# BLAST

Herramienta básica de búsqueda de alineación local BLAST encuentra regiones de similitud entre secuencias biológicas. El programa compara secuencias de nucleótidos o proteínas con bases de datos de secuencias y calcula la significancia estadística.

**BLAST**®

Home Recent Results Saved Strategies Help

Check out the ClusteredNR database on BLAST+

Learn more

Give us feedback

### **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

ClusteredNR database on BLAST+

The ClusteredNR database is now available for BLAST+

Thu, 24 Aug 2023

More BLAST news...

#### **Web BLAST**



#### blastx

translated nucleotide ▶ protein

tblastn
protein > translated nucleotide



BLAST® » blastn suite Recent Results Saved Strategies Help Check out the ClusteredNR database on BLAST+ Learn more Give us feedback Standard Nucleotide BLAST blastp tblastn tblastx blastn blastx BLASTN programs search nucleotide databases using a nucleotide query, more... Reset page Bookmark **Enter Query Sequence** Query subrange ? Enter accession number(s), gi(s), or FASTA sequence(s) ? clear From Or, upload file Examinar... Ningún archivo seleccionado. Job Title Enter a descriptive title for your BLAST search ? Align two or more sequences ? **Choose Search Set Database** ⑤ Standard databases (nr etc.): ○ rRNA/ITS databases ○ Genomic + transcript databases ○ Betacoronavirus Try experimental taxonomic nt databases Download New Experimental databases For more info see What are taxonomic nt databases?

### Identificadores

### AT5G05150

### Formato FASTA:

Es un formato de fichero informático basado en texto, utilizado para representar secuencias de ácidos nucleicos, de péptido, y en el que los pares de bases o los aminoácidos se representan usando códigos de una única letra

#### >AtATG18b.I.AT4G30510.1

MANLSSLPSPIYCVSFNQDNSGFAISWKDSFKIFDSTTGRLCYERAAGAFVIVEMLYSSD LLAIVGAGEQASLSPRRLCLFKTTTGLPLRELNFLTSILAVRMNKKRLVVVLLEKTFVYD LNTLVMLDTIDTVPNPKGLSAFSPSLEGCYLAVPASTTKGSVLVYNVMDLQSHSEIDAHR SPLAAIALSSNGMYIATGSEQGTLIRVHLVSEATKSYSFRRGTYPSTIYSLSFGPSTQLP DILIATSSSGSIHAFSLSLAINQRSKRSTSFLGSVLPDSVSDALDPAHHHVLQNAVSSGI RSYAVVRKIDKLEGTSSPSHFTSLRATVSVITYNGYFQEYTLSINNKNESLWTLEREFNL FSITTG

#### >AT4G30510

ggtgagttcttcttatttatttgaaaatcgttgaagatagcgaattcgattctcttagta gattttcgaggaaaagttcttctcagataacgtaactcagaagctgcccttttattgttg tttgatggctgttcgattattgcagtggttttgcgataagttggaaagattccttcaaaa tattcgattccactacaggcagactctgctatgaacgaggcaagttttatctgttcgttt tggttctgcagtttctgtgttttaatcaaaagatccaacttttgttgtctatatccttta aattgaatccgtgggcgtttataattgatggatttgatagaaaacttggttagatgaatc gaaatcttaagtatgatctgatattgaagacaacttcctgtaatgcagttgcagacatta aacacttattaatgttgtaaattttagattggttagtgtctgaatattgatgatacctgg taaaaaaaactaatgcaattggcataagatcatccttattccaagcttcaatttttactt cctttgttattgttgagatgttgtatagctcagatctacttgcaattgttggagctggtg aacgggtttacctcttagggaattgaattttttaacttctatactcgctgttcgtatgaa gtgtgaattcatggttctgtgattgcttcttatcagacttgttgttgtgctgctagagaa aacattcgtttacgacttgaatactcttgtcatgcttgatactatcgacaccgtgccaaa tccaaaaggtatctcattaagttccctttgtcagtttttaccgtattagttcattgtaaa cagacatgtaaatgtatatcttcgttcaggtctatctgcattctctccatctcttgaagg ctgctacttagctgttccagctagtacgacgaaaggatctgttttggtgtataacgtcat ggatttgcagtctcatagtgaggtttactagatgcttcacctgatttttggtttaaattt aaactcgattacagattgatgcacatcgttctccattggctgcgattgcactgtcttcaa atggaatgtacattgctacaggatctgagcaaggaaccttgatccgagtccatcttgttt cagaagcaactaaggtgaatagtttgctcttttaatacatcaccattttgtttttctaca gattcttaatcgtttctttcctctgtttgcagtcatatagtttcaggagaggaacatacc catctactatatattcactgtcttttggaccatcgacacagcttccagatatattaattg caactagttcttcaggctccattcatgccttctctttaagtttggccataaaccaaaggt taggattaagaatcacactcttacttcccttatacatactgtgattcctaacatttgtaa gagtttttcagaagcaaaaggtctactagctttcttggatcggttcttcccgacagtgtg agtgatgcactagatccagctcatcatcatgtgcttcaaaacgcagtttcttccgggatt agaaggtgaaagttaattcttttagacgctgctaatgacctatcaggttatattccccaa gaggtttattgtctgtctcttgcttgtgtgtacagttatgccgtggttaggaagatagac aaacttgaaggaacatcatctccatctcattttacatctcttaggtaagattttaggata tcttcattttgaaaccatggttacatgacaagttttgataatcttttacctttcacgttc ttttcagagcaacggtatcggtgatcacttataacggatatttccaggagtatacgctaa gtatcaataacaagaatgagtctctatggacactggaacgtgaattcaatctcttctcca taaccaccgggtgatgtatacaaccaaatatcaaatatatgaaagtactgggcaactttt ttatatataacaatttttataccttagattcattaagaatatatttatagtctcatattc tttggttattcgcaaatataattagctcttccgtgctattactgtatcttaatctt

### **BLASTn**

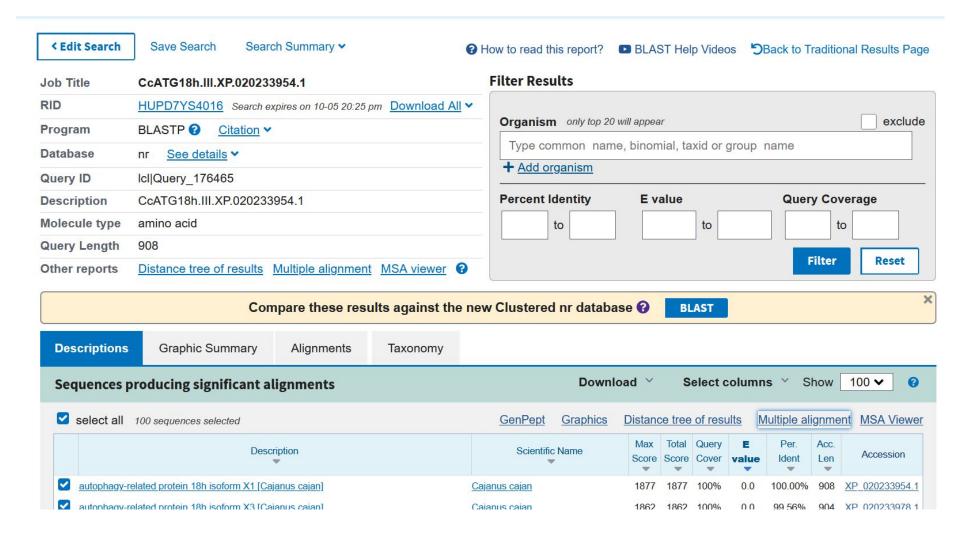
Con nucleotidos

BLASTp

Con Proteínas

# BLAST en otras plataformas

https://www.arabidopsis.org/Blast/index.jsp



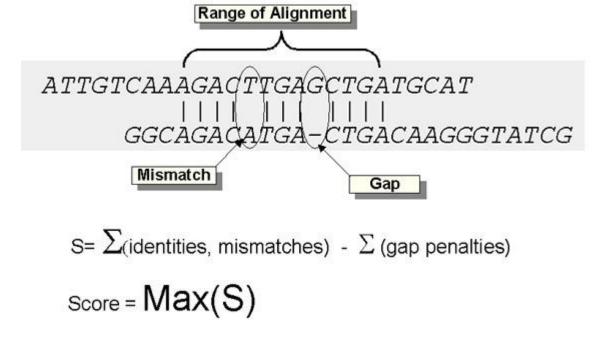
### Descripción

### En los resultados, las columnas en "Descriptions" son:

- 1. Description: el título de la referencia candidata en la base de datos
- 2. Max score: la puntuación de alineación de la mejor coincidencia entre el Query y la Referencia
- 3. Total score: la suma de las puntaciones de todos las coincidencias en la base de datos
- 4. Query cover: el porcentaje de la secuencia del Query que cubre la alineación con la Referencia
- 5. E-value: el valor esperado que se calcula a partir de la puntuación máxima (es decir, se esperaría ver esta alineación por razones aleatorias con una frecuencia del valor indicado)
- 6. Ident: % identidad en la alineación
- 7. Accessión: índice en la base de datos de la referencia candidata

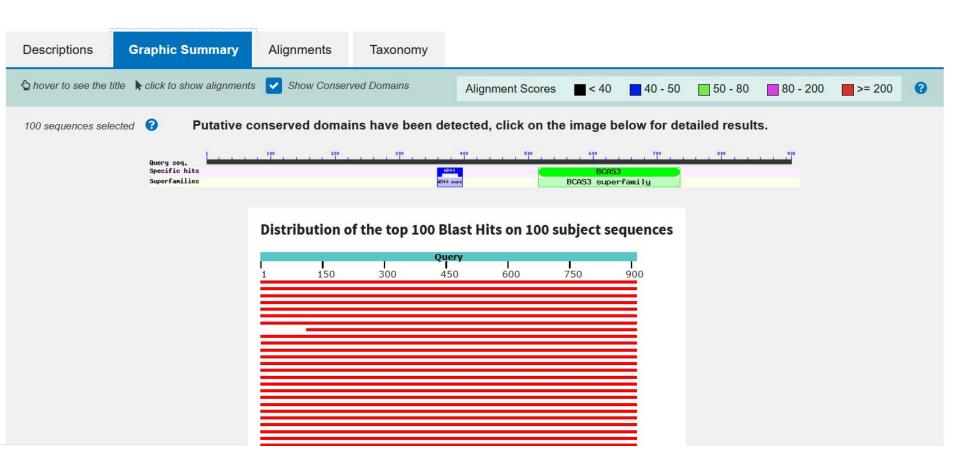
https://ecogenucr.wordpress.com/bioinformatica-4/

Score



### E-value

1. E-value: el valor esperado que se calcula a partir de la puntuación máxima (es decir, se esperaría ver razones aleatorias con una frecuencia del valor indicado)



### hypothetical protein Ahy\_B09g096680 isoform E [Arachis hypogaea]

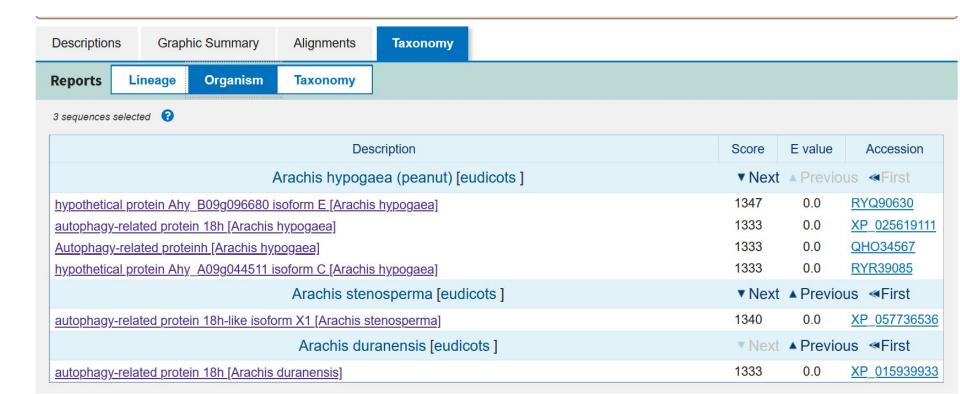
Sequence ID: RYQ90630.1 Length: 909 Number of Matches: 1

Range 1: 3 to 901 GenPept Graphics

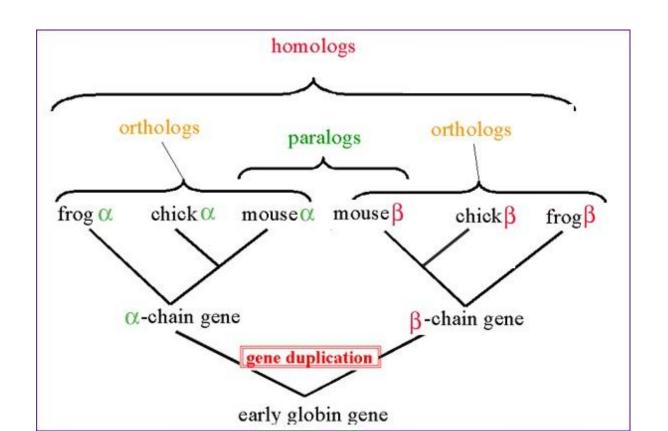
▼ Next Match ▲ Previous Match

Score		Expect	Method		Identities	Positives	Gaps
1347	bits(348	35) 0.0	Composition	nal matrix ad	just. 696/907(7	77%) 776/907(85%	6) 12/907(1%)
Query	6			FISSCIKTASSO FISSCIKTASSO		GDGHDRKDQVAFCVLI	WAC 61
Sbjct	3					GDEDNLKDQILI	
Query	62					DDPVSFLQMQPVPAKS	
Sbjct	59					YPVSFLQMKPVPAKL	
Query	122					SATAVRFYSLRSHTYVI SA ++FYSLRSHTYVI	
Sbjct	119					SARTLQFYSLRSHTYVI	
Query	182					VPQLGGQGMVGVNIG	
Sbjct	179					PVPQLGGQGMVGVNIG	
Query	242					SLVARYAMESSKHLA	
Sbjct	239					G+LVARYAME+SKHL / GNLVARYAMEASKHLT/	

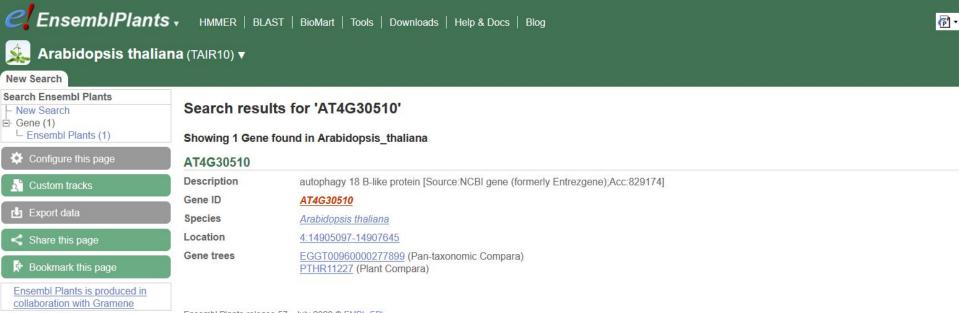
### Homologos



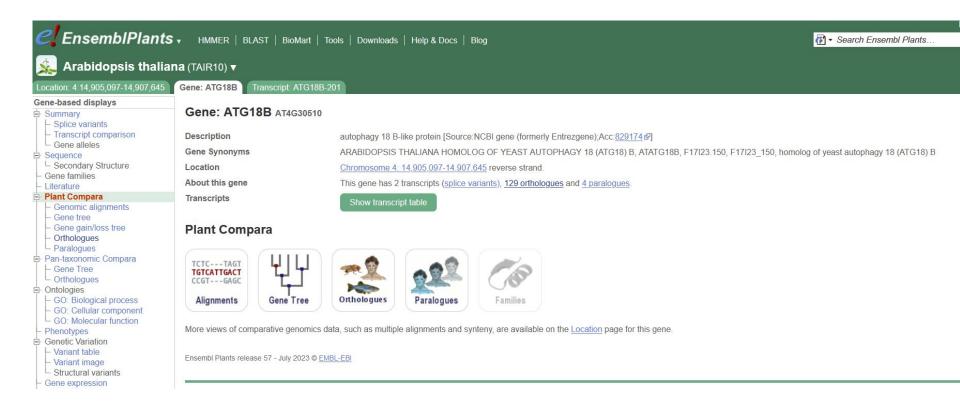
# **HOMÓLOGOS**



# HOMÓLOGOS - ENSEMBL



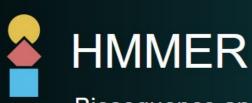
Ensembl Plants release 57 - July 2023 © EMBL-EBI



https://plants.ensembl.org/Arabidopsis\_thaliana/Gene/Compara?db=core;g=AT4G30510;r=4:14905097-14907645;t

### Selected orthologues Hide

Show All v entries		Show/hide columns				Filter	III X
Species	Type	Orthologue	Target %id	Query %id	GOC Score	WGA Coverage	High Confidence
Actinidia chinensis	1-to-1 View Gene Tree	CEY00_Acc10097 Compare Regions (LG9:3,995,957-4,009,056:1) View Sequence Alignments	71.27 %	71.86 %	n/a	94.07	Yes
Aegilops tauschii	1-to-1 View Gene Tree	<u>AET6Gv20872500</u> <u>Compare Regions</u> (6D:461,047,293-461,051,007:-1) <u>View Sequence Alignments</u>	47.22 %	53.28 %	n/a	75.77	Yes
Amborella trichopoda	1-to-1 View Gene Tree	AMTR_s00019p00100510  Compare Regions (AmTr_v1.0_scaffold00019:1,220,882-1,228,704:-1)  View Sequence Alignments	60.06 %	58.74 %	n/a	86.55	Yes
Ananas comosus	1-to-1 View Gene Tree	Aco011590.1.path1  Compare Regions (contig7:316,512-320,202:-1)  View Sequence Alignments	63.54 %	62.84 %	n/a	94.04	Yes
Arabidopsis halleri	1-to-1 View Gene Tree	g04175 Compare Regions (FJVB01000007.1:77,856-80,670:1) View Sequence Alignments	91.35 %	98.09 %	100	98.31	Yes
Arabidopsis lyrata	1-to-1 View Gene Tree	Scaffold 701168.1  Compare Portions (7:4.725.284.4.726.926.1)	98.72 %	84.15 %	100	100.00	Yes



Biosequence analysis using profile hidden Markov Models

# HOMÓLOGOS - HMMER

Un modelo oculto de Márkov o HMM (por sus siglas del inglés, Hidden Markov Model)

https://www.ebi.ac.uk/Tools/hmmer/search/phmmer

#### PHMMER Results



