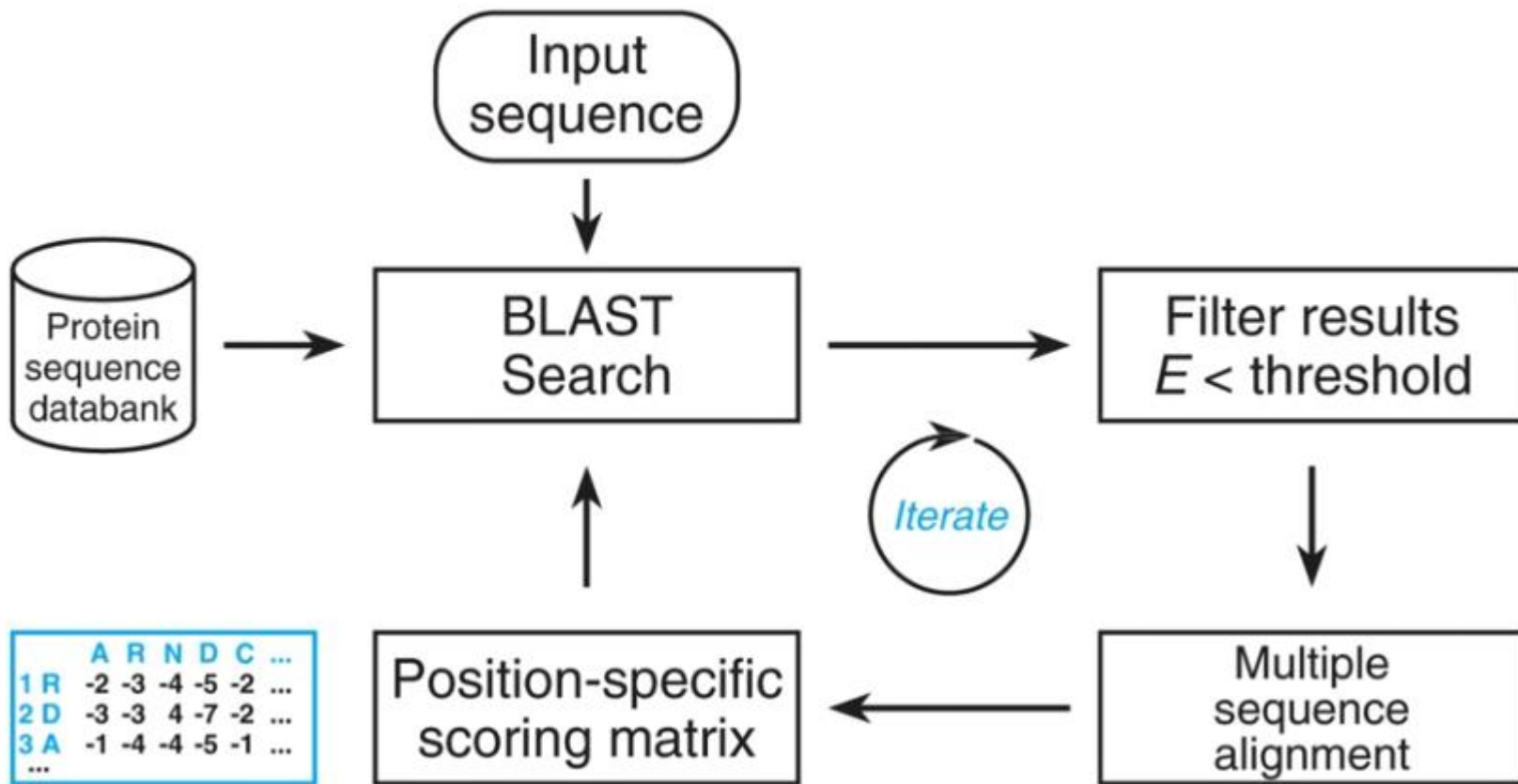
select all 20 sequences selected

[GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#)

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	osmotically-inducible protein OsmY [Rickettsiales bacterium]	137	137	96%	1e-38	39.67%	MBB20570.1
<input checked="" type="checkbox"/>	BON domain-containing protein [Rickettsiales bacterium]	123	123	99%	7e-33	39.69%	MBB66425.1
<input checked="" type="checkbox"/>	transport-associated protein [Rickettsiales bacterium]	101	101	87%	2e-24	37.13%	MBV35553.1
<input checked="" type="checkbox"/>	Uncharacterized protein in gshII 5' region [Candidatus Jidaibacter acanthamoeba]	100	100	98%	1e-23	33.51%	KIE04947.1
<input checked="" type="checkbox"/>	BON domain-containing protein [Candidatus Jidaibacter acanthamoeba]	98.6	98.6	89%	4e-23	35.47%	WP_039456942.1
<input checked="" type="checkbox"/>	BON domain-containing protein [Candidatus Aquarickettsia rohweri]	82.4	82.4	91%	5e-17	32.22%	WP_128044620.1
<input checked="" type="checkbox"/>	Osmotically-inducible protein Y precursor [Candidatus Arcanobacter lacustris]	82.0	82.0	94%	8e-17	30.27%	KKB96100.1
<input checked="" type="checkbox"/>	BON domain-containing protein [Candidatus Phycoarickettsia trachydisci]	80.5	80.5	96%	2e-16	30.16%	WP_108874613.1
<input checked="" type="checkbox"/>	BON domain-containing protein [Candidatus Hepatobacter penaei]	74.3	74.3	87%	7e-14	32.94%	WP_031934417.1
<input checked="" type="checkbox"/>	hypothetical protein [Rickettsiales bacterium]	74.3	74.3	96%	9e-14	31.75%	MBO90079.1
<input checked="" type="checkbox"/>	BON domain protein [Neorickettsia helminthoeca str. Oregon]	70.1	70.1	96%	2e-12	28.57%	AHX11031.1
<input checked="" type="checkbox"/>	BON domain-containing protein [Neorickettsia helminthoeca]	70.1	70.1	96%	3e-12	28.57%	WP_051579462.1
<input checked="" type="checkbox"/>	phospholipid-binding protein [Rickettsiales bacterium TMED174]	68.6	68.6	94%	8e-12	25.41%	OUW24413.1
<input checked="" type="checkbox"/>	BON domain-containing protein [Rickettsia typhi]	68.9	68.9	92%	8e-12	29.12%	WP_011190840.1
<input checked="" type="checkbox"/>	BON domain-containing protein [Candidatus Midichloria mitochondrii]	67.0	67.0	89%	4e-11	28.07%	WP_013951652.1
<input checked="" type="checkbox"/>	hypothetical protein [Rickettsiales bacterium]	61.6	61.6	92%	4e-09	20.79%	MBH43598.1
<input checked="" type="checkbox"/>	hypothetical protein [Rickettsiales bacterium]	60.8	60.8	93%	6e-09	23.76%	MAY90257.1
<input checked="" type="checkbox"/>	hypothetical protein [Rickettsiales bacterium]	56.2	56.2	73%	3e-07	30.28%	MAZ46716.1
<input checked="" type="checkbox"/>	hypothetical protein [Rickettsiales bacterium]	55.1	55.1	96%	8e-07	25.27%	MBS91946.1
<input checked="" type="checkbox"/>	hypothetical protein [Rickettsiales bacterium]	53.1	53.1	86%	4e-06	28.74%	MBC11294.1

PSI-BLAST (Position-Specific Iterative Basic Local Alignment Search Tool)

Psiblast



Extrait de la page résultat du premier run

<input checked="" type="checkbox"/>	BON domain-containing protein [Magnetospira sp. QH-2]	45.8	45.8	80%	0.001	29.84%	WP_048020069.1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [Rhodospirillaceae bacterium RKSG073]	45.8	45.8	94%	0.002	23.63%	WP_155154364.1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [Caenispirillum salinarum]	45.8	45.8	51%	0.002	28.00%	WP_009540523.1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [SAR116 cluster bacterium]	45.8	45.8	78%	0.002	26.00%	RCL78867.1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [Magnetospirillum sp. ME-1]	45.8	45.8	76%	0.002	28.05%	WP_085372472.1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [Alphaproteobacteria bacterium]	45.4	45.4	59%	0.002	30.43%	NCF45983.1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [Alphaproteobacteria bacterium]	45.4	45.4	82%	0.002	25.95%	NBR38031.1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [Pelagibacteraceae bacterium]	45.4	45.4	83%	0.002	21.88%	MBH71125.1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [Alphaproteobacteria bacterium]	45.4	45.4	49%	0.002	29.79%	MQA65521.1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	phospholipid-binding domain-containing protein [Rhodospirillaceae bacterium]	45.1	45.1	83%	0.003	26.06%	MAI12765.1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [Roseospirillum parvum]	45.1	45.1	79%	0.003	27.33%	WP_092617644.1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	hypothetical protein CFH02_01694 [Alphaproteobacteria bacterium MarineAlpha3_Bin1]	44.7	44.7	80%	0.004	27.04%	PPR66808.1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [Magnetospirillum moscoviense]	43.1	78.2	87%	0.004	31.07%	WP_068499403.1	<input checked="" type="checkbox"/>

Run PSI-BLAST Iteration 2 with max number of sequences

Sequences with E-value WORSE than threshold

select all 72 sequences selected

PSI-BLAST iteration 1

	Description	Max score	Total score	Query cover	E value	Per. Ident	Accession	Select for PSI blast	Used to build PSSM	Newly added
<input checked="" type="checkbox"/>	BON domain-containing protein [Ensifer sp. WSM1721]	44.3	44.3	39%	0.005	40.00%	WP_051440915.1	<input type="checkbox"/>		
<input checked="" type="checkbox"/>	Osmotically-inducible protein OsmY, contains BON domain [Roseospirillum parvum]	44.7	44.7	79%	0.005	27.33%	SDH05704.1	<input type="checkbox"/>		
<input checked="" type="checkbox"/>	BON domain-containing protein [Enhydrobacter aerosaccus]	42.7	42.7	38%	0.006	36.00%	SKA25257.1	<input type="checkbox"/>		
<input checked="" type="checkbox"/>	hypothetical protein [Rhodospirillaceae bacterium]	43.9	43.9	50%	0.008	27.45%	MAF49290.1	<input type="checkbox"/>		
<input checked="" type="checkbox"/>	BON domain-containing protein [Alphaproteobacteria bacterium]	42.7	42.7	49%	0.008	29.47%	MPY71064.1	<input type="checkbox"/>		

Extrait de la page résultat du premier run

<input checked="" type="checkbox"/>	BON domain protein [Rickettsia amblyommatis str. Darkwater]	151	151	78%	3e-44	26.67%	KJV94938.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [Rhodospira trueperi]	153	153	84%	3e-44	25.15%	WP_092780429.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [Alphaproteobacteria bacterium]	152	203	95%	3e-44	24.46%	NBQ83556.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [Candidatus Micropelagos thuwalensis]	153	198	98%	4e-44	25.26%	WVP_021777646.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	putative phospholipid-binding domain protein [Wolbachia pipientis wUni]	150	150	73%	4e-44	32.87%	ONI56871.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	21 kDa hemolysin protein, putative [Rhodospirillum centenum SW]	152	152	86%	4e-44	28.12%	ACJ00819.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [alpha proteobacterium IMCC14465]	152	202	96%	6e-44	27.27%	WVP_008518969.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	Transport-associated domain [Ehrlichia minasensis]	151	151	90%	8e-44	28.00%	CEI85043.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain protein [Rickettsia amblyommatis str. Ao/Pa]	150	150	76%	8e-44	26.67%	KJV62027.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [Marinicauda pacifica]	151	197	96%	8e-44	29.57%	WVP_135945713.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	Transport-associated domain [Ehrlichia canis str. Jake]	151	151	90%	9e-44	26.86%	AAZ68703.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [PS1 clade bacterium]	153	199	98%	1e-43	23.68%	RCL83276.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	hypothetical protein [Rhizobiales bacterium]	153	199	98%	1e-43	23.68%	MAV18645.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	hypothetical protein [Rickettsiales bacterium]	151	151	100%	1e-43	23.96%	MAI83460.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [Amphiplicatus metriotherophilus]	150	150	79%	1e-43	24.34%	WVP_089410959.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	Putative phospholipid-binding domain protein (modular protein) [uncultured Alphaproteobacteria bacterium]	152	192	97%	1e-43	22.87%	SBW08860.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RlpA-like protein [Rickettsia endosymbiont of Proechinophthirus fluctus]	149	184	95%	1e-43	23.17%	KYP96621.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	TPA: hypothetical protein [Rhodobiaceae bacterium]	152	199	98%	2e-43	23.68%	HCN31600.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [Oleomonas sp. K1W22B-S]	151	210	80%	2e-43	32.90%	RJF89784.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	hypothetical protein [Rhodospirillaceae bacterium]	149	149	90%	2e-43	24.57%	MAJ92173.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	hypothetical protein COC03_03675 [Robiginitomaculum sp.]	153	198	98%	2e-43	25.79%	PHQ59960.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	hypothetical protein COB92_02500 [Robiginitomaculum sp.]	153	198	98%	2e-43	26.32%	PHQ67920.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [Henriciella marina]	151	151	97%	2e-43	32.96%	WVP_156807951.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	phospholipid-binding protein [Pelagibacteraceae bacterium]	150	150	91%	2e-43	25.84%	MAV49112.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	hypothetical protein [Rickettsiales bacterium]	150	188	99%	3e-43	22.51%	MBS91946.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [alpha proteobacterium HIMB114]	150	150	91%	3e-43	26.97%	WVP_009373058.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	phospholipid-binding protein [Candidatus Pelagibacter sp.]	150	150	89%	3e-43	28.65%	MBL60633.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [Oleisolibacter albus]	150	150	84%	3e-43	27.78%	WVP_114394999.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	hypothetical protein COA91_06815 [Robiginitomaculum sp.]	153	200	98%	3e-43	24.06%	PHS39033.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	hypothetical protein [Rhodospirillaceae bacterium]	150	150	92%	4e-43	18.08%	MBB42584.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	hypothetical protein [Rickettsiales bacterium]	149	149	97%	4e-43	24.06%	MAT32797.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [Oleomonas sp. K1W22B-S]	149	207	80%	7e-43	32.90%	WVP_158596205.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [Novispirillum itersonii]	149	149	89%	1e-42	17.92%	WVP_019645911.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [Marinicauda algicola]	148	148	76%	2e-42	28.19%	WVP_135995419.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	hypothetical protein COB56_07460 [Robiginitomaculum sp.]	150	194	96%	3e-42	26.32%	PHS75645.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [Rhizobiales bacterium]	147	147	100%	3e-42	37.19%	RYE75761.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	MULTISPECIES: BON domain-containing protein [unclassified Azospirillum]	147	147	57%	3e-42	33.33%	WVP_160105555.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	BON domain-containing protein [Haematospirillum jordaniae]	149	149	87%	4e-42	23.86%	WVP_082611849.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	TPA: phospholipid-binding protein [Hyphomonas atlantica]	146	146	68%	4e-42	35.38%	HBH44370.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

En jaune les nouvelles séquences détectées avec la PSSM