

Find a Gene Project (BGGN 213)

Akshara Balachandra, A11933675

May 15, 2019

Question 1: Name of the protein

Name: Sonic Hedgehog protein isoform 1 preprotein (SHH)

Accession: NP_000184.1

Species: Homo sapiens

Question 2: BLAST search

Method: TBLASTN search against Xenopus ESTs

Database: Expressed sequence tags (est)

Organism: Xenopus (Taxid: 8353)

Alignment details

Chosen match: Accession BX704643.1, a 874bp clone from *Xenopus tropicalis*.

>BX704643.1 BX704643 XGC-tadpole *Xenopus tropicalis* cDNA
clone TTPA011g13 5', mRNA sequence

cDNA clone TTPA011g13, mRNA sequence.
Length: 874

Score	Expect	Method	Identities
387 bits(994)	3e-132	Compositional matrix adjust.	197/285(69%)

Positives	Gaps	Frame
220/285(77%)	23/285(8%)	+2

Query	68	RNSERFKELTPNYPDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEG	127
		RNS+RFKELTPNYPDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEG	
Sbjct	41	RNSDRFKELTPNYPDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEG	220
Query	128	WDEGDGHHSEESLHYEGRAVDITTSRDRSKYGMLARLAVEAGFDWVYYESKAHIHCSVKA	187
		WDEGDGHHSEESLHYEGRAVDITTSRDRSKYGMLARLAVEAGFDWVY+ESKAHIHCSVKA	
Sbjct	221	WDEGDGHHSEESLHYEGRAVDITTSRDRSKYGMLARLAVEAGFDWVYFESKAHIHCSVKA	400
Query	188	ENSVAAKSGGCFPGSATVHLEQGGTKLVKDLSPGDRVLAADDQGRLLYSDFLTFDLRDDG	247
		ENSVAAKSGGCFPGSA V +E GGTK V++L PGDRVL++D QG L+YSDFL F+D++	
Sbjct	401	ENSVAAKSGGCFPGSARVMVEPGGTKAVREL RPGRVLS S DPQGNLIYSDFLLFIDKEHD	580

```

Query  248  AKKVFYVIETREPRERLLLTAAHLLFVAPHNDSATXXXXXXXXXXXXXXXXXXXXRALFAS  307
          KK++YVI+T + R R  +TAAHLLFVA  N + +                      +++FAS
Sbjct  581  VKKLYYVIQTSQTRIR--MTAAHLLFVAQSNGTGS-----FKSVFAS  700

Query  308  RVRPGQRVYVVAERDGRRLLPAAVHSVTLSEEAAGAYAPLTAQG  352
          VRPG  +Y    R  D  L  A  V  V  L  EE  G  +  P    G
Sbjct  701  NVRPGDVIYSADDR--DMTLREAMVEKVDL-EEDIGGFCPRNCPG  826

```

Translated BLAST: tblastn

blastn blastp blastx **tblastn** tblastx

TBLASTN search translated nucleotide databases using a protein query. [more...](#)

Enter Query Sequence

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

ref|NP_000184.1

Clear Query subrange [?](#)

From To

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

Job Title (2) - NP_000184:sonic hedgehog protein isoform 1... Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

Choose Search Set

Database [+](#) Expressed sequence tags (est) [?](#)

Organism Optional Xenopus (taxid:8353) ☐ exclude [+](#) Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude Optional ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Limit to Optional ☐ Sequences from type material

Entrez Query Optional Enter an Entrez query to limit search [?](#) [YouTube](#) [Create custom database](#)

BLAST Search database Expressed sequence tags (est) using Tblastn (search translated nucleotide databases using a protein query) ☐ Show results in a new window

[Algorithm parameters](#) **Note:** Parameter values that differ from the default are highlighted in yellow and marked with [+](#) sign

Figure 1: BLAST parameters

Job Title	NP_000184:sonic hedgehog protein isoform 1...
RID	DT912J1E014 Search expires on 05-17 07:43 am Download All ?
Program	TBLASTN ? Citation ?
Database	est See details ?
Query ID	NP_000184.1
Description	sonic hedgehog protein isoform 1 preproprotein [Homo sapiens]
Molecule type	amino acid
Query Length	462
Other reports	?

Filter Results	
Organism <small>only top 20 will appear</small> <input type="checkbox"/> exclude	
Type common name, binomial, taxid or group name	
+ Add organism	
Percent Identity	E value
<input type="text"/> to <input type="text"/>	<input type="text"/> to <input type="text"/>
Filter	Reset

Descriptions	Graphic Summary	Alignments	Taxonomy
---------------------	-----------------	------------	----------

Sequences producing significant alignments		Download ?	Manage Columns ?	Show <input type="text" value="100"/>	?
<input checked="" type="checkbox"/> select all	37 sequences selected				
	Description	Max Score	Total Score	Query Cover	E value
<input checked="" type="checkbox"/>	DC124442 Yamamoto/Hyodo-Miura NIBB/NBRP Xenopus DMZ pCS2p+ cDNA library Xenopus laevis cDNA clone x1304x08 5' mRNA sequen	420	420	51%	4e-146 85.65%
<input checked="" type="checkbox"/>	DC102103 Yamamoto/Hyodo-Miura NIBB/NBRP Xenopus DMZ pCS2p+ cDNA library Xenopus laevis cDNA clone x1228b16 5' mRNA sequen	394	394	43%	7e-136 91.63%
<input checked="" type="checkbox"/>	BX704643 XGC-ladpole Xenopus tropicalis cDNA clone TTPA011g13 5' mRNA sequence	387	430	64%	3e-132 69.12%

Figure 2: BLAST search results

Sequence ID: [BX704643.1](#) Length: 874 Number of Matches: 2

Score	Expect	Method	Identities	Positives	Gaps	Frame
387 bits(994)	3e-132	Compositional matrix adjust.	197/285(69%)	220/285(77%)	23/285(8%)	+2
Query	68	RNSERFKELTPNYPNDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEG		127		
Sbjct	41	RNS+RFKELTPNYPNDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEG		220		
		RNSDRFKELTPNYPNDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEG				
Query	128	WDEEDGHHSSESLHYEGRADVITTSDRDRSKYGMGLARLAVEAGFDWYYFESKAHIHCSVKA		187		
Sbjct	221	WDEEDGHHSSESLHYEGRADVITTSDRDRSKYGMGLARLAVEAGFDWYYFESKAHIHCSVKA		400		
		WDEEDGHHSSESLHYEGRADVITTSDRDRSKYGMGLARLAVEAGFDWYYFESKAHIHCSVKA				
Query	188	ENSVAASKSGGCFPGSATAHDLQGGTKLVKDLSPGDRVLVLYSDFLTFLDRDDG		247		
Sbjct	401	ENSVAASKSGGCFPGSA V +E GGTG V+H PGDRVL+H QG L+YSDFL F+D++		580		
		ENSVAASKSGGCFPGSARVMYFPGTKAVRLRNPGRDRLSSDPQGNLIYSDFLFLIKHED				
Query	248	AKKVFYVIETREPRERLLLTAAHLLFVAPHINDSTXXXXXXXXXXXXXXXXXXRRALFAS		307		
Sbjct	581	KK++YVI+T + R R +TAAHLLFVA N + + ++FAS		700		
		VKKLYYVIQTSTRIR--MTAAHLLFVAQSGNTGS-----FKSVFAS				
Query	308	RVRPGQRVYVVAERDGRRLLPAAHVSYLTSEEAAGAYAPLTAQG		352		
Sbjct	701	VRPG +Y R D L A V L V L E E G + P G		826		
		RVRPGDVIYSADRR--DMTLREAMVKEVDL--EEDIIGGFCPRNCPG				

Question 3: Novel Protein Information

>Xenopus tropicalis protein (from BLAST result)
RNSDRFKELTPNYPDIIFKDEENTGADRLMTQRCKDKLNALISVMNQWPGVKLRVTEG
WDEDGHHSEESLHYEGRAVDITTSDDRDRSKYGMLARLAVEAGFDWVYFESKAHIHCSVKA
ENSVAAKSGGCFPGSARVMVEPGGTAKAVRELRLPGDRVLSSDPQGNLIYSDFLFLIDKEHD
VKKLYYVIQTSQTRIRMTAAHLLFVAQSNGTGSFKSVFASNVRPGDVIYSADRRDMLTRE
AMVEKVDLEEDIGGFCPRNCPG

Species: *Xenopus (Silurana) tropicalis* (Common name: western clawed frog, Taxid: 8364)

Question 4: Proof that gene/protein are novel

3

blastn **blastp** **blastx** **tblastn** **tblastx**

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

Enter Query Sequence

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

From

To

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

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

☐ Align two or more sequences [?](#)

Choose Search Set

Database ?

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

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

Program Selection

Algorithm

☐ Quick BLASTP (Accelerated protein-protein BLAST)

☒ blastp (protein-protein BLAST)

☐ PSI-BLAST (Position-Specific Iterated BLAST)

☐ PHI-BLAST (Pattern Hit Initiated BLAST)

☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm [?](#)

Figure 4: BLASTP parameters

Sequences producing significant alignments [Download](#) [Manage Columns](#) [Show](#) [?](#)

☐ select all 0 sequences selected [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#)

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input type="checkbox"/>	PREDICTED: sonic hedgehog protein [Xenopus tropicalis]	533	533	100%	0.0	97.71%	XP_017950257.1
<input type="checkbox"/>	RefName: Full=Sonic hedgehog protein; AltName: Full=Shh unprocessed N-terminal signaling and C-terminal autoprocessing domains; S	465	465	100%	7e-162	83.71%	Q92000.1
<input type="checkbox"/>	PREDICTED: sonic hedgehog protein-like [Xenopus laevis]	462	462	99%	7e-161	84.03%	XP_018124514.1
<input type="checkbox"/>	sonic hedgehog protein precursor [Xenopus laevis]	442	442	100%	1e-152	80.38%	NP_001081782.1
<input type="checkbox"/>	shh protein [Xenopus tropicalis]	423	423	80%	3e-147	97.14%	AAI66395.1
<input type="checkbox"/>	PREDICTED: sonic hedgehog protein [Nanorana parkeri]	424	424	100%	5e-146	76.32%	XP_018409491.1

Figure 5: BLASTP results page

Download ▼ GenPept Graphics						
PREDICTED: sonic hedgehog protein [Xenopus tropicalis]						
Sequence ID: XP_017950257.1 Length: 416 Number of Matches: 1						
Range 1: 68 to 329 GenPept Graphics ▼ Next Match ▲ Previous Match						
Score	Expect	Method	Identities		Positives	Gaps
533 bits(1373)	0.0	Compositional matrix adjust.	256/262(98%)		256/262(97%)	0/262(0%)
Query 1		RNSDRFKELTPNYPDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEG				60
Sbjct 68		RNSDRFKELTPNYPDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEG				127
Query 61		WDEdGHSEESLHYEGrAVDITTSdRDRsKYGMLARLAVEAGFDWVYFESKAHIHCSVKA				120
Sbjct 128		WDEdGHSEESLHYEGrAVDITTSdRDRsKYGMLARLAVEAGFDWVYFESKAHIHCSVKA				187
Query 121		ENSVAaKSGGCFPGsARVMVEPGGTkAVRELrPGDrVLSsDPQGNLIYsDFLLFIDKEHD				180
Sbjct 188		ENSVAaKSGGCFPGsARVMVEPGGTkAVRELrPGDrVLSsDPQGNLIYsDFLLFIDKEHD				247
Query 181		VKKLYYVIQTSQTRIRMTAAHLLFVAQsNGTGSfKsVFASnVRPGDVIYsADRRDmTLRE				240
Sbjct 248		VKKLYYVIQTSQTRIRMTAAHLLFVAQsNGTGSfKsVFASnVRPGDVIYsADRRDmTLRE				307
Query 241		AMVEKVDLEEDIGGfCPRNCpG	262			
		AMVEKVDLEEDIG F P G				
Sbjct 308		AMVEKVDLEEDIGAFAPVTAQg	329			

Figure 6: Alignment for top result