

# Recherche par similarité dans les banques/bases de données

## La suite Blast (Basic Local Alignment Search Tool)



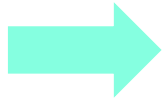
Vous voulez savoir si votre séquence d'intérêt (par exemple que vous venez de séquencer) est déjà connue où que des séquences similaires sont présentes dans les bases de données.

Cette recherche permet :

- d'inférer la fonction de la séquence à partir de séquences similaires qui auraient été étudiées expérimentalement
- de construire un échantillon de séquences homologues pour inférer des relations évolutives entre les séquences.

Pour réaliser la recherche de séquences similaires à une séquence d'intérêt (appelée séquence requête en français et query en anglais), la méthode doit passer au crible des millions de séquences et ne retenir que les séquences similaires à votre séquence requête et cela rapidement.

Donc ces approches sont plutôt basées sur la notion de filtrage que d'alignement proprement dit. En effet, les approches classiques d'alignement local de deux séquences prennent beaucoup trop de temps pour être utilisable.



Développement de méthodes utilisant des heuristiques\*

Deux logiciels :

- FASTA
- BLAST

\*Une heuristique est une méthode de calcul qui fournit rapidement une solution réalisable, pas nécessairement optimale

Définition de Blast sur le site du NCBI : The Basic Local Alignment Search Tool (BLAST) trouve des **régions de similarités locales** entre des séquences. Le programme compare des séquences nucléotidiques ou protéiques à des bases de données de séquences et calcule la **significativité statistique** des alignements.

Blast peut être utilisé pour inférer des relations fonctionnelles et évolutives entre les séquences et peut aussi aider à identifier des membres d'une famille de gènes.

Blast recherchant des alignements locaux pourra :

- Identifier les différents domaines fonctionnels de la protéine requête (cf. exemple ci-dessous) dans des séquences différentes



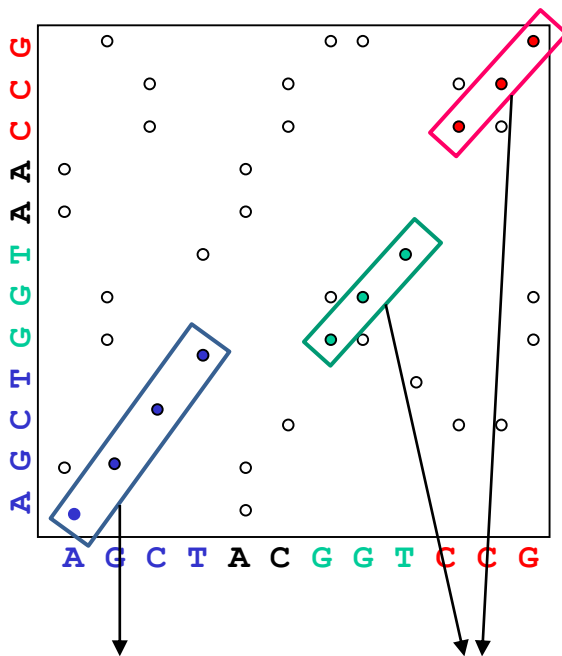
- Identifier des exons conservés au sein de différents gènes
- Par contre, si la similarité s'étend sur toute la longueur des deux séquences, Blast peut ne pas renvoyer l'équivalent de l'alignement global mais simplement l'alignement des parties les plus conservées des deux séquences.

- ✓ 1990 - Première version produite par le NCBI. Cette première version ne prend pas en compte les événements d'insertion/délétion donc ne réalise que des alignements sans gaps. Elle fournit une p-value permettant de juger de la significativité des résultats.
- ✓ 1997 - Une nouvelle version autorise les **insertions/délétions** (Blast2) et inclue le **PSI Blast** (fin du cours)
- ✓ 2009 - Nouvelle version de Blast (**BLAST+**) est produite par le NCBI (Camacho et al. 2009))
- ✓ Actuellement : version 2.11.0 du Blast+

# 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 les raccourcissant.

MSP : score 20

HSP : score 15

# Algorithme de Blast (version 1)

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

- problème avec les HSP:
  - ✓ 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  $S_{HSP}$
  
- 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

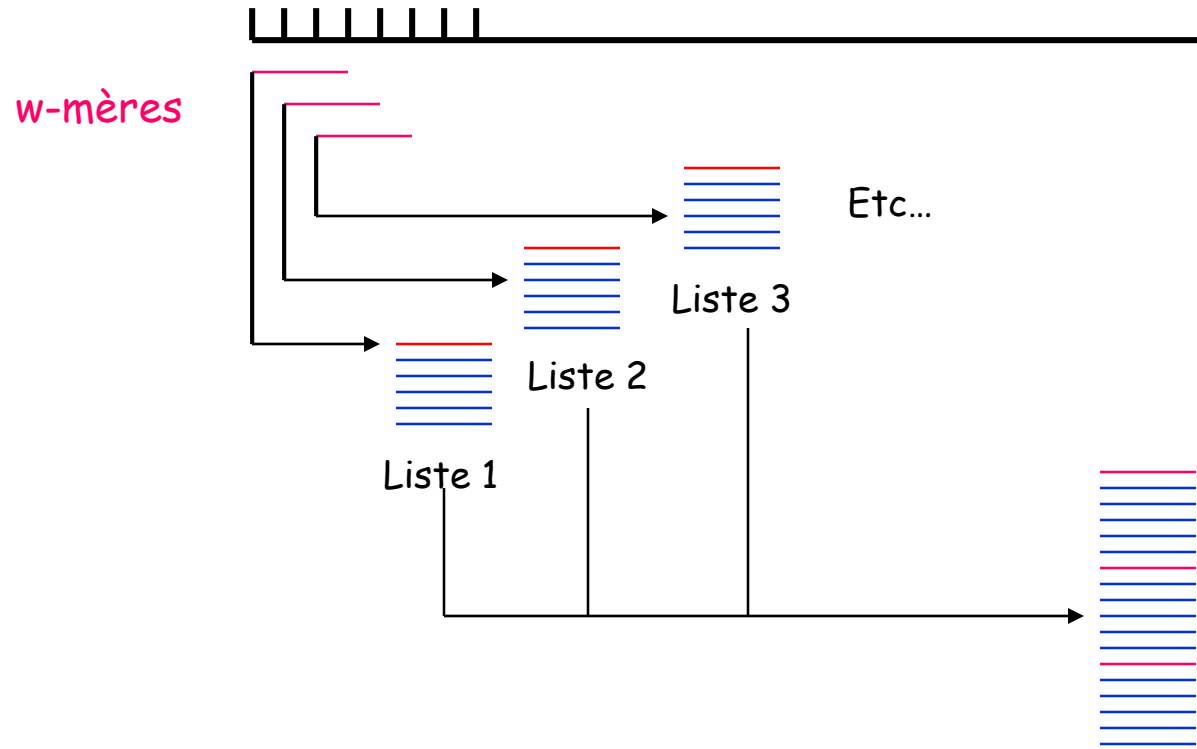
-> ce score est utilisé comme seuil  $S_{HSP}$
  
- Sélection des HSP
  - ✓ Paire de segments dont le score est  $<$  à  $S_{HSP}$ , similarité non significative due au hasard. Cette paire n'est pas retenue.
  - ✓ Paire de segments dont le score est  $\geq$  à  $S_{HSP}$ , similarité due à une histoire évolutive commune -> séquences apparentées -> 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



Exemple : RGH

mots similaires :

RDH

LGH

KGH

Liste des « mots voisins »  
mots de longueurs w  
ayant un score  
d'alignement avec le  
mot de la séquence  
requête d'au moins T

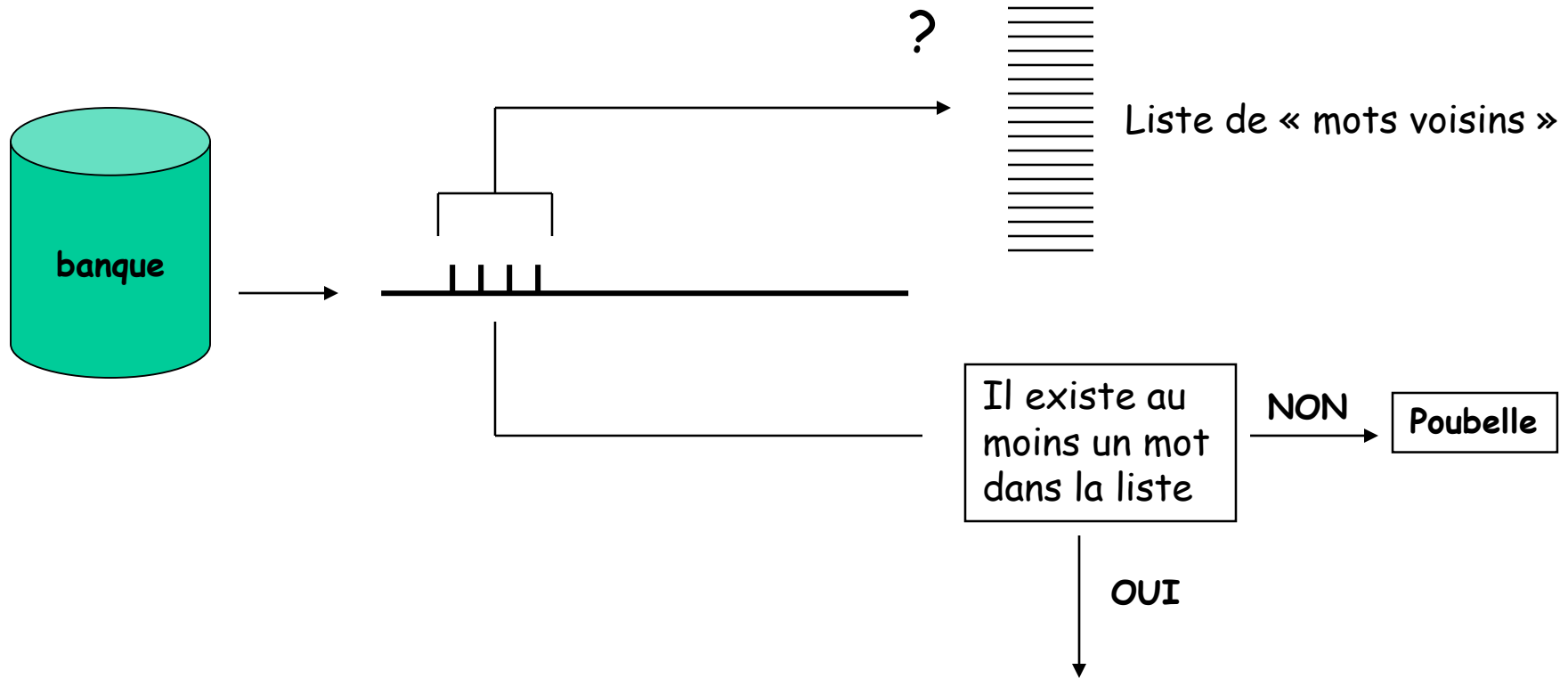
w-mères : mot de taille w (w = 3 pour les protéines et w = 11 pour les séquences d'acides nucléiques)





# 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

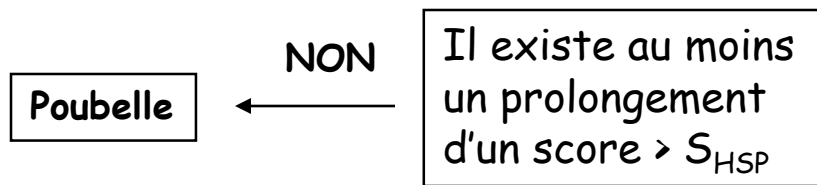
Séquence requête Q



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

# Algorithme de Blast (version 1)

## Significativité des alignements

### 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

$\lambda$  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.

# Algorithme de Blast (version 1) :

## Significativité des alignements

### La p-value

On compare une séquence de longueur  $m$  avec une banque de données de longueur  $n$ , 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 :

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

Avec  $\lambda$  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

# Résultat d'une recherche avec la version 1 de BlastP

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.YJCW "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** IEMKDIHKTFGKNQVLSGVSFQLMPGEVHALMGENGAGKSTLMNILTGLHKADKQISIN 62  
+++K I K F + LSG + + PG V AL+GENGAGKST+M +LTG++ D G +  
Sbjct: **5** LQLKGIDKAPFGVKALSGAALNVYPGRVMALVGENGAGKSTMMKVLTDGIYTRDAGTLLWL 64

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

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

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

Query: 243 RYPKRTPSLGDQVFEVKN **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** DVSFYVRSGEIVGVSGLMGAGRTEMMRALFGVDRLDTGEIWIAGKKTAIKNPQEA VVKKGL 328  
DVSF +R GEI+GVSGLMGAGRTE+M+ L+G +G + + G + ++PQ+ + G+  
Sbjct: **270** DVSFTRKKEILGVSGLMGAGRTEMLMKVLYGALPRTSGYVTLDGHEVVTRSPQDGLANGI 329

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

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+ +KT S E LSGGNQKQV IA+ + PKVLILDEPT  
Sbjct: **364** GSLKHADEQQAVSDFIRLFNVKTPSMEQAIGLLSGGNQKQVAIARGLMTRPKVLILDEPT 423

Query: 422 RGVVDVGAKREIYTLMLNELTERGVVAIIMVSELPEILGMSDRIIVVHEGRISGEIHAREAT 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**

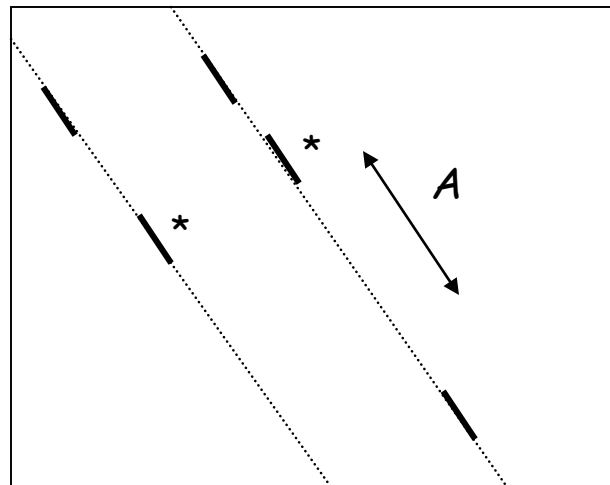
# Gapped Blast ou Blast2

(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  $A$ .



## 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)

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

**Search**

Human

Mouse

Rat

Microbes

# Exemple d'interface du programme BlastP (site NCBI)

The screenshot shows the NCBI Protein BLAST search interface. The browser address bar displays the URL: `https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&`. The page header includes the NIH logo and the text "U.S. National Library of Medicine" and "NCBI National Center for Biotechnology Information". The main heading is "BLAST® >> blastp suite" and "Standard Protein BLAST".

The interface is divided into several sections:

- Enter Query Sequence:** This section contains a large text input field for the query sequence. A blue arrow points from this field to the text "Votre séquence sonde". To the right of the input field are "From" and "To" fields under the "Query subrange" label.
- Choose Search Set:** This section includes a "Database" dropdown menu currently set to "Non-redundant protein sequences (nr)". A blue arrow points from this dropdown to the text "Choix de la banque". Below the database selection are fields for "Organism" and "Exclude" options.
- Program Selection:** This section lists various BLAST algorithms. The "blastp (protein-protein BLAST)" option is selected with a radio button.

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

# 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

**BLAST** » blastp suite

Standard Protein BLAST

blastn blastp blastx tblastn tblastx

Enter Query Sequence

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

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

Optional 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

Optional

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

BLAST<sup>®</sup> » 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"/>	<a href="#">osmotically-inducible protein OsmY [Thalassospira xiamenensis]</a>	159	159	93%	5e-47	48.80%	<a href="#">PTC01376.1</a>
<input checked="" type="checkbox"/>	<a href="#">osmotically-inducible protein OsmY [Rickettsiales bacterium]</a>	137	137	96%	1e-38	39.67%	<a href="#">MBB20570.1</a>
<input checked="" type="checkbox"/>	<a href="#">BON domain-containing protein [Rickettsiales bacterium]</a>	123	123	99%	7e-33	39.69%	<a href="#">MBB86425.1</a>
<input checked="" type="checkbox"/>	<a href="#">MULTISPECIES: BON domain-containing protein [Proteobacteria]</a>	107	107	99%	1e-26	36.32%	<a href="#">WP_124735979.1</a>
<input checked="" type="checkbox"/>	<a href="#">phospholipid-binding protein [Rhodospirillaceae bacterium]</a>	105	105	94%	9e-26	34.62%	<a href="#">MBD21055.1</a>
<input checked="" type="checkbox"/>	<a href="#">BON domain-containing protein [Alphaproteobacteria bacterium]</a>	103	103	99%	2e-25	33.88%	<a href="#">NBR40331.1</a>
<input checked="" type="checkbox"/>	<a href="#">transport-associated protein [Rickettsiales bacterium]</a>	101	101	87%	2e-24	37.13%	<a href="#">MBV35553.1</a>
<input checked="" type="checkbox"/>	<a href="#">Uncharacterized protein in gshII_5' region [Candidatus Jidaibacter acanthamoeba]</a>	100	100	98%	1e-23	33.51%	<a href="#">KIE04947.1</a>
<input checked="" type="checkbox"/>	<a href="#">hypothetical protein BTY25_04340 [Alphaproteobacteria bacterium 16-39-46]</a>	100	100	90%	3e-23	33.91%	<a href="#">OYZ37433.1</a>
<input checked="" type="checkbox"/>	<a href="#">BON domain-containing protein [Candidatus Jidaibacter acanthamoeba]</a>	98.8	98.8	89%	4e-23	35.47%	<a href="#">WP_039456942.1</a>
<input checked="" type="checkbox"/>	<a href="#">BON domain-containing protein [endosymbiont of Acanthamoeba sp. UWCS]</a>	98.8	98.8	89%	4e-23	34.88%	<a href="#">WP_038537893.1</a>
<input checked="" type="checkbox"/>	<a href="#">hypothetical protein COV36_02350 [Alphaproteobacteria bacterium CG11_big_fil_rev_8_21_14_0_20_44_7]</a>	96.7	96.7	92%	2e-22	34.46%	<a href="#">PIR33644.1</a>



# Page de résultat : résumé graphique

Descriptions

Graphic Summary

Alignments

Taxonomy

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

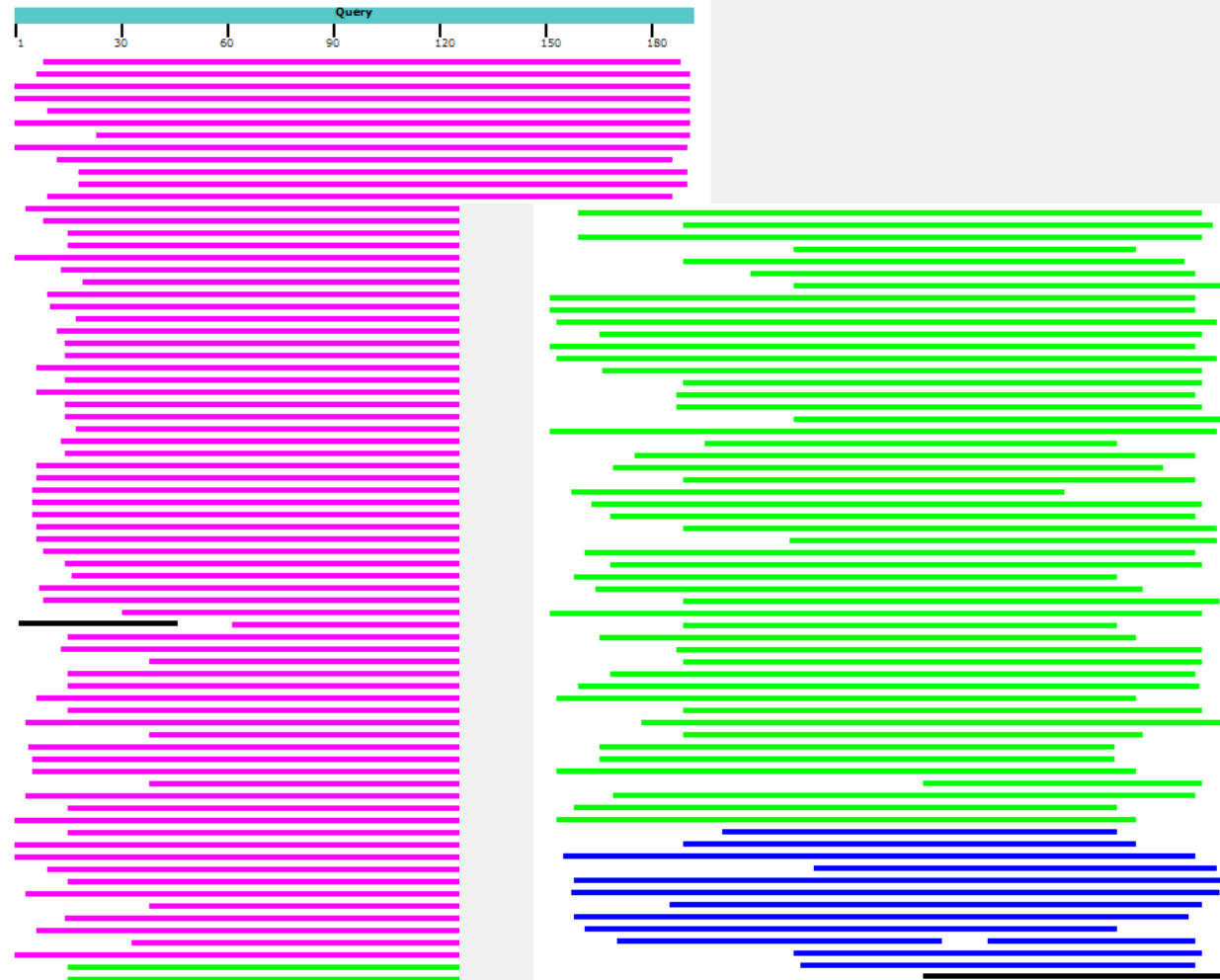
Alignment Scores  < 40  40 - 50  50 - 80  80 - 200  >= 200

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  ?

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 LISALLLQGCVAAAVVGTAAVGTKAATDPRSVGTQVDDGTLVVRVNSALS 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	RINVTAYQGKVLVVGQSPNAEL 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 VFLMGRIVDAEEGR 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 AVLISALLLQGCVAAAVVGTAAVGTKAATDPRSVGTQVDDGTLVVRVNSALS KDEQIKK	66	I L++ LLQGC A V AA D RS+GTQ+DD T+ R++ ALS ++ I +		
Sbjct 10	IGALLALSLLQGC AALVVGAGVGA-ASAADRRLSLGTQLDDKTVSSRISIALSDNKAI DE	68			
Query 67	EARINVTAYQGKVLVVGQSPNAEL SARAKQIAMGV D GANEVYNEIRQGQPIGLGEASNDT	126	+A I+V +G VLLVGQ+PN L +A+ A V +++N+IR G P+ ++D		
Sbjct 69	QANIHVHVFN GSVLLVGQAPNEALINQAQVAATSVKNIEKLNQIRL GSPVSASTTT HDV	128			
Query 127	WITTKVRSQLLTS DLVKSSNVKVT TENG E VFLMGLVTEREAKAAADIASRVSGVKRVTTA	186	W+T+K+++ L+ + ++ V EN EVFLMGLV+ +EA+ AA +A + GVK+V A		
Sbjct 129	WLT SKIKANLVADKRIDGFHIVAVENSEVFLMGLVSTKEA 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 type="checkbox"/> Bacteria	508	246	
<input type="checkbox"/> Proteobacteria	1	246	<a href="#">Proteobacteria hits</a>
<input type="checkbox"/> Alphaproteobacteria	507	245	
<input type="checkbox"/> Rhodospirillales	236	94	
<input type="checkbox"/> Rhodospirillaceae	222	88	
<input type="checkbox"/> Thalassospira	12	7	
<input type="checkbox"/> Thalassospira xiamenensis	1	1	<a href="#">Thalassospira xiamenensis hits</a>
<input type="checkbox"/> Thalassospira marina	3	1	<a href="#">Thalassospira marina hits</a>
<input type="checkbox"/> unclassified Thalassospira	4	2	
<input type="checkbox"/> Thalassospira sp. MCCC 1A01428	2	1	<a href="#">Thalassospira sp. MCCC 1A01428 hits</a>
<input type="checkbox"/> Thalassospira sp. TSL5-1	2	1	<a href="#">Thalassospira sp. TSL5-1 hits</a>
<input type="checkbox"/> Thalassospira australica	1	1	<a href="#">Thalassospira australica hits</a>
<input type="checkbox"/> Thalassospira profundimaris	1	1	<a href="#">Thalassospira profundimaris hits</a>
<input type="checkbox"/> Thalassospira mesophila	2	1	<a href="#">Thalassospira mesophila hits</a>
<input type="checkbox"/> unclassified Rhodospirillaceae	52	5	
<input type="checkbox"/> Rhodospirillaceae bacterium	48	1	<a href="#">Rhodospirillaceae bacterium hits</a>
<input type="checkbox"/> Rhodospirillaceae bacterium LM-1	1	1	<a href="#">Rhodospirillaceae bacterium LM-1 hits</a>
<input type="checkbox"/> Rhodospirillaceae bacterium TMED23	1	1	<a href="#">Rhodospirillaceae bacterium TMED23 hits</a>
<input type="checkbox"/> Rhodospirillaceae bacterium TMED256	1	1	<a href="#">Rhodospirillaceae bacterium TMED256 hits</a>
<input type="checkbox"/> Rhodospirillaceae bacterium TMED63	1	1	<a href="#">Rhodospirillaceae bacterium TMED63 hits</a>
<input type="checkbox"/> Azospirillum	1	28	<a href="#">Azospirillum hits</a>
<input type="checkbox"/> unclassified Azospirillum	1	16	<a href="#">unclassified Azospirillum hits</a>
<input type="checkbox"/> Azospirillum sp. TSH58	3	1	<a href="#">Azospirillum sp. TSH58 hits</a>
<input type="checkbox"/> Azospirillum sp. TSO5	2	1	<a href="#">Azospirillum sp. TSO5 hits</a>
<input type="checkbox"/> Azospirillum sp. TSH7	1	1	<a href="#">Azospirillum sp. TSH7 hits</a>
<input type="checkbox"/> Azospirillum sp. TSH20	1	1	<a href="#">Azospirillum sp. TSH20 hits</a>
<input type="checkbox"/> Azospirillum sp. TSA2s	1	1	<a href="#">Azospirillum sp. TSA2s hits</a>
<input type="checkbox"/> Azospirillum sp. Sh1	2	1	<a href="#">Azospirillum sp. Sh1 hits</a>
<input type="checkbox"/> Azospirillum sp. TSA6c	2	1	<a href="#">Azospirillum sp. TSA6c hits</a>
<input type="checkbox"/> Azospirillum sp. M2T2B2	2	1	<a href="#">Azospirillum sp. M2T2B2 hits</a>
<input type="checkbox"/> Azospirillum sp. TSH100	3	1	<a href="#">Azospirillum sp. TSH100 hits</a>
<input type="checkbox"/> Azospirillum sp. TSH64	2	1	<a href="#">Azospirillum sp. TSH64 hits</a>
<input type="checkbox"/> Azospirillum sp. B508	1	1	<a href="#">Azospirillum sp. B506 hits</a>
<input type="checkbox"/> Azospirillum sp. B21	2	1	<a href="#">Azospirillum sp. B21 hits</a>
<input type="checkbox"/> Azospirillum sp. TSO35-2	2	1	<a href="#">Azospirillum sp. TSO35-2 hits</a>
<input type="checkbox"/> Azospirillum sp. K2W22B-5	2	1	<a href="#">Azospirillum sp. K2W22B-5 hits</a>

# 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	<a href="#">38G2CGA901R</a> Search expires on 02-01 21:57 pm <a href="#">Download All</a> <span>▼</span>
Program	BLASTP <a href="#">?</a> <a href="#">Citation</a> <span>▼</span>
Database	nr <a href="#">See details</a> <span>▼</span>
Query ID	lcl Query_4796
Description	EcolA01.YRAP
Molecule type	amino acid
Query Length	191
Other reports	<a href="#">Distance tree of results</a> <a href="#">Multiple alignment</a> <a href="#">MSA viewer</a> <a href="#">?</a>

## 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  [?](#)

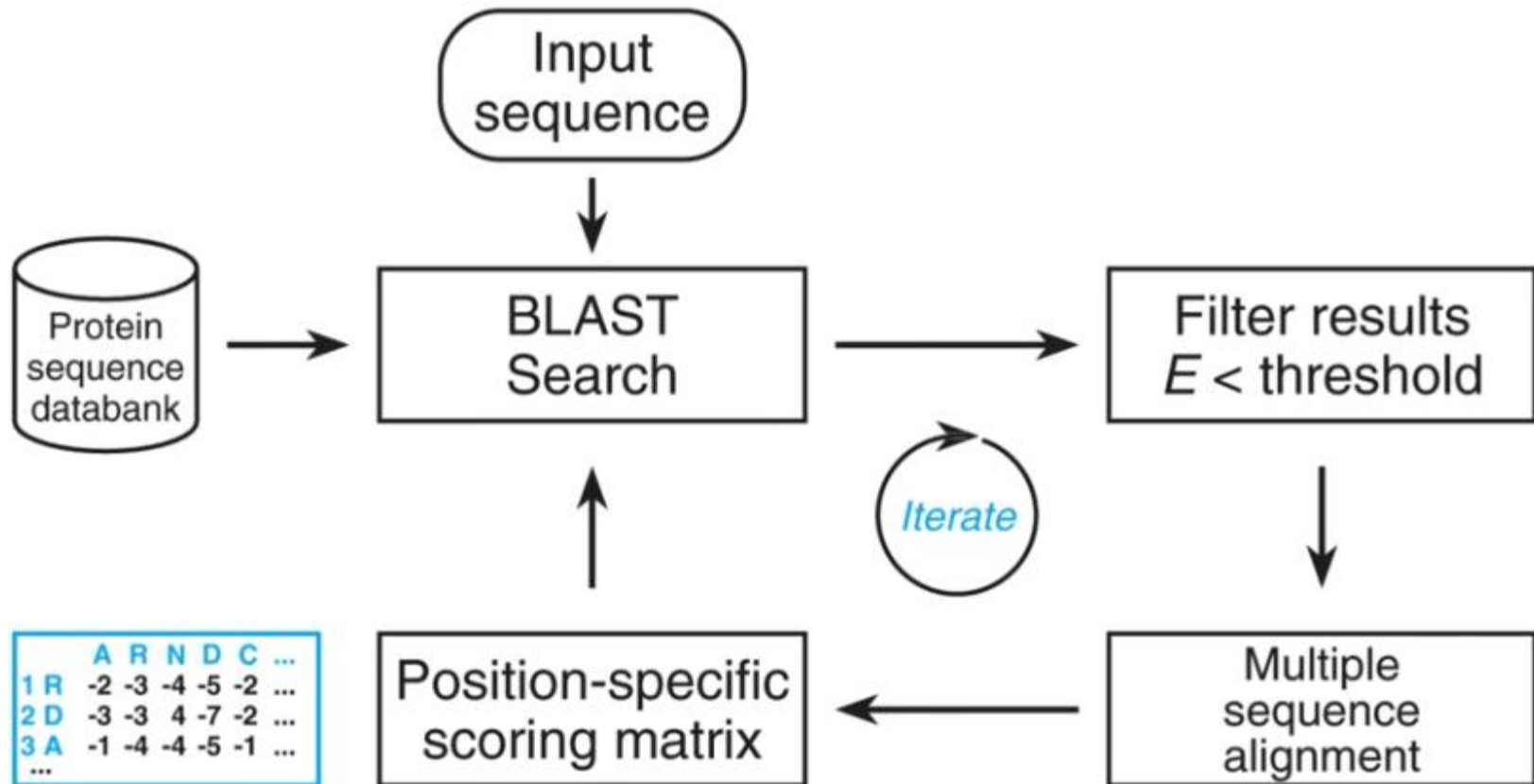
select all 20 sequences selected

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

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	<a href="#">osmotically-inducible protein OsmY [Rickettsiales bacterium]</a>	137	137	96%	1e-38	39.67%	<a href="#">MBB20570.1</a>
<input checked="" type="checkbox"/>	<a href="#">BON domain-containing protein [Rickettsiales bacterium]</a>	123	123	99%	7e-33	39.69%	<a href="#">MBB66425.1</a>
<input checked="" type="checkbox"/>	<a href="#">transport-associated protein [Rickettsiales bacterium]</a>	101	101	87%	2e-24	37.13%	<a href="#">MBV35553.1</a>
<input checked="" type="checkbox"/>	<a href="#">Uncharacterized protein in gshII 5' region [Candidatus Jidaibacter acanthamoeba]</a>	100	100	98%	1e-23	33.51%	<a href="#">KIE04947.1</a>
<input checked="" type="checkbox"/>	<a href="#">BON domain-containing protein [Candidatus Jidaibacter acanthamoeba]</a>	98.6	98.6	89%	4e-23	35.47%	<a href="#">WP_039456942.1</a>
<input checked="" type="checkbox"/>	<a href="#">BON domain-containing protein [Candidatus Aquarickettsia rohweri]</a>	82.4	82.4	91%	5e-17	32.22%	<a href="#">WP_128044820.1</a>
<input checked="" type="checkbox"/>	<a href="#">Osmotically-inducible protein Y precursor [Candidatus Arcanobacter lacustris]</a>	82.0	82.0	94%	8e-17	30.27%	<a href="#">KKB96100.1</a>
<input checked="" type="checkbox"/>	<a href="#">BON domain-containing protein [Candidatus Phycoarickettsia trachydisci]</a>	80.5	80.5	96%	2e-16	30.16%	<a href="#">WP_108874613.1</a>
<input checked="" type="checkbox"/>	<a href="#">BON domain-containing protein [Candidatus Hepatobacter penaei]</a>	74.3	74.3	87%	7e-14	32.94%	<a href="#">WP_031934417.1</a>
<input checked="" type="checkbox"/>	<a href="#">hypothetical protein [Rickettsiales bacterium]</a>	74.3	74.3	96%	9e-14	31.75%	<a href="#">MBO90079.1</a>
<input checked="" type="checkbox"/>	<a href="#">BON domain protein [Neorickettsia helminthoeca str. Oregon]</a>	70.1	70.1	96%	2e-12	28.57%	<a href="#">AHX11031.1</a>
<input checked="" type="checkbox"/>	<a href="#">BON domain-containing protein [Neorickettsia helminthoeca]</a>	70.1	70.1	96%	3e-12	28.57%	<a href="#">WP_051579482.1</a>
<input checked="" type="checkbox"/>	<a href="#">phospholipid-binding protein [Rickettsiales bacterium TMED174]</a>	68.6	68.6	94%	8e-12	25.41%	<a href="#">OUW24413.1</a>
<input checked="" type="checkbox"/>	<a href="#">BON domain-containing protein [Rickettsia typhi]</a>	68.9	68.9	92%	8e-12	29.12%	<a href="#">WP_011190840.1</a>
<input checked="" type="checkbox"/>	<a href="#">BON domain-containing protein [Candidatus Midichloria mitochondrii]</a>	67.0	67.0	89%	4e-11	28.07%	<a href="#">WP_013951652.1</a>
<input checked="" type="checkbox"/>	<a href="#">hypothetical protein [Rickettsiales bacterium]</a>	61.6	61.6	92%	4e-09	20.79%	<a href="#">MBH43598.1</a>
<input checked="" type="checkbox"/>	<a href="#">hypothetical protein [Rickettsiales bacterium]</a>	60.8	60.8	93%	6e-09	23.76%	<a href="#">MAY90257.1</a>
<input checked="" type="checkbox"/>	<a href="#">hypothetical protein [Rickettsiales bacterium]</a>	56.2	56.2	73%	3e-07	30.28%	<a href="#">MAZ46716.1</a>
<input checked="" type="checkbox"/>	<a href="#">hypothetical protein [Rickettsiales bacterium]</a>	55.1	55.1	96%	8e-07	25.27%	<a href="#">MBS91946.1</a>
<input checked="" type="checkbox"/>	<a href="#">hypothetical protein [Rickettsiales bacterium]</a>	53.1	53.1	86%	4e-06	28.74%	<a href="#">MBC11294.1</a>

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

## Psiblast



# Interface du programme PSI-Blast (site NCBI)

**Choose Search Set**

Database: Reference proteins (refseq\_protein) ?

Organism: Enterobacteraceae (taxid:543)  exclude Add organism  
Escherichia coli (taxid:562)  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:  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 using PSI-BLAST (Position-Specific Iterated BLAST)  
 Show results in a new window

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

**Algorithm parameters**

**General Parameters**

Max target sequences: 5000 ?  
Select the maximum number of aligned sequences to display ?

Short queries:  Automatically adjust parameters for short input sequences ?

Expect threshold: 0.001 ?

Word size: 3 ?

Max matches in a query range: 0 ?

**Scoring Parameters**

Matrix: BLOSUM62 ?

Gap Costs: Existence: 11 Extension: 1 ?

Compositional adjustments: Conditional compositional score matrix adjustment ?

**Filters and Masking**

Filter:  Low complexity regions ?

Mask:  Mask for lookup table only ?  
 Mask lower case letters ?

**PSI/PHI/DELTA BLAST**

Upload PSSM: Parcourir... Aucun fichier sélectionné. ?

PSI-BLAST Threshold: 0.005 ?

Pseudocount: 0 ?

Seuil E-value pour sélectionner la séquence lors de la recherche

Seuil E-value pour retenir la séquence pour la construction de la PSSM

# Extrait de la page résultat du premier run

[< 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: Enterobacteraceae (taxid:543) ; and exclude: Escherchia coli (taxid:562)

Job Title **EcoIA01.YRAP**  
 RID [ZU2WF9NT013](#) *Search expires on 02-05 18:18 pm* [Download All](#) ▾  
 Program PSI-BLAST Iteration 1 [Citation](#) ▾  
 Database refseq\_protein [See details](#) ▾  
 Query ID Icl|Query\_23531  
 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

**PSI-BLAST incl. threshold**

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

### Run PSI-Blast iteration 2

Number of sequences  [Run](#)

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

### Sequences producing significant alignments

[Download](#) ▾ [New](#) [Select columns](#) ▾ [Show](#)  [?](#)

1327 sequences selected

## 1327 séquences

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

### Sequences with E-value BETTER than threshold

select all 1327 sequences selected

PSI-BLAST iteration 1

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	Select for PSI blast	Used to build PSSM	Newly added
<input checked="" type="checkbox"/> <a href="#">divisome-associated lipoprotein YraP [Enterobacteriaceae]</a>	<a href="#">Enterobacteriaceae</a>	381	381	100%	5e-135	100.00%	191	<a href="#">WP_000848033.1</a>	<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/> <a href="#">divisome-associated lipoprotein YraP [Escherichia]</a>	<a href="#">Escherichia</a>	380	380	100%	1e-134	99.48%	191	<a href="#">WP_000848037.1</a>	<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/> <a href="#">divisome-associated lipoprotein YraP [Enterobacterales]</a>	<a href="#">Enterobacterales</a>	380	380	100%	1e-134	99.48%	191	<a href="#">WP_000848047.1</a>	<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/> <a href="#">divisome-associated lipoprotein YraP [unclassified Escherichia]</a>	<a href="#">unclassified Escherichia</a>	379	379	100%	2e-134	98.95%	191	<a href="#">WP_130204622.1</a>	<input checked="" type="checkbox"/>		

<input checked="" type="checkbox"/> <a href="#">BON domain-containing protein [Enterobacteriaceae]</a>	<a href="#">Enterobacteriaceae</a>	44.3	44.3	58%	7e-04	30.36%	104	<a href="#">WP_106928311.1</a>	<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/> <a href="#">BON domain-containing protein [Kluyvera]</a>	<a href="#">Kluyvera</a>	44.3	44.3	58%	7e-04	29.46%	104	<a href="#">WP_064543215.1</a>	<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/> <a href="#">molecular chaperone OsmY [Klebsiella pneumoniae]</a>	<a href="#">Klebsiella pneumoniae</a>	44.7	44.7	54%	7e-04	34.23%	127	<a href="#">WP_154718391.1</a>	<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/> <a href="#">BON domain-containing protein [Pantoea]</a>	<a href="#">Pantoea</a>	43.9	43.9	34%	8e-04	41.79%	104	<a href="#">WP_059011630.1</a>	<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/> <a href="#">BON domain-containing protein [Enterobacter ludwigii]</a>	<a href="#">Enterobacter ludwigii</a>	43.9	43.9	34%	8e-04	40.30%	104	<a href="#">WP_084738827.1</a>	<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/> <a href="#">molecular chaperone OsmY [Salmonella enterica]</a>	<a href="#">Salmonella enterica</a>	43.9	43.9	43%	8e-04	37.93%	94	<a href="#">WP_079948429.1</a>	<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/> <a href="#">molecular chaperone OsmY [Klebsiella pneumoniae]</a>	<a href="#">Klebsiella pneumoniae</a>	44.3	44.3	54%	0.001	34.23%	133	<a href="#">WP_102059708.1</a>	<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/> <a href="#">BON domain-containing protein [Trabulsisiella odontotermitis]</a>	<a href="#">Trabulsisiella odontotermitis</a>	43.9	43.9	31%	0.001	44.26%	104	<a href="#">WP_054177877.1</a>	<input checked="" type="checkbox"/>		

Run PSI-BLAST Iteration 2 with max number of sequences  [Run](#)

# Extrait de la page résultat du second run

[← 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: Enterobacteraceae (taxid:543); and exclude: Escherchia coli (taxid:562)

Job Title **EcolA01.YRAP**

RID [ZU33N1WJ013](#) Search expires on 02-05 18:22 pm [Download All](#) ▾

Program PSI-BLAST Iteration 2 [Citation](#) ▾

Database **refseq\_protein** [See details](#) ▾

Query ID **Id|Query\_23531**

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

Type common name, binomial, taxid or group name

[+ Add organism](#)

**Percent Identity**  to

**E value**  to

**Query Coverage**  to

**PSI-BLAST incl. threshold**

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

### Run PSI-Blast iteration 3

Number of sequences  [Run](#)

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

**Sequences producing significant alignments** [Download](#) ▾ [Select columns](#) ▾ Show  **?**

**1471 séquences**

[1471 sequences selected](#)  sequences newly added this iteration **?**

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

**Sequences with E-value BETTER than threshold**

select all 1471 sequences selected [Skip to the first new sequence](#)

PSI-BLAST iteration 2

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	Select for PSI blast	Used to build PSSM	Newly added
<input checked="" type="checkbox"/>	<a href="#">divisome-associated lipoprotein YraP [Kosakonia oryzae]</a>	<a href="#">Kosakonia oryzae</a>	251	251	100%	9e-84	84.82%	191	<a href="#">WP_064563421.1</a>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	<a href="#">divisome-associated lipoprotein YraP [Kosakonia]</a>	<a href="#">Kosakonia</a>	251	251	100%	1e-83	85.34%	191	<a href="#">WP_043956417.1</a>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	



# Extrait de la page résultat du second run

En jaune les  
nouvelles  
séquences  
détectées  
avec la PSSM

<input checked="" type="checkbox"/>	molecular chaperone OsmY [Raoultella terrigena]	Raoultella terrigena	143	143	96%	4e-41	28.72%	203	WP_134626226.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	molecular chaperone OsmY [Raoultella]	Raoultella	143	143	96%	4e-41	29.23%	203	WP_088911868.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	molecular chaperone OsmY [Raoultella]	Raoultella	143	143	96%	5e-41	28.72%	203	WP_041143944.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	divisome-associated lipoprotein YraP [Klebsiella aerogenes]	Klebsiella aerogenes	140	140	59%	9e-41	83.19%	113	WP_021468202.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	molecular chaperone OsmY [Cedecea neteri]	Cedecea neteri	142	142	95%	1e-40	25.91%	203	WP_038475795.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	molecular chaperone OsmY [Shigella flexneri]	Shigella flexneri	142	142	95%	1e-40	26.98%	192	WP_025760760.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	molecular chaperone OsmY [Buttiauxella sp. B2]	Buttiauxella sp. B2	142	142	95%	1e-40	27.98%	205	WP_139881715.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	molecular chaperone OsmY [Enterobacter sp. Colony194]	Enterobacter sp. Colony194	142	142	86%	1e-40	26.74%	207	WP_220267448.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	molecular chaperone OsmY [Mangrovibacter vixingensis]	Mangrovibacter vixingensis	142	142	95%	2e-40	28.12%	205	WP_226571978.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	molecular chaperone OsmY [Klebsiella aerogenes]	Klebsiella aerogenes	142	191	96%	2e-40	29.69%	203	WP_047065619.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	molecular chaperone OsmY [Cedecea neteri]	Cedecea neteri	142	142	95%	2e-40	25.39%	203	WP_039290058.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	molecular chaperone OsmY [Klebsiella aerogenes]	Klebsiella aerogenes	142	191	96%	2e-40	29.69%	203	WP_125992599.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	molecular chaperone OsmY [Klebsiella aerogenes]	Klebsiella aerogenes	142	191	96%	2e-40	29.69%	203	WP_045358894.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	molecular chaperone OsmY [Mangrovibacter]	Mangrovibacter	142	142	95%	2e-40	28.12%	205	WP_036106434.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	molecular chaperone OsmY [Cedecea sp. NFIX57]	Cedecea sp. NFIX57	142	142	95%	2e-40	24.87%	203	WP_085541367.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	molecular chaperone OsmY [Klebsiella aerogenes]	Klebsiella aerogenes	141	191	96%	2e-40	29.69%	203	WP_083413924.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	molecular chaperone OsmY [Klebsiella michiganensis]	Klebsiella michiganensis	141	141	95%	3e-40	25.91%	203	WP_045781267.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	molecular chaperone OsmY [Klebsiella aerogenes]	Klebsiella aerogenes	141	191	96%	3e-40	29.69%	203	WP_047740078.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	molecular chaperone OsmY [Cedecea neteri]	Cedecea neteri	141	141	95%	3e-40	25.39%	203	WP_061272729.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	molecular chaperone OsmY [Klebsiella aerogenes]	Klebsiella aerogenes	141	191	96%	3e-40	29.17%	203	WP_015368385.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	molecular cha	BON domain-containing protein [Edaphovirga cremea]	Edaphovirga cremea	63.1	122	76%	5e-11	31.94%	104	WP_114193732.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	molecular cha	BON domain-containing protein [Enterobacter hormaechei]	Enterobacter hormaechei	61.5	61.5	40%	1e-10	21.43%	83	WP_163284324.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	molecular cha	BON domain-containing protein [Buttiauxella]	Buttiauxella	61.5	118	87%	2e-10	27.17%	104	WP_034492695.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	molecular cha	BON domain-containing protein [Klebsiella pneumoniae]	Klebsiella pneumoniae	60.4	60.4	39%	3e-10	20.48%	82	WP_117047656.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	molecular cha	BON domain-containing protein [Enterobacter kobei]	Enterobacter kobei	60.4	60.4	39%	4e-10	28.57%	78	WP_163310730.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	molecular cha	BON domain-containing protein [Buttiauxella sp. A111]	Buttiauxella sp. A111	60.4	114	97%	6e-10	27.03%	104	WP_183269294.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	molecular cha	BON domain-containing protein [Raoultella planticola]	Raoultella planticola	58.9	58.9	19%	2e-09	100.00%	38	WP_163968914.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	molecular cha	BON domain-containing protein [Klebsiella aerogenes]	Klebsiella aerogenes	58.0	58.0	39%	3e-09	22.89%	91	WP_155958965.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	molecular cha	BON domain-containing protein [Klebsiella pneumoniae]	Klebsiella pneumoniae	58.8	58.8	72%	1e-08	23.57%	184	WP_136666254.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	molecular cha	hypothetical protein [Klebsiella pneumoniae]	Klebsiella pneumoniae	55.7	55.7	41%	1e-08	31.65%	69	WP_117064949.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	molecular cha	BON domain-containing protein [Klebsiella pneumoniae]	Klebsiella pneumoniae	55.7	105	64%	1e-08	31.75%	64	WP_117085659.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	molecular cha	BON domain-containing protein [Candidatus Regiella insecticola]	Candidatus Regiella insecticola	57.7	57.7	38%	2e-08	32.47%	152	WP_006704449.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	molecular cha	BON domain-containing protein [Klebsiella pneumoniae]	Klebsiella pneumoniae	54.6	54.6	35%	5e-08	21.43%	73	WP_116431086.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	molecular cha	BON domain-containing protein [Trabulsiiella odontotermitis]	Trabulsiiella odontotermitis	54.2	107	65%	7e-08	40.00%	75	WP_049857391.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	molecular cha	BON domain-containing protein [Candidatus Regiella insecticola]	Candidatus Regiella insecticola	55.4	55.4	37%	1e-07	32.00%	161	WP_178467774.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	molecular cha	BON domain-containing protein [Klebsiella pneumoniae]	Klebsiella pneumoniae	52.3	99.1	58%	2e-07	35.05%	58	WP_117068330.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	molecular cha	BON domain-containing protein [Trabulsiiella odontotermitis]	Trabulsiiella odontotermitis	52.7	104	70%	3e-07	40.00%	76	WP_049857428.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	molecular cha	BON domain-containing protein [unclassified Shigella]	unclassified Shigella	52.7	103	98%	7e-07	16.22%	114	WP_069369725.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	molecular cha	BON domain-containing protein [Klebsiella michiganensis]	Klebsiella michiganensis	50.0	50.0	40%	7e-06	26.58%	125	WP_220378607.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	molecular cha	BON domain-containing protein [Klebsiella pneumoniae]	Klebsiella pneumoniae	47.6	91.4	49%	1e-05	34.04%	51	WP_146700819.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	molecular cha	BON domain-containing protein [Enterobacter ludwigii]	Enterobacter ludwigii	47.3	90.7	76%	3e-05	23.46%	85	WP_049868897.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

# Résultats suite runs consécutifs PSI-BLAST

[← Edit Search](#) [Save Search](#) [Search Summary](#) **Itération 3** [How to](#)

**i** Your search is limited to records that include: Enterobacteraceae (taxid:543) ; and exclude: Escherchia coli (taxid:562)

Job Title	EcoIA01.YRAP
RID	<a href="#">ZUE9WSDZ016</a> Search expires on 02-05 21:33 pm <a href="#">Download All</a> <b>v</b>
Program	PSI-BLAST Iteration 3 <a href="#">Citation</a> <b>v</b>
Database	refseq_protein <a href="#">See details</a> <b>v</b>
Query ID	Icl Query_23867
Description	EcoIA01.YRAP
Molecule type	amino acid
Query Length	191
Other reports	<a href="#">Distance tree of results</a> <a href="#">Multiple alignment</a> <a href="#">MSA viewer</a> <b>?</b>

### Filter Results

**Organism** *only top 20 will appear*

[+ Add organism](#)

**Percent Identity**  to  **E value**  to

**PSI-BLAST incl. threshold**

**Run PSI-Blast iteration 4**

**Number of sequences**

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

**Sequences producing significant alignments**

## 1474 séquences

1474 sequences selected  sequences newly added this iteration **?** [GenPept](#)

**Sequences with E-value BETTER than**

select all 1474 sequences selected

[← Edit Search](#) [Save Search](#) [Search Summary](#) **Itération 4** [How to read this report?](#) [BLAST Help Videos](#) [Back to Tradition](#)

**i** Your search is limited to records that include: Enterobacteraceae (taxid:543) ; and exclude: Escherchia coli (taxid:562)

Job Title	EcoIA01.YRAP
RID	<a href="#">ZUEG4SAF016</a> Search expires on 02-05 21:36 pm <a href="#">Download All</a> <b>v</b>
Program	PSI-BLAST Iteration 4 <a href="#">Citation</a> <b>v</b>
Database	refseq_protein <a href="#">See details</a> <b>v</b>
Query ID	Icl Query_23867
Description	EcoIA01.YRAP
Molecule type	amino acid
Query Length	191
Other reports	<a href="#">Distance tree of results</a> <a href="#">Multiple alignment</a> <a href="#">MSA viewer</a> <b>?</b>

## Itération 4

### Filter Results

**Organism** *only top 20 will appear*

[+ Add organism](#)

**Percent Identity**  to  **E value**  to  **Query Coverage**  to

**PSI-BLAST incl. threshold**

[Filter](#)

**Run PSI-Blast iteration 5**

**Number of sequences**

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

**Sequences producing significant alignments** [Download](#) [New Select columns](#) [Show](#)

1474 sequences selected  sequences newly added this iteration **?** [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#) **New**

**⚠** No new sequences were found above the 0.005 threshold