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 (2.7.1) 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	RNSERFKELTPNYNPDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEG RNS+RFKELTPNYNPDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEG	127
Sbjct	41	RNSDRFKELTPNYNPDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEG	220
Query	128	WDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLAVEAGFDWVYYESKAHIHCSVKA WDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLAVEAGFDWVY+ESKAHIHCSVKA	187
Sbjct	221	WDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLAVEAGFDWVYFESKAHIHCSVKA	400
Query	188	ENSVAAKSGGCFPGSATVHLEQGGTKLVKDLSPGDRVLAADDQGRLLYSDFLTFLDRDDG ENSVAAKSGGCFPGSA V +E GGTK V++L PGDRVL++D QG L+YSDFL F+D++	247
Sbjct	401	${\tt ENSVAAKSGGCFPGSARVMVEPGGTKAVRELRPGDRVLSSDPQGNLIYSDFLLFIDKEHD}$	580

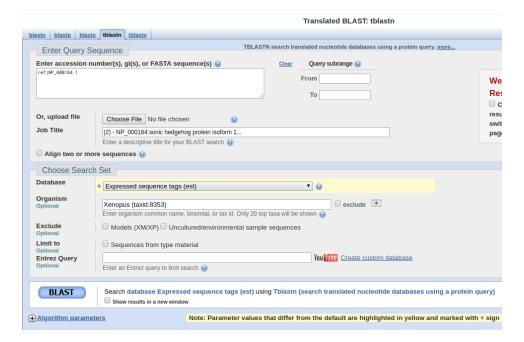


Figure 1: BLAST parameters

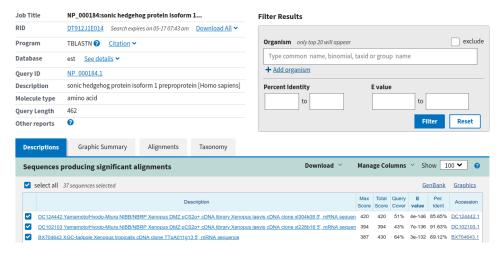


Figure 2: BLAST search results

BX704643 XGC-tadpole Xenopus tropicalis cDNA clone TTpA011g13 5', mRNA sequence

Sequence ID: BX704643.1 Length: 874 Number of Matches: 2

Score Expect Method Identities Positives Gaps 387 bits(994) 3e-132 Compositional matrix adjust. 197/285(69%) 220/285(77%) 23/285(8 Query 68 RNSERFKELTPNYNPDIJFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEG RNS+RFKELTPNYNPDIJFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEG 220 Query 128 WDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLAVEAGFDWVYYESKAHIHCSVKA WDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLAVEAGFDWVYYESKAHIHCSVKA WDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLAVEAGFDWVYYESKAHIHCSVKA 400 Query 188 ENSVAAKSGGCFPGSATVHLEQGGTKLVKDLSPGGRVLADDQGRLLYSDFLTFLDRDDG ENSVAAKSGGCFPGSATVMVEPGGTKLVFLDGDVLADDQGRLLYSDFLTFLDRDDG 247 ENSVAAKSGGCFPGSATVMVEPGGTKLVKDLSPGDRVLADDQGRLLYSDFLTFLDRDDG 247 BONG AKKVFYVIETREPRERLLTAAHLLFVAPHNDSATXXXXXXXXXXXXXXXXXXXXXXXXXXXABLFAS 307	Frame %) +2
uery 68 RNSERFKELTPNYNPDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEG RNS+RFKELTPNYNPDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEG bjct 41 RNSDRFKELTPNYNPDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEG 220 uuery 128 WDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLAVEAGFDWVYYESKAHIHCSVKA WDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLAVEAGFDWVYYESKAHIHCSVKA 400 uuery 188 ENSVAAKSGGCFPGSATVHLEQGGTKLVKDLSPGDRVLAADDQGRLLYSDFLTFLDRDDG 247 ENSVAAKSGGCFPGSA V +E GGTK V++L PGDRVL++D QG L+YSDFL F+D++ bjct 401 ENSVAAKSGGCFPGSARVMVEPGGTKAVRELRPGDRVLSSDPQGNLIYSDFLLFIDKEHD 580 uuery 248 AKKVFYVIETREPRERLLLTAAHLLFVAPHNDSATXXXXXXXXXXXXXXXXXXXXXXXXXARALFAS 307	%) +2
RNS+RFKELTPHYNPDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEG RNSDRFKELTPHYNPDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEG RNSDRFKELTPHYNPDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEG 220 uery 128 WDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLAVEAGFDWVYYESKAHIHCSVKA bjct 221 WDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLAVEAGFDWVYFESKAHIHCSVKA 400 uery 188 ENSVAAKSGGCFPGSATVHLEQGGTKLVKDLSPGDRVLAADDQGRLLYSDFLTFLDRDDG ENSVAAKSGGCFPGSA V +E GGTK V++L PGDRVL++D QG L+YSDFL F+D++ bjct 401 ENSVAAKSGGCFPGSARVMVEPGGTKAVRELRPGDRVLSSDPQGNLIYSDFLLFIDKEHD 580 uery 248 AKKVFYVIETREPRERLLLTAAHLLFVAPHNDSATXXXXXXXXXXXXXXXXXXXXXXXXXARALFAS 307	
bjct 41 RNSDRFKELTPNYNPDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEG 220 uery 128 WDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLAVEAGFDWVYYESKAHIHCSVKA WDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLAVEAGFDWVYYESKAHIHCSVKA 400 uery 188 ENSVAAKSGGCFPGSATVHLEQGGTKLVKDLSPGDRVLAADDQGRLLYSDFLTFLDRDDG 247 ENSVAAKSGGCFPGSA V +E GGTK V++L PGDRVL++D QG L+YSDFL F+D++ bjct 401 ENSVAAKSGGCFPGSARVMVEPGGTKAVRELRPGDRVLSSDPQGNLIYSDFLLFIDKEHD 580 uery 248 AKKVFYVIETREPRERLLLTAAHLLFVAPHNDSATXXXXXXXXXXXXXXXXXXXXXXXXXXARALFAS 307	
WDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLAVEAGFDWVY+ESKAHIHCSVKA WDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLAVEAGFDWVYYESKAHIHCSVKA 400 WDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLAVEAGFDWVYFESKAHIHCSVKA 400 WDEDGHHSEESLHYEGRAVDHS 400 WDEDGHAVENDAM 400 WDEDGHHSEESLHYEGRA	
bjct 221 WDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLAVEAGFDWVYFESKAHIHCSVKA 400 uery 188 ENSVAAKSGGCFPGSATVHLEQGGTKLVKDLSPGDRVLAADDQGRLLYSDFLTFLDRDDG 247 ENSVAAKSGGCFPGSA V +E GGTK V++L PGDRVL++D QG L+YSDFL F+D++ bjct 401 ENSVAAKSGGCFPGSARVMVEPGGTKAVRELRPGDRVLSSDPQGNLIYSDFLLFIDKEHD 580 uery 248 AKKVFYVIETREPRERLLLTAAHLLFVAPHNDSATXXXXXXXXXXXXXXXXXXXXXXXXXXXARALFAS 307	
ENSVAAKSGGCFPGSA V +E GGTK V++L PGDRVL++D QG L+YSDFL F+D++ bjct 401 ENSVAAKSGGCFPGSARVMVEPGGTKAVRELRPGDRVLSSDPQGNLIYSDFLLFIDKEHD 580 LUEY 248 AKKVFYVIETREPRERLLLTAAHLLFVAPHNDSATXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	
bjct 401 ENSVAAKSGGCFPGSARYMVEPGGTKAVRELRPGDRVLSSDPQGNLIYSDFLLFIDKEHD 580 uery 248 AKKVFYVIETREPRERLLLTAAHLLFVAPHNDSATXXXXXXXXXXXXXXXXXXXXXXXARALFAS 307	
KK++YVI+T + R R +TAAHLLFVA N + + +++FAS	
bjct 581 VKKLYYVIQTSQTRIRMTAAHLLFVAQSNGTGSFKSVFAS 700	
uery 308 RVRPGQRVYVVAERDGDRRLLPAAVHSVTLSEEAAGAYAPLTAQG 352 VRPG +Y R D L A V V L EE G + P G	
bjct 701 NVRPGDVIYSADRRDMTLREAMVEKVDL-EEDIGGFCPRNCPG 826	

Figure 3: Alignment details

Question 3

Chosen Sequence:

>Xenopus tropicalis protein (from BLAST result)
RNSDRFKELTPNYNPDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEG
WDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLAVEAGFDWVYFESKAHIHCSVKA
ENSVAAKSGGCFPGSARVMVEPGGTKAVRELRPGDRVLSSDPQGNLIYSDFLLFIDKEHD
VKKLYYVIQTSQTRIRMTAAHLLFVAQSNGTGSFKSVFASNVRPGDVIYSADRRDMTLRE
AMVEKVDLEEDIGGFCPRNCPG

Name: Xenopus sonic hedgehog

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

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Pipoidea; Pipidae; Xenopodinae; Xenopous; Silurana

Question 4

A BLASTP search against the NR database yielded a top hit to a protein from $Xenopus\ tropicalis$ with 100% coverage and 97.71% identity. The alignment details are given in the screenshots below.

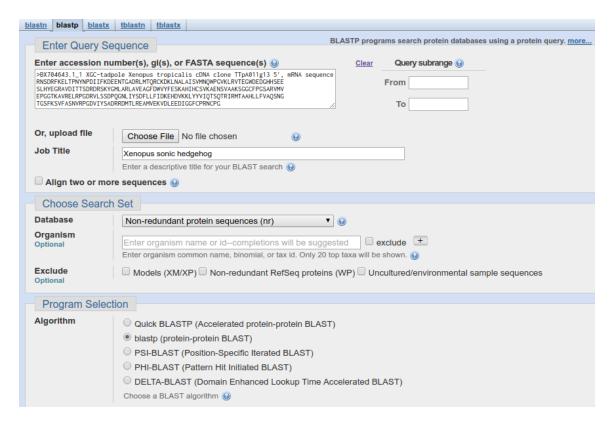


Figure 4: BLASTP parameters

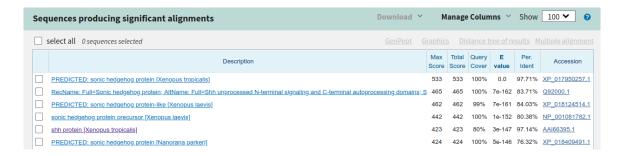


Figure 5: BLASTP results page

<u>♣ Download</u> ✓ <u>GenPept</u> <u>Graphics</u>

PREDICTED: sonic hedgehog protein [Xenopus tropicalis]

Sequence ID: XP_017950257.1 Length: 416 Number of Matches: 1

Range	1: 68	3 to 329 <u>G</u>	GenPept Graphics				▼ <u>N</u>	ext Match A	Prev
Score		Expect	Method		Identities	Positives		Gaps	_
533 b	its(13	73) 0.0	Compositional	matrix adjust.	256/262(98%)	256/262(97	7%)	0/262(0%)
uery	1		LTPNYNPDIIFKDEE LTPNYNPDIIFKDEE				60		
Sbjct	68		LTPNYNPDIIFKDEE				127	7	
)uery	61		EESLHYEGRAVDITT				120	9	
Sbjct	128	WDEDGHHS	EESLHYEGRAVDITT	SDRDRSKYGMLA	RLAVEAGFDWVYFES	KAHIHCSVKA	187	7	
uery	121		GGCFPGSARVMVEPG GGCFPGSARVMVEPG				180	9	
Sbjct	188	ENSVAAKS	GGCFPGSARVMVEPG	GTKAVRELRPGDI	RVLSSDPQGNLIYSD	FLLFIDKEHD	247	7	
uery	181		QTSQTRIRMTAAHLL QTSQTRIRMTAAHLL				240)	
Sbjct	248		QTSQTRIRMTAAHLL				307	7	
Query	241		EEDIGGFCPRNCPG EEDIG F P G	262					
Sbjct	308	AMVEKVDL	EEDIGAFAPVTAQG	329					

Figure 6: Alignment for top result