

Find a Gene Project

Part 1

Student: Yushi Li

Email: yul012@ucsd.edu

PID: A15639705

Questions:

[Q1] Tell me the name of a protein you are interested in. Include the species and the accession number. This can be a human protein or a protein from any other species as long as it's function is known.

If you do not have a favorite protein, select human RBP4 or KIF11. Do not use beta globin as this is in the worked example report that I provide you with online.

Name: CLOCK

Accession: AAB83969

Species: Homo sapiens

Functions: Involved in circadian regulation of gene expression through a variety of cellular processes such as protein acetylation and DNA damage checkpoint signaling.

[Q2] Perform a BLAST search against a DNA database, such as a database consisting of genomic DNA or ESTs. The BLAST server can be at NCBI or elsewhere. Include details of the BLAST method used, database searched and any limits applied (e.g. Organism).

Method: tBLASTn against frog TSA.

Database: Transcriptome Shotgun Assembly (TSA)

Organism: Frogs (taxid:8342)

Also include the output of that BLAST search in your document. If appropriate, change the font to `Courier size 10` so that the results are displayed neatly. You can also screen capture a BLAST output (e.g. alt print screen on a PC or on a MAC press ⌘ -shift-4. The pointer becomes a bulls eye. Select the area you wish to capture and release. The image is saved as a file called `Screen Shot [].png` in your Desktop directory). It is **not** necessary to print out all of the blast results if there are many pages.

BLAST® » tblastn Home Recent Results Saved Strategies Help

blastn blastp blastx **tblastn** tblastx

Translated BLAST: tblastn

TBLASTN search translated nucleotide databases using a protein query. more...

[Reset page](#) [Bookmark](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) Query subrange [?](#)

AAB83969 From

To

Or, upload file [Choose File](#) No file chosen [?](#)

Job Title [?](#)

AAB83969:CLOCK [Homo sapiens]

Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

Choose Search Set

Database [?](#)

Transcriptome Shotgun Assembly (TSA)

Limit by [?](#)

☒ Organism ☐ BioProjectID ☐ TSA Project

frogs (taxid:8342) ☐ exclude [Add Organism](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Limit to [?](#)

Optional ☐ Sequences from type material

BLAST Search database tsa using Tblastn (search translated nucleotide databases using a protein query)

☐ Show results in a new window

On the BLAST results, clearly indicate a match that represents a protein sequence, encoded from some DNA sequence, that is homologous to your query protein. I need to be able to inspect the pairwise alignment you have selected, including the E value and score. It should be labeled a "genomic clone" or "mRNA sequence", etc. - but include nonfunctional annotation.

Chosen match: Accession GIVH01185592.1, a 2514bp segment of a piece of transcribed RNA from *Xenopus tropicalis* obtained through transcriptome shotgun assembly (TSA).

? Your search is limited to records that include: frogs (taxid:8342)

Job Title **AAB83969:CLOCK [Homo sapiens]**

RID [ZTMRDF1A01R](#) Search expires on 02-05 14:17 pm [Download All](#) [v](#)

Program TBLASTN [?](#) [Citation](#) [v](#)

Database tsa (45 databases) [See details](#) [v](#)

Query ID [AAB83969.1](#)

Description CLOCK [Homo sapiens]

Molecule type amino acid

Query Length 846

Other reports [?](#)

Filter Results

Percent Identity to E value to Query Coverage to

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

Descriptions Graphic Summary Alignments

Sequences producing significant alignments Download [v](#) [New](#) Manage columns [v](#) Show 100 [v](#) [?](#)

☒ select all 100 sequences selected [GenBank](#) [Graphics](#)

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	TSA: Xenopus tropicalis TRINITY_DN3887_c0_g1_i2.transcribed RNA sequence	1151	1151	100%	0.0	68.31%	5403	GIVH01185592.1
<input checked="" type="checkbox"/>	TSA: Xenopus tropicalis TRINITY_DN3887_c0_g1_i1.transcribed RNA sequence	1151	1151	100%	0.0	68.31%	5325	GIVH01185591.1
<input checked="" type="checkbox"/>	TSA: Xenopus tropicalis TRINITY_DN3887_c0_g1_i3.transcribed RNA sequence	1151	1151	100%	0.0	68.31%	5365	GIVH01185590.1
<input checked="" type="checkbox"/>	TSA: Rana catesbeiana CCH-0011-C_S5565634.transcribed RNA sequence	1143	1143	99%	0.0	68.20%	6294	GFBS01317984.1
<input checked="" type="checkbox"/>	TSA: Rana catesbeiana CCH-BS01-R16271920.transcribed RNA sequence	1143	1143	99%	0.0	68.20%	3716	GFBS01009157.1
<input checked="" type="checkbox"/>	TSA: Odorrana tormota CL18005.Contig2_All.transcribed RNA sequence	1139	1139	99%	0.0	68.00%	4000	GGLB01058044.1

See below for alignment:

Descriptions

Graphic Summary

Alignments

Alignment view

Pairwise

?

Restore defaults

100 sequences selected

Download

GenBank

Graphics

TSA: *Xenopus tropicalis* TRINITY_DN3887_c0_g1_i2, transcribed RNA sequence

Sequence ID: [GIVH01185592.1](#) Length: 5403 Number of Matches: 1

Range 1: 309 to 2822 [GenBank](#) [Graphics](#)

▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
1151 bits(2978)	0.0	Compositional matrix adjust.	662/852(78%)	725/852(85%)	20/852(2%)	+3
Query 1	MLFTVSCSKMSSIVDRDDSSIFDGLVEEDDKDKAKRVS RNKSEKKRRDQFNVLIKELGSM	60				
Sbjct 309	+L ++S KMSS DRDD SIFDGLVEEDDKDKAKRVS RNKSEK+RRDQFN+LIKELGSM	488				
Query 61	LPGNARKMDKSTVLQKSIDFLRKHKETIAQSDASEIRQDWKPTFLSNEEFTQLMLEALDG	120				
Sbjct 489	LPGNAR+MDKSTVLQKSIDFLRKHKETIAQSDASEIRQDWKPTFLSNEEFTQLMLEALDG	668				
Query 121	FFLAIMTDGSIIVVSESVTSLLLEHLPDLVDQSIQNFPIPEGEHSEVYKILSTHLLSEDSL	180				
Sbjct 669	FFLA+MTDG+IIVVSESVTSLLLEHLPDLVDQSIQNF+PEGEHSEVYKILST +LES SL	848				
Query 181	TPEYLKSNQLEFCCHMLRGTIDPKEPSTYEVVKFIGNFKSLNSVSSAHNGFEGTIQRT	240				
Sbjct 849	+EYLK+KN+LEFCCHMLRGT DPKEPSTYE+VKFIGNFKSLN+V SS HNGF+G +QR+	1028				
Query 241	HRPSYEDRVCFVATVRLATPQFIKEMCTVEEPNEEFTSRHSLEWKFLFLDHRAPPIIGYL	300				
Sbjct 1029	LRPPYEERVCVATVRLATPQFIKEMCTVEE EFTSRHSLEWKFLFLDHRAPPIIGYL	1208				
Query 301	PFEVLGTSGVDYYHVDDLLENLAKCHEHLMQYGGKSCYYRFLTKGQQIWLQTHYYITYH	360				
Sbjct 1209	PFEVLGTSGVDYYHVDDLLENLAKCHEHLMQYGGKSCYYRFLTKGQQIWLQT YYITYH	1388				
Query 181	TPEYLKSNQLEFCCHMLRGTIDPKEPSTYEVVKFIGNFKSLNSVSSAHNGFEGTIQRT	240				
Sbjct 849	+EYLK+KN+LEFCCHMLRGT DPKEPSTYE+VKFIGNFKSLN+V SS HNGF+G +QR+	1028				
Query 241	HRPSYEDRVCFVATVRLATPQFIKEMCTVEEPNEEFTSRHSLEWKFLFLDHRAPPIIGYL	300				
Sbjct 1029	LRPPYEERVCVATVRLATPQFIKEMCTVEE EFTSRHSLEWKFLFLDHRAPPIIGYL	1208				
Query 301	PFEVLGTSGVDYYHVDDLLENLAKCHEHLMQYGGKSCYYRFLTKGQQIWLQTHYYITYH	360				
Sbjct 1209	PFEVLGTSGVDYYHVDDLLENLAKCHEHLMQYGGKSCYYRFLTKGQQIWLQTRYITYH	1388				
Query 361	QWNSRPEFIVCTHTVVSYA EVAERRELGI EESLPETAADKSQDSGSDNRINTVSLKEA	420				
Sbjct 1389	QWNSRPEFIVCTHTVVSYA EV AERRRE GIE+S P ADK+QDS SDN +NTVSLKEA	1568				
Query 421	LERFDHspptpsassrsrkssh-tavsdpsstptkiptdtstpprqlhpahekmvorrss	479				
Sbjct 1569	LERFD S TPS SS+SS KSS TAVSDPSSTPTKI T+TSTPPRQ + +K RRSS	1739				
Query 480	fssqssinsqsvgssLTOPVMSQATNLPiPQMSQ--FQFSAQLGAMQHLKDQLEQRTRMI	537				
Sbjct 1740	SSQS++SOSV +L+Q VM Q ++ + QGM+Q FQF+AQ GAM+HLKDQLEQRTRMI	1919				
Query 538	EANTHRQOEELRKIQEQQLMVHGQGLQMFLOQSNPGLNFGSVQLSSGNSNIQQLAPINM	597				
Sbjct 1920	EENIQRQOEELRKIQEQQLMVHGQGIQMFLOQPGPLNFGPVQVSSANSASTIQQLPQLTM	2099				
Query 598	QGQVVPTNIOISGMNTGHIGTTQHMtqqgtlqststasqqNVLSGHsagqtslpsqtqstl	657				
Sbjct 2100	QGQVV TNQ+Q+GMNTGH+G TQH++QQO LQ+TS Q QON+ GH5Q0TSL SQT TL	2276				
Query 658	tAPLYNTMVISQPAAGSMVQIPSSMPONSTQSAAVTTFTQDRQIRFSQGGQOLVTIKLVTA	717				
Sbjct 2277	T+PLYNTHVISQP +GSMVQ+PS++ Q S Q A+VTTF QDRQIRFSQ QQ+VTKLVT P	2453				
Query 718	VACGAVMVPSMLMGQVVTAYPTFatqqqqsgtllsvtqqqqqqssqeqqltsv---qqps	774				
Sbjct 2454	VACG VMVPSM MG VVTAYPTFATQ000 Q T Q +0 S+ QQP+	2633				
Query 775	qaqltqppqgflqtsRLHLGNPSTQLILSAAFPLqqstfpgshhqqhqsqqqqalsrhRT	834				
Sbjct 2634	QAQL Q PQQFLQTSRL+HGN STQLILS FP+ Q00LS HRT	2786				
Query 835	DSLPPDPKSVQPO 846					
Sbjct 2787	DS+ DPKSVQ Q DSMSPDKSVQ00 2822					

Alignment details:

>TSA: *Xenopus tropicalis* TRINITY_DN3887_c0_g1_i2, transcribed RNA sequence

Sequence ID: GIVH01185592.1 Length: 5403

Range 1: 309 to 2822

Score:1151 bits(2978), Expect:0.0,

Method: Compositional matrix adjust.,
Identities:662/852(78%), Positives:725/852(85%), Gaps:20/852(2%)

Query	1	MLFTVSCSKMSSIVDRDDSSIFDGLVEEDDKDKAKRVSRNKSEKKRRDQFNVLIKELGSM	60
		+L ++S KMSS DRDD SIFDGLVEEDDKDKAKRVSRNKSEK+RRDQFN+LIKELGSM	
Sbjct	309	LLLSISIHKMSSTADRDDGSIFDGLVEEDDKDKAKRVSRNKSEKKRRDQFNILIKELGSM	488
Query	61	LPGNARKMDKSTVLQKSIDFLRKHKETAQSDASEIRQDWKPTFLSNEEFTQLMLEALDG	120
		LPGNAR+MDKSTVLQKSIDFLRKHKEI+AQSDASEIRQDWKPTFLSNEEFTQLMLEALDG	
Sbjct	489	LPGNARRMDKSTVLQKSIDFLRKHKESAQSDASEIRQDWKPTFLSNEEFTQLMLEALDG	668
Query	121	FFLAIMTDGSIIVSVESVTSLEHLPSDLVDQSIFNFIPEGEHSEVYKILSTHLESDSL	180
		FFLA+MTDG+IIYSVESVTSLEHLPSDLVDQSIFNF+PEGEHSEVYKILST +LES SL	
Sbjct	669	FFLAVMTDGNIIYSVESVTSLEHLPSDLVDQSIFNFVPEGEHSEVYKILSTRMLESGL	848
Query	181	TPEYLKSKNQLEFCCHMLRGTIDPKEPSTYEYVKFIGNFKSLNSVSSSAHNGFEGTIQRT	240
		+ EYLK+KN+LEFCCHMLRGT DPKEPSTYE+VKFIGNFKSLN+V SS HNGF+G +QR+	
Sbjct	849	SSEYLKTKNELEFCCHMLRGTADPKEPSTYEFVKFIGNFKSLNNVPSSTHNGFDGALQRS	1028
Query	241	HRPSYEDRVCFVATVRLATPQFIKEMCTVEEPNEEFTSRHSLEWKFLFLDHRAPPIIGYL	300
		RP YE+RVCFVATVRLATPQFIKEMCTVEE EFTSRHSLEWKFLFLDHRAPPIIGYL	
Sbjct	1029	LRPPYEERVCFVATVRLATPQFIKEMCTVEESTEEFTSRHSLEWKFLFLDHRAPPIIGYL	1208
Query	301	PFEVLGTSGYDYYHVDDLENLAKCHEHLMQYGKGKSCYYRFLTKGQQWIWLQTHYYITYH	360
		PFEVLGTSGYDYYHVDDLENLAKCHEHLMQYGKGKSCYYRFLTKGQQWIWLQT YYITYH	
Sbjct	1209	PFEVLGTSGYDYYHVDDLENLAKCHEHLMQYGKGKSCYYRFLTKGQQWIWLQTRYITYH	1388
Query	361	QWNSRPEFIVCTHTVVSYA EVAERRELIGIEESLPETAADKSQDSGSDNRINTVSLKEA	420
		QWNSRPEFIVCTHTVVSYA EV AERRRE GIE+S P ADK+QDS SDN +NTVSLKEA	
Sbjct	1389	QWNSRPEFIVCTHTVVSYA EVGAERRRERGIEDSPPAITADKNQDSVSDNHMNTVSLKEA	1568
Query	421	LERFDhsptpsassrskssh-tavsDPSSTPTKIPTDTSTPPRQHLP AHEKMOVRRss	479
		LERFD S TPS SS+SS KSS TAVSDPSSTPTKI T+TSTPPRQ + +K RRSS	
Sbjct	1569	LERFDDSRTPSPSSKSSIKSSSHTAVSDPSSTPTKITETSTPPRQAITGLDK---RRSS	1739
Query	480	fssqsinsqsvgssLTQPVMSQATNLPIPQGMSQ--FQFSAQLGAMQHLKDQLEQRTRMI	537
		SSQS++SQSV +L+Q VM Q ++ + QGM+Q FQF+AQ GAM+HLKDQLEQRTR+I	
Sbjct	1740	VSSQSMSSQSVSQALSQSVMKQTASMQLOQGMQPMQFTAQFGAMKHLKDQLEQRTRII	1919
Query	538	EANIHRQQEELRKIQEQQLQMVHGQGLQMFLQQSNPGLNFGSVQLSSGNSSNIQQ LAPINM	597
		E NI RQQEELRKIQEQQL MVHGQG+QMFLQQ PGLNFG VQ+SS NS++IQQ L + M	
Sbjct	1920	EENIQRQQEELRKIQEQQLHVMHGQGIQMFLQQPGPGLNFGPVQVSSANSASIQQLPQLTM	2099
Query	598	QGQVPTNQIQSGMNTGHIGTTQHMIqqqtlqststqsgqNVLSGHsqqtslpsqtqstl	657
		QGQVV TNQ+Q+GMNTGH+G TQH++QQQ LQ+TS Q QQN+ GHSQQTSL SQT TL	
Sbjct	2100	QGQVVQTNQLQAGMNTGHVG-TQHIMQQQQLQTTSQQGQQNIHGHSQQTSLSSQTSCTL	2276
Query	658	tAPLYNTMVISQPAAGSMVQIPSSMPQNSTQSAAVTFTQDRQIRFSQGQQLVTKLVTAP	717
		T+PLYNTMVISQP +GSMVQ+PS++ Q S Q A+VTTF QDRQIRFSQ QQ+VTKLVT P	
Sbjct	2277	TSPLYNTMVISQPPSGSMVQMPNSNI-QQSNQGASVTTFQDRQIRFSQAQQIVTKLVTP	2453
Query	718	VACGAVMPSTMLMGQVVTAYPTFatqqqqsqtlsvtqqqqqqssqeqqltsv---qqps	774
		VACG VMVPSTM MG VVTAYPTFATQQQQ Q T Q +Q S+ QQP+	
Sbjct	2454	VACGTMVPSTMFMGPVVTAYPTFATQQQQQQQPQTLSITQHQPQQDQQSMPTVQQPA	2633

```

Query 775   qaqltqppqqflqtsRLLHGNPSTQLILSAAFPLqqstfpqshhqqhqsqqqqqlsrhRT 834
           QAQL Q PQQFLQTSRL+HGN STQLILS FP+ QQQLS HRT
Sbjct 2634  QAQLAQQPQQFLQTSRLVHGNQSTQLILS-PFPV-----QQNTFAPSHQQQLSHHRT 2786

Query 835   DSLPDPSKVQPQ 846
           DS+ DPSKVQ Q
Sbjct 2787  DSMSDPSKVQQQ 2822

```

In general, [Q2] is the most difficult for students because it requires you to have a “feel” for how to interpret BLAST results. You need to distinguish between a perfect match to your query (i.e. a sequence that is not “novel”), a near match (something that might be “novel”, depending on the results of [Q4]), and a non-homologous result.

If you are having trouble finding a novel gene try restricting your search to an organism that is poorly annotated.

[Q3] Gather information about this “novel” protein. At a minimum, show me the protein sequence of the “novel” protein as displayed in your BLAST results from [Q2] as FASTA format (you can copy and paste the aligned sequence subject lines from your BLAST result page if necessary) or translate your novel DNA sequence using a tool called EMBOSS Transeq at the EBI. Don’t forget to translate all six reading frames; the ORF (open reading frame) is likely to be the longest sequence without a stop codon. It may not start with a methionine if you don’t have the complete coding region. Make sure the sequence you provide includes a header/subject line and is in traditional FASTA format.

Chosen sequence:

```

>GIVH01185592.1:309-2822 TSA: Xenopus tropicalis TRINITY_DN3887_c0_g1_i2, transcribed
RNA sequence
LLLSISIHKMSSTADRDGSI F DGLVEEDDKDKAKRVSRNKSEKRRRDQFNILIKELGSMLPGNARRMDKSTVLQKSIDFLRKH
KEISAQSDASEIRQDWKPTFLSNEEFTQLMLEALDGFFLAVMTDGNIIYVSESVTSLLEHLPSDLVDQSIFNFVPEGEHSEVYK
ILSTRMLESGLSSEYLKTKNELEFCCHMLRGTADPKPESTYEFVKFIGNFKSLNNVPSSSTHNGFDGALQRLRPPYEERVCV
ATVRLATPQFIKEMCTVEESTEEFTSRHSLEWKFLFLDHRAPPIIGYLPFEVLGTSGYDYYHVDDLENLAKCHEHLMQYKGKGS
CYRFLTKGQQWIWLQTRYIITYHQWNSRPEFIVCTHTVVSYA EVGAERRRERGIEDSPPAITADKNQDSVSDNHMNTVSLKEA
LERFDDSRTPSPSSKSSIKSSSHTAVSDPSSTPTKITTTETSTPPRQAITGLDKRRSSVSSQSMSSQSVSQALSQSVMKQTASMQ
LQQGMAQPMFQFTAQFGAMKHLKDQLEQRTRIEENIQRQQEELRKIQEQLHMHVHGQGIQMFLQQPGPGLNFGPVQVSSANSAS
IQQLPQLTMQGQVVQTNQLQAGMNTGHVGTQHIMQQQQQLQTTSQQGQQNIHGGHSQQTSLSSQTSGLTSLSPYNTMVISQPPSG
SMVQMPNSNIQQSNQGASVTTFPQDRQIRFSQAQQIVTKLVTTTPVACGTMVPMSTMFMPVVTAYPTFATQQQQQQQQPQTLSIT
QHQPQQDQQQSMPTVQQPAQAQLAQPPQQFLQTSRLVHGNQSTQLILSPFPVQQNTFAPSHQQQLSHHRTDSMSDPSKVQQQ

```

Here, tell me the name of the novel protein, and the species from which it derives. It is very unlikely (but still definitely possible) that you will find a novel gene from an organism such as *S. cerevisiae*, human or mouse, because those genomes have already been thoroughly annotated. It is more likely that you will discover a new gene in a genome that is currently being sequenced, such as bacteria or plants or protozoa.

Name: Kermit's CLOCK

Species: *Xenopus tropicalis*

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia;
Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi;
Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amphibia; Batrachia;
Anura; Pipioidea; Pipidae; Xenopodinae; Xenopus; Silurana

[Q4] Prove that this gene, and its corresponding protein, are novel. For the purposes of this project, “novel” is defined as follows. Take the protein sequence (your answer to [Q3]), and use it as a query in a blastp search of the nr database at NCBI.

- If there is a match with 100% amino acid identity to a protein in the database, from the same species, then your protein is NOT novel (even if the match is to a protein with a name such as “unknown”). Someone has already found and annotated this sequence, and assigned it an accession number.
- If the top match reported has less than 100% identity, then it is likely that your protein is novel, and you have succeeded.
- If there is a match with 100% identity, but to a different species than the one you started with, then you have likely succeeded in finding a novel gene.
- If there are no database matches to the original query from [Q1], this indicates that you have partially succeeded: yes, you may have found a new gene, but no, it is not actually homologous to the original query. You should probably start over.

Details:

BLASTP against NR database yielded a top hit from *X. tropicalis* with less than 100% identity.

blastnblastblastxtblastntblastx

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

Reset pageBookmark

Enter Query Sequence

Enter accession number(s), g(i), or FASTA sequence(s) [?](#) [Clear](#)

Query subrange [?](#)

>EMBOSS_001_1
LLLSISIHQMSSTADRDGGSIFDGLVEEDKDKAKRVSRNKSEKRRDQFNIL
IKELGSM
LPGNARRMDKSTVLQKSIDFLRKHKEISQSDASEIRQDWKPTFLSNEEFTQ

From
To

Or, upload file

Choose FileNo file chosen [?](#)

Job Title

6 sequences (EMBOSS_001_1)
Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

Choose Search Set

Database

Non-redundant protein sequences (nr) [?](#)

Organism

Homo sapiens (taxid:9606) [?](#) ☐ exclude [Add organism](#)
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. [?](#)

Exclude

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

Program Selection

Algorithm

☐ Quick BLASTP (Accelerated protein-protein BLAST)
☒ blastp (protein-protein BLAST)
☐ PSI-BLAST (Position-Specific Iterated BLAST)
☐ PHI-BLAST (Pattern Hit Initiated BLAST)
☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)
Choose a BLAST algorithm [?](#)

BLAST

Search database nr using Blastp (protein-protein BLAST)

DescriptionsGraphic SummaryAlignmentsTaxonomy

Sequences producing significant alignmentsDownload [New](#) Select columns [Show](#) 100 [?](#)

☒ select all100 sequences selected

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

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	circadian locomotor output cycles protein kaput [Xenopus tropicalis]	Xenopus tropicalis	1715	1715	98%	0.0	99.88%	829	NP_001122127.1
<input checked="" type="checkbox"/>	hypothetical protein XENTR_v10000539 [Xenopus tropicalis]	Xenopus tropicalis	1714	1714	98%	0.0	99.76%	829	KAF8629596.1
<input checked="" type="checkbox"/>	circadian rhythmicity protein CLOCK [Xenopus laevis]	Xenopus laevis	1396	1396	86%	0.0	95.60%	778	AAF12827.1
<input checked="" type="checkbox"/>	clock circadian regulator L homeolog isoform X2 [Xenopus laevis]	Xenopus laevis	1427	1427	92%	0.0	94.35%	775	XP_041431955.1
<input checked="" type="checkbox"/>	circadian locomotor output cycles protein kaput [Xenopus laevis]	Xenopus laevis	1466	1466	98%	0.0	92.79%	827	XP_018098638.1
<input checked="" type="checkbox"/>	hypothetical protein XELAEV_18005292mg [Xenopus laevis]	Xenopus laevis	1466	1466	98%	0.0	91.68%	798	OCT99510.1
<input checked="" type="checkbox"/>	PREDICTED: circadian locomotor output cycles protein kaput [Nanorana parkeri]	Nanorana parkeri	1371	1371	98%	0.0	86.47%	826	XP_018431209.1
<input checked="" type="checkbox"/>	circadian locomotor output cycles protein kaput isoform X1 [Rana temporaria]	Rana temporaria	1368	1368	98%	0.0	85.30%	827	XP_040190735.1
<input checked="" type="checkbox"/>	circadian locomotor output cycles protein kaput isoform X2 [Rana temporaria]	Rana temporaria	1320	1320	95%	0.0	84.99%	807	XP_040190741.1
<input checked="" type="checkbox"/>	hypothetical protein GDO81_000351 [Engystomops pustulosus]	Engystomops pustulosus	1352	1352	98%	0.0	84.56%	820	KAG8591906.1
<input checked="" type="checkbox"/>	circadian locomotor output cycles protein kaput isoform X1 [Bufo gargarizans]	Bufo gargarizans	1347	1347	98%	0.0	84.56%	821	XP_044155187.1
<input checked="" type="checkbox"/>	circadian locomotor output cycles protein kaput isoform X3 [Bufo gargarizans]	Bufo gargarizans	1301	1301	95%	0.0	84.56%	801	XP_044155213.1
<input checked="" type="checkbox"/>	hypothetical protein GDO81_000351 [Engystomops pustulosus]	Engystomops pustulosus	1304	1304	95%	0.0	84.35%	800	KAG8591913.1
<input checked="" type="checkbox"/>	circadian locomotor output cycles protein kaput isoform X2 [Bufo gargarizans]	Bufo gargarizans	1338	1338	98%	0.0	84.32%	817	XP_044155206.1
<input checked="" type="checkbox"/>	circadian locomotor output cycles protein kaput isoform X1 [Bufo bufo]	Bufo bufo	1333	1333	98%	0.0	84.32%	821	XP_040274781.1
<input checked="" type="checkbox"/>	circadian locomotor output cycles protein kaput isoform X3 [Bufo bufo]	Bufo bufo	1286	1286	95%	0.0	84.31%	801	XP_040274783.1
<input checked="" type="checkbox"/>	hypothetical protein GDO81_000351 [Engystomops pustulosus]	Engystomops pustulosus	1338	1338	98%	0.0	84.20%	816	KAG8591908.1
<input checked="" type="checkbox"/>	clock [Homo sapiens]	Homo sapiens	1294	1294	99%	0.0	78.24%	845	AAF13733.1

circadian locomoter output cycles protein kaput [Xenopus tropicalis]

Sequence ID: [NP_001122127.1](#) Length: 829 Number of Matches: 1

[See 1 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)

Range 1: 1 to 829 [GenPept](#) [Graphics](#) [▼ Next Match](#) [▲ P](#)

Score	Expect	Method	Identities	Positives	Gaps
1715 bits(4442)	0.0	Compositional matrix adjust.	828/829(99%)	828/829(99%)	0/829(0%)
Query 10	MSSTADRDDGSIFDGLVEEDDKAKRVS RNKSEKRRRDQFNILIKELGSMLPGNARRMD				69
Sbjct 1	MSSTADRDDGSIFDGLVEEDDKAKRVS RNKSEKRRRDQFNILIKELGSMLPGNARRMD				60
Query 70	KSTVLQKSIDFLRKHKELISAQSDASEIRQDWKPTFLSNEEFTQLMLEALDGFFLAVMTDG				129
Sbjct 61	KSTVLQKSIDFLRKHKELISAQSDASEIRQDWKPTFLSNEEFTQLMLEALDGFFLAVMTDG				120
Query 130	NIIYVSESVTSLLEHLPSDLVDQSFNFVPEGEHSEVYKILSTRMLES6SLSSSEYLKTKN				189
Sbjct 121	NIIYVSESVTSLLEHLPSDLVDQSFNFVPEGEHSEVYKILSTRMLES6SLSSSEYLKTKN				180
Query 190	ELEFCCHMLRGTDADPKEPSTYEFVKFIGNFKSLNNVPSSTHNGFDGALQRSLRPPYEERV				249
Sbjct 181	ELEFCCHMLRGTDADPKEPSTYEFVKFIGNFKSLNNVPSSTHNGFDGALQRSLRPPYEERV				240
Query 250	CFVATVRLATPQFIKEMCTVEESTEEFTSRHSLEWKFLFLDHRAPPIIGVLPFEVLGTSG				309
Sbjct 241	CFVATVRLATPQFIKEMCTVEESTEEFTSRHSLEWKFLFLDHRAPPIIGVLPFEVLGTSG				300
Query 310	YDYVHVDLENLAKCHEHLMQYGGKSCYYRFLTKGQQWIWLQTRYITYYHQWNSRPEFI				369
Sbjct 301	YDYVHVDLENLAKCHEHLMQYGGKSCYYRFLTKGQQWIWLQTRYITYYHQWNSRPEFI				360
Query 370	VCHTHTVSYAEVGAERRRREGIEDSPPAITADKNQDSVSDNHMNTVSLKEALERFDDSR				429
Sbjct 361	VCHTHTVSYAEVGAERRRREGIEDSPPAITADKNQDSVSDNHMNTVSLKEALERFDDSR				420
Query 430	PSPSSKSSIKSSSHTAVSDPSSTPTKITTTETSTPPRQAITGLDKRRSSVSSQSMSSQSVS				489
Sbjct 421	PSPSSKSSIKSSSHTAVSDPSSTPTKITTTETSTPPRQAITGLDKRRSSVSSQSMSSQSVS				480
Query 490	QALSQSVMKQTASMQLQGGMAQPMFQFTAQFGAMKHLKDQLEQRTRIIEENIQRQOEELR				549
Sbjct 481	QALSQSVMKQTASMQLQGGMAQPMFQFTAQFGAMKHLKDQLEQRTRIIEENIQRQOEELR				540
Query 550	KIQEQLHMHVGQGIQMFLLQPPGGLNFGPVQVSSANSASIQQLPQLTMQOGVVQTNQLQA				609
Sbjct 541	KIQEQLHMHVGQGIQMFLLQPPGGLNFGPVQVSSANSASIQQLPQLTMQOGVVQTNQLQA				600
Query 610	GMNTGHVGTQHIMQ000LQTTSQGGQNIHGGHSQQTSLSSQTSGLTSPLYNTMVISQP				669
Sbjct 601	GMNTGHVGTQHIMQ000LQTTSQGGQNIHGGHSQQTSLSSQTSGLTSPLYNTMVISQP				660
Query 670	PSGSMVMQPSNIQQSNQGASVTTFPQDRQIRFSQAQQIVTKLVTTTPVACGTVMVPSTMF				729
Sbjct 661	PSGSMVMQPSNIQQSNQGASVTTFPQDRQIRFSQAQQIVTKLVTTTPVACGTVMVPSTMF				720
Query 730	GPVVTAYPTFATQ0000000PQTL SITQHOPNQDQ0SMPTVQQPAQAQLAQ0PQQFLQT				789
Sbjct 721	GPVVTAYPTFATQ0000000PQTL SITQHOPNQDQ0SMPTVQQPAQAQLAQ0PQQFLQT				780
Query 790	SRLVHGNQSTQLILSPFPVQNTFAPSHQQQLSHHRTDSMSDPKVKVQQQ				838
Sbjct 781	SRLVHGNQSTQLILSPFPVQNTFAPSHQQQLSHHRTDSMSDPKVKVQQQ				829