


# Blast

- BLAST : Basic Local Alignment Search Tool
- Références
  - Altschul et al. Basic local alignment search tool. *J Mol Biol.* 1990 Oct 5;215(3):403-10.
  - Altschul et al. (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402.
- Accessibilité
  - ftp, packages (dont GCG)
  - Nombreux serveurs Web
    - Au NCBI : <http://www.ncbi.nlm.nih.gov/BLAST/>
    - ...

Odile Lecompte -IGBMC ASMI

# Algorithme de blast

## Construction de la liste des mots similaires aux mots de la query


 Séquence requête (query)  
 tous les mots de longueur w

Pour les protéines : w= 3 par défaut  
 Pour l'ADN : w = 11 par défaut

**Mot 1**

Parmi tous les mots possibles de longueur w (20x20x20=8000 possibilités si w=3) :  
 Sélection des mots dont l'alignement avec le mot de la séquence query donne un score  $\geq T$   
 => protéines : matrice de scores (blosum62 par défaut)  
 => ADN : match = 2, mismatch -3

Ex : PQG	→	PQG	Score = 7 + 5 + 6 = 18	Si T=13 → Élimination de PQA
		PEG	Score = 15	
		PQA	Score = 12	
		...	...	

**Mot 2** →

Remarque : dans le blast original, T=13  
~ 50 mots atteignent ce seuil

Etc...

Odile Lecompte -IGBMC ASMI

# Algorithme de Blast

## Localisation des mots sur les séquences de la banque

Chaque « hit » est identifié

Il s'agit d'une liste finie et pré-établie.

Odile Leconte -IGBMC ASMI

# Algorithme de Blast

## Construction des HSP (Blast1)

**Construction des HSPs (Blast original)**

Pour chaque hit, extension de l'alignement (dans les 2 directions) tant que:

- l'extrémité d'une des séquences n'est pas atteinte
- le score ne chute pas plus d'une valeur X en dessous du meilleur score obtenu jusque-là pour cet alignement

L'alignement ainsi obtenu est appelé **HSP** (High-scoring segment pair) et a un score supérieur au score du hit initial

Seules les HSPs dont le score est supérieur à un seuil fixé sont retenues.

Il peut exister plusieurs HSPs entre deux séquences (leur score est alors combiné).

absence de gap dans les alignements de Blast1 ⇒ « **ungapped** » blast

Odile Leconte -IGBMC ASMI

# Algorithme de Blast

## Construction des HSP (Blast2=gapped blast)

**Extension sans gap : « Two-hits method »**  
 au moins 2 hits non chevauchants  
 situés sur la même diagonale (pas de gaps)  
 situés à une distance  $\leq A$

Dans blast2 : T=11

**Exemple:**  
 + 15 hits avec score > 13  
 . 22 hits additionnels avec un score > 11

2 paires de hits sont sur une même diagonale et à une distance < 40

Blast original (T=13) : 15 tentatives d'extension  
 Blast2 (T=11) : 2 tentatives d'extension  
 => gain de temps

ASMI

# Algorithme de Blast

## Alignement optimal (Smith & Waterman)

- Sélection des HSPs avec score suffisant
- Choix d'une paire de résidus (amorce) dans l'HSP : paire centrale du meilleur segment de 11 paires
- Alignement optimal « limité » à partir de l'amorce => introduction de **gaps**

L'alignement optimal se fait en considérant seulement les alignements qui ne tombent pas plus de Xg en dessous du meilleur score obtenu.

ASMI

Odile Lecompte -IGBMC

## Algorithme de Blast

Alignement optimal limité

	L	N	A	K	S	I	M	W	Q	A	T	R	C	I	S	V
Y									/	/	/	/	/	/	/	/
C									/	/	/	/	/	/	/	/
W									/	/	/	/	/	/	/	/
Q	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
A	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
T	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
D	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
S	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
G	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/

Odile Lecompte -IGBMC ASMI

## Algorithme de Blast

Alignement optimal obtenu

Lettre : identité  
+ : score positif

Alignements obtenus par Blast1

```

Leghemoglobin 43 FSPFKDSAGVVDSPKLGAAAEKVFQGMVRDSAVOLRATGEV--LDGKDGSD----- 90
                   F L + V+ +PK+ AH +KV L + GE V LD G+
Beta globin 45 FGDLSNPGAVMGNPKVKAHGKKV-----LHSPGEGVHLDNLKGTFAALSE 90
                   ^
                   |
                   | amorce
Leghemoglobin 91 IHIQGVLDP-HFVVVKEALLKTIKEASGDKWSEELSAAAEVAYDGLATAT 140
                   +H K +DP +F ++ L+ + G ++ EL A+++ G+A A+
Beta globin 91 LHCDKLEVDPENFRLLGNVLVVVLARHFGKDFPELQASYQKVYAGVANAL 141
    
```

Odile Lecompte -IGBMC ASMI



# Les différentes comparaisons

Programmes	Requête	Banque	Comparaison	Exemples d'utilisation
Blastn	ADN	ADN	nucléique	Recherche d'ARN structuraux, d'éléments régulateurs
Blastp	Protéine	protéines	protéique	Recherche de protéines homologues
Tblastn	Protéine	ADN (traduit dans les 6 cadres)	protéique	Recherche de similarités entre une protéine et une séquence génomique mal annotée
Blastx	ADN (traduit dans les 6 cadres)	protéines	protéique	Recherche des phases de lecture dans une séquence codante
Tblastx	ADN (traduit dans les 6 cadres)	ADN (traduit dans les 6 cadres)	protéique	Avantages de tblastn et blastx mais très long

Megablast => optimisé pour des séquences nucléiques quasi-identiques (>95% identité)  
(taille des mots = 28), très rapide



## BLAST Assembled Genomes

Choose a species genome to search, or [list all genomic BLAST databases](#)

- [Human](#)
- [Mouse](#)
- [Rat](#)
- [Arabidopsis thaliana](#)
- [Oryza sativa](#)
- [Bos taurus](#)
- [Danio rerio](#)
- [Drosophila melanogaster](#)
- [Gallus gallus](#)
- [Pan troglodytes](#)
- [Microbes](#)
- [Apis mellifera](#)

## Basic BLAST

Choose a BLAST program to run.

- [nucleotide blast](#) Search a **nucleotide** database using a **nucleotide** query  
*Algorithms:* blastn, megablast, discontinuous megablast
- [protein blast](#) Search **protein** database using a **protein** query  
*Algorithms:* blastp, psi-blast, phi-blast
- [blastx](#) Search **protein** database using a **translated nucleotide** query
- [tblastn](#) Search **translated nucleotide** database using a **protein** query
- [tblastx](#) Search **translated nucleotide** database using a **translated nucleotide** query

## Specialized BLAST

Choose a type of specialized search (or database name in parentheses.)

- Make specific primers with [Primer-BLAST](#)
- Search [trace archives](#)
- Find [conserved domains](#) in your sequence (cds)
- Find sequences with similar [conserved domain architecture](#) (cdart)
- Search sequences that have [gene expression profiles](#) (GEO)
- Search [immunoglobulins](#) (IgBLAST)
- Search for [SNPs](#) (snp)
- Screen sequence for [vector contamination](#) (vecscreen)
- [Align](#) two (or more) sequences using BLAST (bl2seq)
- Search [protein](#) or [nucleotide](#) targets in PubChem BioAssay
- Search SRA [transcript libraries](#)
- Constraint Based Protein [Multiple Alignment Tool](#)

# Interface Blast (NCBI)

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

Enter Query Sequence

Enter accession number, gi, or FASTA sequence  Clear Query subrange

Or, upload file  Percentage

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 (MM/P)  Uncultured/environmental sample sequences

Optional

Entrez Query

Optional

Program Selection

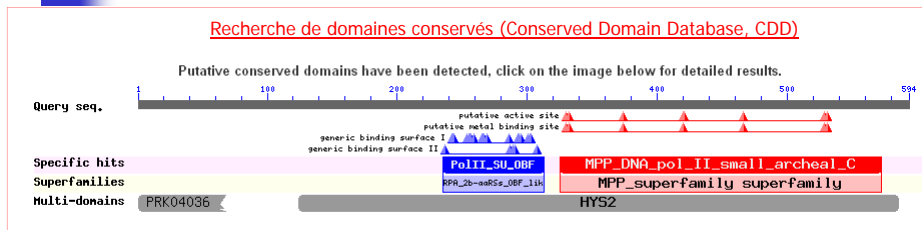
Algorithm  blastp (protein-protein BLAST)  PSI-BLAST (Position-Specific Iterated BLAST)  PHI-BLAST (Pattern Hit Initiated BLAST)

Search database Swissprot protein sequences(swissprot) using Blastp (protein-protein BLAST)  Show results in a new window

Odile Lecompte -IGBMC

ASMI

# Sortie Blast



<http://www.ncbi.nlm.nih.gov/sites/entrez?db=cdd>

[NCBI-curated domains](#)  
(use 3D-structure information)

- Domain models from
- [Pfam](#)
- [SMART](#)
- [COG](#)
- [PRK](#)
- [TIGRFAM](#)

Odile Lecompte -IGBMC

ASMI

# Sortie Blast

View concise result ?

**Conserved domains on** [gi|2833545|sp|Q58113|]

RecName: Full=DNA polymerase II small subunit; Short=Pol II

**Graphical summary** show options >

**Query seq.** 100 200 300 400 500 594

**Specific hits**

**Non-specific hits**

**Superfamilies**

**Multi-domains** PRK04036 PRK04036 HYS2 Metallophos Icc

[Search for similar domain architectures](#) ? [Refine search](#) ?

Hit	Description	Pssmid	Multi-dom	Evalue
H MPP_DNA_pol_II_small_archaeal_C[ <a href="#">cd07386</a> ]	The small subunit of the archaeal DNA polymerase II contains a C-terminal ...	163629	no	9.65e-122
H PolII_SU_OBF[ <a href="#">cd04490</a> ]	PolII_SU_OBF: A subfamily of OB folds corresponding to the OB fold found in Pyrococcus ...	72962	no	5.74e-21
H MPP_YydB[ <a href="#">cd07400</a> ]	YydB (BSU40220) is an uncharacterized Bacillus subtilis protein that belongs to the ...	163643	no	3.36e-06
H MPP_PoID2_C[ <a href="#">cd07387</a> ]	PoID2 (DNA polymerase delta, subunit 2) is an auxiliary subunit of the eukaryotic DNA ...	163630	no	4.40e-06
H DNA_poI_E_B[ <a href="#">fam04043</a> ]	This family contains a number of DNA polymerase subunits. The B subunit of the DNA ...	148591	no	1.92e-08
H MPP_YbbF-LpxH[ <a href="#">cd07390</a> ]	YbbF/LpxH is an Escherichia coli UDP-2,3-diacetylglucosamine hydrolase thought to catalyze ...	163641	no	2.13e-03
H MPP_Mre11_N[ <a href="#">cd00840</a> ]	Mre11 (also known as SbcD in Escherichia coli) is a subunit of the MRX protein complex ...	163616	yes	7.33e-03
H PRK04036[ <a href="#">PRK04036</a> ]	DNA polymerase II small subunit; Validated	167687	yes	0e+00
H HYS2[ <a href="#">COG1311</a> ]	Archaeal DNA polymerase II, small subunit/DNA polymerase delta, subunit B [DNA ...	31502	yes	6.74e-135
H PRK04036[ <a href="#">PRK04036</a> ]	DNA polymerase II small subunit; Validated	167687	yes	2.72e-10
H Metallophos[ <a href="#">pfam00149</a> ]	This family includes a diverse range of phosphoesterases, including protein phosphoserine ...	143917	yes	2.04e-05
H Icc[ <a href="#">COG1409</a> ]	Predicted phosphohydrolases [general function prediction only]	31599	yes	5.76e-03

Odile Lecompte -IGBMC

ASMI

# Sortie Blast

Banque et séquence requête utilisées

**sp|Q58113|** (594 letters)

<b>Query ID</b>	gi 2833545 sp Q58113.1 DP2S_METJA	<b>Database Name</b>	swissprot
<b>Description</b>	DNA polymerase II small subunit [Methanocaldococcus jannaschii DSM 2661] >gi 2833545 sp Q58113.1 DP2S_METJA RecName: Full=DNA polymerase II small subunit; Short=Pol II >gi 1591417 gb AAB98694.1  DNA polymerase delta small subunit [Methanocaldococcus jannaschii DSM 2661]	<b>Description</b>	Non-redundant SwissProt sequences
<b>Molecule type</b>	amino acid	<b>Program</b>	BLASTP 2.2.24+ <a href="#">Citation</a>
<b>Query Length</b>	594		

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [Multiple alignment](#)

**Distribution of 12 Blast Hits on the Query Sequence** ?

Mouse-over to show define and scores, click to show alignments

**Color key for alignment scores**

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

Odile Lecompte -IGBMC

ASMI

## Sortie

---

**Sequences producing significant alignments:**

Accession	Description	Max score	Total score	Query coverage	E value
<a href="#">Q58113.1</a>	DNA polymerase II small subunit [Methanocaldococcus jannaschii DSM 2661]	1189	1189	100%	0.0
<a href="#">Q28484.1</a>	RecName: Full=DNA polymerase II small subunit; Short=Pol II	347	347	65%	9e-95
<a href="#">Q27456.1</a>	RecName: Full=DNA polymerase II small subunit; Short=Pol II	324	324	65%	1e-87
<a href="#">P81412.1</a>	RecName: Full=DNA polymerase II small subunit; Short=Pol II; AltName: Full=	294	294	98%	1e-78
<a href="#">Q57863.1</a>	RecName: Full=DNA polymerase II small subunit; Short=Pol II	288	288	58%	9e-77
<a href="#">Q9HMR7.2</a>	RecName: Full=DNA polymerase II small subunit; Short=Pol II >sp B0R7U1.1	283	283	65%	2e-75
<a href="#">Q9VDF3.1</a>	RecName: Full=DNA polymerase II small subunit; Short=Pol II	282	282	63%	5e-75
<a href="#">Q9HLK5.1</a>	RecName: Full=DNA polymerase II small subunit; Short=Pol II	281	281	63%	1e-74
<a href="#">Q48520.1</a>	RecName: Full=DNA polymerase delta small subunit	65.9	65.9	44%	7e-10
<a href="#">Q9LE85.1</a>	RecName: Full=DNA polymerase delta small subunit	62.0	62.0	38%	1e-08
<a href="#">Q6G071.2</a>	RecName: Full=Kinetochore protein Nuf2-B; AltName: Full=Cell division cycle-	33.1	33.1	15%	5.4
<a href="#">Q9CE62.1</a>	RecName: Full=2',3'-cyclic-nucleotide 2'-phosphodiesterase	32.7	32.7	23%	7.9

Odile Lecompte -IGBMCASMI

## Blast : alignements

Séquence requête (query)

Séquence de la banque

```

>|sp|027456.1|DP2S_METH|G RecName: Full=DNA polymerase II small subunit; Short=Pol II
Length=482

GENE ID: 1471122_MTH1405 | DNA polymerase II small subunit
[Methanothermobacter thermautotrophicus str. Delta H] (10 or fewer PubMed links)

Score = 324 bits (830), Expect = 1e-87, Method: Compositional matrix adjust.
Identities = 164/392 (42%), Positives = 251/392 (65%), Gaps = 8/392 (2%)

Query 183 SDVSGKSTCTGTIEDIVKYYFRDRFERLKVFIERKAQRKGY-PLKDIKMKGQKDFVVGVI 241
          SD S +S +G I D + YF R+ LK + + + + RG+ P+ D+ + G+ + + +G+
Sbjct 87  SPTSNSRSTSGEIGMIAVFNRSVSSLKMLLSKRPFLKCHIFIAIDLRG--GEDVVSIIIGM 144

Query 242 VSDVDSTRGNLIVRIEDTEDEATLLPREKIEAGKIPDDILLDEVIGAIQTVSKSGSSI 301
          V+DV +I+N + I+ +ED E + + + + E + + + + I+ DEV+G GT K G +
Sbjct 145 VMDVRNTKMNRHRIELEDTGEISVVVHNEHMKLFEKSEKIVRDEVVGVHGT--KGRFV 202

Query 302 YVDEIIRPALPKPEKRIDEEIIYMAFLSDIHVGSKEFLHKEFEKFIIRFLNGDVDNELEEK 361
          EI P +P + K +D +AF+SD+H+G3+ FL F KF+++HGD +E +
Sbjct 203 VASEIFHPGVPRIQEREMD--FSVAFISDVHIGSQTFLLEDAPMKFVKWINGDFGSEEQRS 260

Query 362 VVSRKLYICLAGDLVDCGVVPGQEDLYEVDLIEQYREIAMVLDQIPEHISIIISPGNH 421
          + + +KY+ +AGD+VDG+G+YPGQE++L DI EQY E A I I I+ PGNH
Sbjct 261 LAADVKTLLVAGDIDVGIQIYPGQEKLLIRDHEQYEEAARLFGDIRSDIRIVMIPGNH 320

Query 422 DAVRPAEPQPKLPEKITK-LFNRMNLYFVGNPCTLNHIGFDLLYHGRSFDLDLVGGIRAA 480
          D+ R AEPQP +PE+ K L+ + NI F+ NP + + + G TL+YHGRSFD+ +
Sbjct 321 DSSRIAEQPAIPEEYAKSLYSIRNIEFLSNFSLVSLDGVRTLIVHGRSFDMDAMSVMGL 380

Query 481 SYENPVTIMKELIKRLLCPTYGRCPIAPEHKDYLVIDRDIIDLHTGHIMHNGYIYRG 540
          S+E IM+EL+++R L P YG R P+A E +D+LVID +LHTGH+HIN Y Y+G
Sbjct 381 SHERSOLIMEELLEKRHLAPYIGERTPLASEIEDHLVIDEVHVLHTGHVHINAYKRYG 440

Query 541 VVMNSGTFOGQTDFOKRWGISPTPALVPIIN 572
          V ++NSGTFO QT+FOK I FT VP++N
Sbjct 441 VHLINSGTFOGQTEFKQIYNIYPTCGQVPLN 472

```

Odile Lecompte -IGBMCASMI





# Blast : alignements

```

>|_sp|06G071.2|NUF2B_XENLA| RecName: Full=Kinetochore protein Nuf2-B; AltName: Full=Cell
division cycle-associated protein 1-B
Length=462

GENE ID: 443588 muF2-b | NUF2, MDC80 kinetochore complex component, homolog
[Xenopus laevis] (10 or fewer PubMed links)

Score = 33.1 bits (74), Expect = 5.4, Method: Compositional matrix adjust.
Identities = 26/99 (27%), Positives = 51/99 (52%), Gaps = 7/99 (7%)

Query 129 EKLKQL-ISKEEKEDFDAERAKSYE---HITKIKESVNSRIKWIANDIDAVIYIYE--- 181
EKL ++ I ++K+EDF+ ++ + E HI + ++ ++ R+ + ++ I I E
Sbjct 350 EKLAKVDIKMKKKQEDFMQKQELLEVCSHIQEKQKQVINGRVAQVLQEIQTISKKEQLL 409

Query 182 DSDVSGKSTCTGTIEDFVKYFRDRFERLKVFIERKAQRK 220
++ +GK+ C I DF + L+ ER A R+
Sbjct 410 ETTEAGMKCQEVITDFFAALEKYHDSLQKASERSADRR 448

>|_sp|09CEE2.1|CNPD_LACLA| RecName: Full=2',3'-cyclic-nucleotide 2'-phosphodiesterase
Length=531

Score = 32.7 bits (73), Expect = 7.9, Method: Compositional matrix adjust.
Identities = 37/144 (26%), Positives = 74/144 (52%), Gaps = 15/144 (10%)

Query 78 YTTGEEEEKPKPEVSGTEIKKETEKIEKIEKIEFVKKEKEQPIKKSDEDVEEKLKQLISK 137
Y EEE+++ +E++ E K+E +E EK + +++E+ + + D+ + +K + L SK
Sbjct 71 YTLREEEQKQREIEDEFKQERQELKETEK----RLKQREILDRKDDTLTKGEENLDSK 126

Query 138 EE---KREDFDAERAKRYEHITKIKESVNSRIKWIANDIDAVIYIYEDSDVSGKSTCTGT 194
EE +K D +HR ++ HI + K RI ++ D I + E D G
Sbjct 127 EENLVKRTDLSKREEQLAHIEEKRFLELERISNLSTDDAREIILSETRD-----GL 178

Query 195 IEDFVKYFRDRFERLKVFIERKAQ 218
++ + R E+ +Y ++++
Sbjct 179 TKEMAQIIRQSEEAQVEADKRAK 202

```

**Biais en composition**



# Filtres

General Parameters

Max target sequences: 100  
Select the maximum number of aligned sequences to display

Short queries:  Automatically adjust parameters for short input sequences

Expect threshold: 10

Word size: 3

Max matches in a query range: 0

Scoring Parameters

Matrix: BLOSUM62

Gap Costs: Existence: 11 Extension: 1

Compositional adjustments: No adjustment

Filters and Masking

Filter:  Low complexity regions

Mask:  Mask for lookup table only  
 Mask lower case letters

## Filtres

Certaines régions peuvent être filtrées :

- faible complexité
- courts motifs répétés
- éléments répétés dispersés connus (ALUs repeats...)

=> non prises en compte dans la recherche

Odile Lecompte -IGBMC ASMI

## Filtres

>ref|NP\_247686.1| **G** DNA polymerase II small subunit [Methanocaldococcus jannaschii DSM 2661]

sp|Q58113.1|DP23\_METVA RecName: Full=DNA polymerase II small subunit; Short=Pol II

gb|AA098694.1| **G** DNA polymerase delta small subunit [Methanocaldococcus jannaschii DSM 2661]

Length=594

Score = 1093 bits (2828), Expect = 0.0

Identities = 594/594 (100%), Positives = 594/594 (100%), Gaps = 0/594 (0%)

```

Query 1  MEIINKFLDLEALLSPTVYEKLNFDDEKLRLLIQKIREFKKYNNAFILLDEKFLDIFLQ 60
Sbjct 1  MEIINKFLDLEALLSPTVYEKLNFDDEKLRLLIQKIREFKKYNNAFILLDEKFLDIFLQ 60
Query 61  KDLDLITNEVVDLELETTGeeekpkpkvkkkaiikketeekiekekiefvkkkeekqfi 120
FIFYYTGeeekpkpkvkkkaiikketeekiekekiefvkkkeekqfi
FIFYYTGEEEEKPKPEVKKEIKKETEKKIEKIEKIEFVKKEEKEQFI
FIFYYTGEEEEKPKPEVKKEIKKETEKKIEKIEKIEFVKKEEKEQFI
Query 181  EDSVSGKSTCTGTIEDFVKYFDRFRERLKVFIERRAQRKGYPLKD IKKMGQKDIFVVG 240
Sbjct 181  EDSVSGKSTCTGTIEDFVKYFDRFRERLKVFIERRAQRKGYPLKD IKKMGQKDIFVVG 240
Query 241  IVSDVSDTRNGNLIVRIEDTEDEATLLPKEKIEAGKIPDDILLDEVIGAIQTVSKSGSS 300
Sbjct 241  IVSDVSDTRNGNLIVRIEDTEDEATLLPKEKIEAGKIPDDILLDEVIGAIQTVSKSGSS 300
Query 301  IYVDEIIRFALPPEKPRIDEEIYMAFLSDIHVGSKEFLHKEFEKFIIRFLNGVDNELEE 360
Sbjct 301  IYVDEIIRFALPPEKPRIDEEIYMAFLSDIHVGSKEFLHKEFEKFIIRFLNGVDNELEE 360
Query 361  KVSRLKYICLAGDLVDGCVYPGQEDLVEVDLIEQYREIANYLDQIPEHISIIISPGN 420
Sbjct 361  KVSRLKYICLAGDLVDGCVYPGQEDLVEVDLIEQYREIANYLDQIPEHISIIISPGN 420
Query 421  HDAVRPAEPQPKLPEKITKLFNRDNIYFVGNPCTLMIHGFDTLLYHGRSFDLVGQIRAA 480
Sbjct 421  HDAVRPAEPQPKLPEKITKLFNRDNIYFVGNPCTLMIHGFDTLLYHGRSFDLVGQIRAA 480
Query 481  SYENPVTIMKELIKRRLCPTYGGRCPAIEPKHGYLVIDRDIILHTGHINGYGIYRG 540
Sbjct 481  SYENPVTIMKELIKRRLCPTYGGRCPAIEPKHGYLVIDRDIILHTGHINGYGIYRG 540
Query 541  VVMNSGTFQEQTD FQKRMGISPTFAIVPIINNAKVGEKHYLEMDRGVLEVRY 594
Sbjct 541  VVMNSGTFQEQTD FQKRMGISPTFAIVPIINNAKVGEKHYLEMDRGVLEVRY 594

```

Odile Lecompte -IGBMC ASMI

# Blast : taxonomy report

## Lineage Report

```

cellular organisms
├── Euryarchaeota [euryarchaeotes]
│   ├── Methanocaldococcus jannaschii ..... 1189 1 hit [euryarchaeotes] RecName: Full-DNA polymerase II small subunit; Short-Pol II
│   ├── Archaeoglobus fulgidus ..... 347 1 hit [euryarchaeotes] RecName: Full-DNA polymerase II small subunit; Short-Pol II
│   ├── Methanothermobacter thermautotrophicus str. Delta H ..... 324 1 hit [euryarchaeotes] RecName: Full-DNA polymerase II small subunit; Short-Pol II
│   ├── Pyrococcus furiosus ..... 294 1 hit [euryarchaeotes] RecName: Full-DNA polymerase II small subunit; Short-Pol II
│   ├── Pyrococcus horikoshii ..... 288 1 hit [euryarchaeotes] RecName: Full-DNA polymerase II small subunit; Short-Pol II
│   ├── Halobacterium salinarum ..... 283 1 hit [euryarchaeotes] RecName: Full-DNA polymerase II small subunit; Short-Pol II
│   ├── Halobacterium salinarum RI ..... 283 1 hit [euryarchaeotes] RecName: Full-DNA polymerase II small subunit; Short-Pol II
│   ├── Pyrococcus abyssi ..... 282 1 hit [euryarchaeotes] RecName: Full-DNA polymerase II small subunit; Short-Pol II
│   ├── Thermoplasma acidophilum ..... 281 1 hit [euryarchaeotes] RecName: Full-DNA polymerase II small subunit; Short-Pol II
│   ├── Arabidopsis thaliana (thale-cress) ..... 65 1 hit [eudicots] RecName: Full-DNA polymerase delta small subunit
│   ├── Oryza sativa Japonica Group (Japanese rice) ..... 62 1 hit [monocots] RecName: Full-DNA polymerase delta small subunit
│   ├── Xenopus laevis (common platanna) ..... 33 1 hit [frogs & toads] RecName: Full-Kinetochore protein Ndc-80; AltName: Full-Cel
│   └── Lactococcus lactis subsp. lactis ..... 32 1 hit [firmicutes] RecName: Full-2',3'-cyclic-nucleotide 5'-phosphodiesterase
  
```

## Organism Report

```

Methanocaldococcus jannaschii [euryarchaeotes] taxid 2100
pp125911.1|D223 Ntd2 RecName: Full-DNA polymerase II sma... 1189 0.0

Archaeoglobus fulgidus [euryarchaeotes] taxid 2234
pp1282484.1|D223 ARFV RecName: Full-DNA polymerase II sma... 347 9e-95

Methanothermobacter thermautotrophicus str. Delta H [euryarchaeotes] taxid 187420
pp1287455.1|D223 MntH RecName: Full-DNA polymerase II sma... 324 1e-87

Pyrococcus furiosus [euryarchaeotes] taxid 2261
pp1281815.1|D223 PFRF RecName: Full-DNA polymerase II sma... 294 1e-78

Pyrococcus horikoshii [euryarchaeotes] taxid 53953
pp1357863.1|D223 PYPB RecName: Full-DNA polymerase II sma... 288 9e-77

Halobacterium salinarum [euryarchaeotes] taxid 2242
pp1298967.1|D223 HALS RecName: Full-DNA polymerase II sma... 283 2e-75

Halobacterium salinarum RI [euryarchaeotes] taxid 478009
pp1298971.1|D223 HALR RecName: Full-DNA polymerase II sma... 283 2e-75

Pyrococcus abyssi [euryarchaeotes] taxid 29292
pp1297473.1|D223 PYAB RecName: Full-DNA polymerase II sma... 282 5e-75

Thermoplasma acidophilum [euryarchaeotes] taxid 2303
pp1298185.1|D223 THAC RecName: Full-DNA polymerase II sma... 281 1e-74

Arabidopsis thaliana (thale-cress, ...) [eudicots] taxid 3702
pp1248520.1|D223 ATTH RecName: Full-DNA polymerase delta... 65 7e-10

Oryza sativa Japonica Group (Japanese rice) [monocots] taxid 39947
pp129245.1|D223 ORS2 RecName: Full-DNA polymerase delta... 62 1e-08

Xenopus laevis (common platanna, ...) [frogs & toads] taxid 8355
pp1265971.1|D223 XENLA RecName: Full-Kinetochore protein ... 33 5.4

Lactococcus lactis subsp. lactis [firmicutes] taxid 1360
pp129222.1|D223 LACTA RecName: Full-2',3'-cyclic-nucleotide... 32 7.9
  
```

## Taxonomy Report

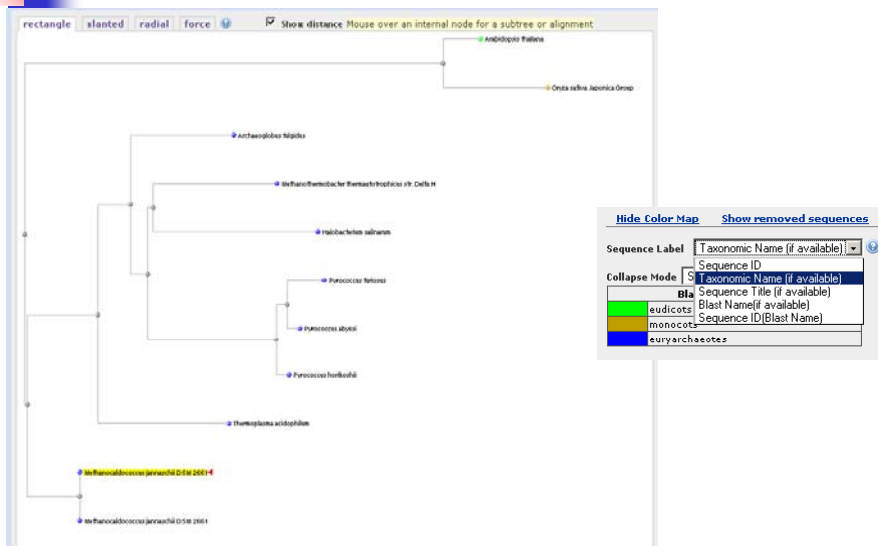
```

cellular organisms ..... 13 hits 13 orgs [root]
├── Euryarchaeota ..... 9 hits 9 orgs [Archaea]
│   ├── Methanocaldococcus jannaschii ..... 1 hits 1 orgs [Methanococci; Methano]
│   ├── Archaeoglobus fulgidus ..... 1 hits 1 orgs [Archaeoglobi; Archae]
│   ├── Methanothermobacter thermautotrophicus str. Delta H ..... 1 hits 1 orgs [Methanothermobact; Meth]
│   ├── Pyrococcus furiosus ..... 1 hits 1 orgs [Pyrococci; Pyrococ]
│   ├── Pyrococcus horikoshii ..... 1 hits 1 orgs [Pyrococci]
│   ├── Pyrococcus abyssi ..... 1 hits 1 orgs [Pyrococci]
│   ├── Halobacterium salinarum RI ..... 1 hits 1 orgs [Halobacteri; Haloba]
│   ├── Halobacterium salinarum ..... 1 hits 1 orgs [Halobacteri; Haloba]
│   ├── Thermoplasma acidophilum ..... 1 hits 1 orgs [Thermoplasmata; Ther]
│   ├── Kinetoplastida ..... 3 hits 3 orgs [Kinetoplastida]
│   ├── Arabidopsis thaliana ..... 1 hits 1 orgs [Eudicotyledons; core]
│   ├── Oryza sativa Japonica Group ..... 1 hits 1 orgs [Liliopsida; communi]
│   ├── Xenopus laevis ..... 1 hits 1 orgs [Fungi; Metazoa group]
│   └── Lactococcus lactis subsp. lactis ..... 1 hits 1 orgs [Bacteria; Firmicutes]
  
```

Odile Lecomte -IGBMC

ASMI

# Blast : arbre des distances



Odile Lecomte -IGBMC

ASMI

# Sorties BLAST

Formatting options Reformat

Show  as   Advanced View  Use old BLAST report format [Reset form to defaults](#)

Alignment View

Display  Graphical Overview  Linkout  Sequence Retrieval  NCB-gi

Masking Character:  Color:


Limit results Descriptions:  Graphical overview:  Alignments:

Organism

Entrez query:

Expect Min:  Expect Max:

Format for  PSI-BLAST with inclusion threshold:



## Pairwise with dots for identities

```

>|_sp|028484.1|DP23_ARCFU RecName: Full=DNA polymerase II small subunit; Short=Pol II
Length=488

Score = 347 bits (80%), Expect = 9e-05, Method: Compositional matrix adjust.
Identities = 180/387 (47%), Positives = 248/387 (65%), Gaps = 11/387 (2%)

Query 184 DVSQKSTCTGTIEDFVNYFFDRFERLKVFIERKAQRKGYPLDKIQEKGQKDLFVVGIVS 243
Sbjct 98 .IT...S.Q.NV...LH..HS.L.K.SKI.RSRVHTT...LHGQ.VR.--NVS...M.N 153

Query 244 DVDSTRNGNLIVRIEDTEATLLPKEKIEAGKIPDDILLDEVIGAIGTVSKSGSIIYV 303
Sbjct 154 E.-YE.GDKCYI.L...TGTI.CVATGQNA.VBR--EL.G.....VT.LLK--...L.A 207

Query 304 DEIIRPALPPEPKRIDEEIYMAFLSDIHVGSKEFLHKEFEKFRFLNGVDNLEEKVY 363
Sbjct 208 NR.VF.DV.DHNGEKKQDF.IV....T.F.....E..W.M.V.W.K.E.GGRKSONLA 267

Query 364 SRLKYICIAQDLVDGQGVYFGQEDLYEVDIIQYREIAMYLDQIPEHISIIISFGMHDA 423
Sbjct 268 ERV...V....I...I.....D..AIS..YG..EFA.SH..E..KE.K..V..... 327

Query 424 VRPAEPQPKLPEKITLPMRDNIYVGNPCTLMIHGFDLLVHGRSFDLLVQIRAAASYE 483
Sbjct 328 ..Q.....RFEQE.RS..PK--VEH....RYVD.E.VKV.I.....I..LISK.PRL..D 386

Query 484 NPVTIMKELIKRRLLCPTVGRCPAIEPEHDTLVDRDIDILHTGHINGYGIYRGVYM 543
Sbjct 387 E.QKV.E..L...H.S.I....T.L...DE.....EDVP....C....TV.T.F...F. 446

Query 544 VNSGTFQEQTDFOKRMGISPTPAIVFI 570
Sbjct 447 ...S.W.A..E...KVMLN.H.GN.RV 473

```

Odile Lecompte -IGBMCASMI

## Query-anchored

<pre> Query 1      HEIIMKFLDEALLSPTVYKLNDFDEKRLQIKRIFPKYNNAFILLDEKFLDIFLQ  60 P81412  3      .FVKS.L.KAMY.IT.SA.YL.REYY.RGEFSIVELVK-.ARSRESY.IT.ALATEFLVK  60 Query 61     KDLDRIINEYVDPDFPFFVYTGEEKEKPKVQSGEIKGTEKIEKIEKIEFVQGEKEQFI  120 P81412  61     .G.EP.LPVETKGG.V-S...SQ...OSY.ESGFT.E.ISQE.K-----G.S...  108 Q9CEE2  71     .TLR...QKQRR.IED.F.Q.RQ.LK.T.----RLKQR.EIL  109 Query 121    KKSDEDVEEKLQGLISKEEKEDFAERAKRYEHITKIKESVNSRIKWIAIDDAVIEIY  180 P81412  109    STGS.PL..E.NSIGIE.IGANE..EKEYSK..DL.-.PMMFDVNVVE.RE.Y.V.FDVL  202            \ \            \ \            \ \ Q6GQ71  350     ELVSNQNDNGGEAIV KYGYVNYAPEEIEVE          RMVK            ...AKV.KNK.K.Q...NQKQEIL...QEKQVING.VAOVLOE.QQT.SKK  405            \ \            \ \            \ \ Q9CEE2  110    DRK.DTLTR.EEN.D...R.T.TLSK.EEQLA..EEK.RLELE..SNLST.DARE.ILS  172            \ \            \ \            \ \            NLV </pre>	<pre> Query 1      HEIIMKFLDEALLSPTVYKLNDFDEKRLQIKRIFPKYNNAFILLDEKFLDIFLQ  60 P81412  3      EFVKSLLKAMYLITPSAYILLREYKGEFSIVELVK-FARSRESYITDALATEFLVK  60 Query 61     KDLDRIINEYVDPDFPFFVYTGEEKEKPKVQSGEIKGTEKIEKIEKIEFVQGEKEQFI  120 P81412  61     KGLEPILPVETKGGFY--STGSQREKQSTEEFQTKIEIQEK-----EGESFI  108 Q9CEE2  71     YTLREKQKQSGEIEDEPKQERQELKETER---RLKQREIL  109 Query 121    KKSDEDVEEKLQGLISKEEKEDFAERAKRYEHITKIKESVNSRIKWIAIDDAVIEIY  180 P81412  109    STGSFLEELLSIGIEIGANE.FDEKEYSKVEDLT-IPMMFDVNVVEKEDVYVDFVL  202            \ \            \ \            \ \ Q6GQ71  350     ELVSNQNDNGGEAIV KYGYVNYAPEEIEVE          RMVK            EKLAQVINKNGQKQEDFNQKQKQILKHTQEKQVINGRVAQVLOEQQTISK  405            \ \            \ \            \ \            D            \ \            \ \            \ \ Q9CEE2  110    DPKDQTLTQGEKLD-SKKEKSTDTLSKREKOLAHIEKQGLERISNLSTDAREIILS  172            \ \            \ \            \ \            NLV </pre>
---	--

Odile Lecompte -IGBMCASMI

## Flat Query-anchored

<pre> Query 1      HEIIMKFLDEALLSPTVYKLNDFDEKRLQIKRIFPKYNNAFILLDEKFLDIFLQ  60 P81412  3      .FVKS.L.KAMY.IT.SA.YL.REYY.RGEFSIVELVK-.ARSRESY.IT.ALATEFLVK  60 Query 61     KDLDRIINEYVDPDFPFFVYTGEEKEKPKVQSGEIKGTEKIEKIEKIEFVQGEKEQFI  120 P81412  61     .G.EP.LPVETKGG.V-S...SQ...OSY.ESGFT.E.ISQE.K-----G.S...  108 Q9CEE2  71     .TLR...QKQRR.IED.F.Q.RQ.LK.T.----RLKQR.EIL  109 Query 121    KKSDEDVEEKLQGLISKEE---KED-----FD-----  145 P81412  109    STGS.PL..E.NSI-QIE.I---GAMELVSNQNDNGGEAIV..KYGYVNYAPEEIEVE  164 Q6GQ71  350     ...AKVD.KNKK--..Q-----N-----  367 Q9CEE2  110    DRK.DTLTR.EEN.-D...NLVR.T.-----TL-----  137 Query 146    AERAKRYE---HITKIKESVNSRIKWIAIDDAVIEI---YE---DSDVSGKSTCTTPI  195 Q28484  98     .IT...S.Q.NV  109 Q27456  87     ..T.NR.VYS.E.  99 P81412  165    EKEYSK.----DL.-.PMMFDVNVVE.RE.Y.V.FDVSNQNLK---PK.KNGSGRE.E.  217 Q9HMR7  114    .G.MT.A..G..EV  127 Q9V2F3  225     V  225 Q9HLK3  101     NAS.  104 Q6GQ71  368    QKQKQILKVC.S..QEKQVING.VAOVLOE.QQT.SK---H.OLLETTEA..NK.QEY.  423 Q9CEE2  130    SK.EQOLA---.EEK.RLELE..SNLST.DARE.IE.L---S---TR-----LT  179 </pre>	<pre> Query 1      HEIIMKFLDEALLSPTVYKLNDFDEKRLQIKRIFPKYNNAFILLDEKFLDIFLQ  60 P81412  3      EFVKSLLKAMYLITPSAYILLREYKGEFSIVELVK-FARSRESYITDALATEFLVK  60 Query 61     KDLDRIINEYVDPDFPFFVYTGEEKEKPKVQSGEIKGTEKIEKIEKIEFVQGEKEQFI  120 P81412  61     KGLEPILPVETKGGFY--STGSQREKQSTEEFQTKIEIQEK-----EGESFI  108 Q9CEE2  71     YTLREKQKQSGEIEDEPKQERQELKETER---RLKQREIL  109 Query 121    KKSDEDVEEKLQGLISKEE---KED-----FD-----  145 P81412  109    STGSFLEELLSI-QIEI---GAMELVSNQNDNGGEAIVKYGYVNYAPEEIEVE  164 Q6GQ71  350     EKLAQVINKNK--KED-----FN-----  367 Q9CEE2  110    DPKDQTLTQGEKLD-SKKEKLVGKTD-----TL-----  137 Query 146    AERAKRYE---HITKIKESVNSRIKWIAIDDAVIEI---YE---DSDVSGKSTCTTPI  195 Q28484  98     DITKUSCCQMV  109 Q27456  87     SITSNRSVTQEI  99 P81412  165    EKEYSKYE---DLT-IPMMFDVNVVEKEDVYVDFVSNQNLK---PFKVNQNGREGEI  217 Q9HMR7  114    DQNTGASTGTGEY  127 Q9V2F3  225     V  225 Q9HLK3  101     BASI  104 Q6GQ71  368    QKQKQILKVC.SKHTQEKQVINGRVAQVLOEQQTISK---SKQILETTEACKQKQEYI  423 Q9CEE2  138    SKREKOLA---HIEKQKQKLELERSNLSTDAREIIL---SE---TRG-----GLT  179 </pre>
--	---

Odile Lecompte -IGBMCASMI