

FIND A GENE PROJECT

Jessica Le (PID: A17321021)
lil044@ucsd.edu

Q1)

Name: PARP1 - poly(ADP-ribose) polymerase 1

Accession: NP_001609.1

Species: Homo Sapiens

Function: enables DNA, RNA, and NAD binding for involvement in DNA damage response, ATP generation from poly-ADP-D-ribose, double-strand break repair, and activation of cellular response during oxidative stress

Q2)

Method: tblastn search against the Expressed sequence tags database excluding Homo Sapiens

Database: Expressed sequence tags (est)

Organism (to exclude): homo sapien (taxid: 9606)

Translated BLAST: tblastn

blastn blastp blastx **tblastn** tblastx

Enter Query Sequence

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

REF|NP_001609.1

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

Organism [Optional](#) ☒ exclude [Add organism](#)

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 est using Tblastn (search translated nucleotide databases using a protein query)

☐ Show results in a new window

Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign

The output of the BLAST search is:

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
✓	Ferret_002_Q15 Ferret library from spleen, liver, and thymus Mustela putorius furo cDNA, mRNA sequence	Mustela putorius...	628	628	35%	0.0	85.79%	1232	GD181183.1
✓	Elaeis guineensis EST, clone (6lib)_isotg11417, mRNA sequence	Elaeis guineensis	623	623	98%	0.0	37.69%	3195	LM613445.1
✓	LB0165.CR_D07_GC_BGC-16 Bos taurus cDNA clone IMAGE:8079417 5', mRNA sequence	Bos taurus	617	617	31%	0.0	95.21%	940	DT816785.1
✓	LB02755.CR_D08_GC_BGC-27 Bos taurus cDNA clone IMAGE:8630554, mRNA sequence	Bos taurus	604	604	29%	0.0	96.66%	928	EE349379.1
✓	HX430312 full-length enriched common marmoset ES cells cDNA library Callithrix jacchus cDNA clone MES-161J...	Callithrix jacchus	579	579	28%	0.0	98.23%	849	HX430312.1
✓	DKFZp470P2427_r1 470 (synonym: pliv1) Pongo abelii cDNA clone DKFZp470P2427 5', mRNA sequence	Pongo abelii	577	577	28%	0.0	97.54%	855	CR765049.1
✓	LB01627.CR_G11_GC_BGC-16 Bos taurus cDNA clone IMAGE:8087941 5', mRNA sequence	Bos taurus	576	576	29%	0.0	96.90%	872	DT815491.1
✓	LB03012.CR_J17_GC_BGC-30 Bos taurus cDNA clone IMAGE:8136691 5', mRNA sequence	Bos taurus	573	573	29%	0.0	96.19%	940	DV928123.1
✓	FY610885 full-length enriched tammar ovary cDNA library Notamacropus eugenii cDNA clone MEQC-012L10 5', ...	Notamacropus e...	566	566	30%	0.0	89.54%	927	FY610885.1
✓	ILLUMIGEN_MCQ_28629 Katze MMPB Macaca mulatta cDNA clone IBIUW:7518 5' similar to Bases 3 to 1054 hi...	Macaca mulatta	561	561	28%	0.0	97.86%	1054	CN647510.1
✓	LB02341.CR_A21_GC_BGC-23 Bos taurus cDNA clone IMAGE:8605151 5', mRNA sequence	Bos taurus	558	558	27%	0.0	96.76%	846	EV634267.1
✓	LB03018.CR_G18_GC_BGC-30 Bos taurus cDNA clone IMAGE:8138924 5', mRNA sequence	Bos taurus	551	551	28%	0.0	94.31%	846	DV930029.1
✓	AGENCOURT_11295825 NIH_MGC_164 Mus musculus cDNA clone IMAGE:30146975 5', mRNA sequence	Mus musculus	550	550	29%	0.0	89.86%	900	CA977714.1
✓	CANFLECK_JINDO_BR_80_C09.ab1 Jindo brain cDNA library Canis lupus familiaris cDNA clone 80_C09, mRNA...	Canis lupus famil...	548	548	27%	0.0	95.67%	833	GR888037.1
✓	CJ029757 full-length enriched swine cDNA library, adult trachea Sus scrofa cDNA clone TCH01E070100 5', mRN...	Sus scrofa	545	545	27%	0.0	95.64%	828	CJ029757.1
✓	Mus musculus mRNA 5-prime sequence from clone LA0AAA13YB13 (LA0AAA13YB13CM1), mRNA sequence	Mus musculus	543	543	27%	0.0	95.22%	818	FO737852.1
✓	BovGen_15742 normal cattle brain Bos taurus cDNA clone RZPDp1056E0422Q 5', mRNA sequence	Bos taurus	534	534	26%	0.0	95.15%	816	CQ887417.1
✓	Rattus norvegicus LA0ACA6YA02 mRNA sequence	Rattus norvegicus	429	827	43%	0.0	94.50%	2002	FO801981.1
✓	JGI_CABH9820.fwd NIH_XGC_tropSkeMus1 Xenopus tropicalis cDNA clone IMAGE:7852834 5', mRNA sequence	Xenopus tropicalis	530	530	29%	4e-179	82.61%	903	DT393336.1
✓	LB03015.CR_D01_GC_BGC-30 Bos taurus cDNA clone IMAGE:8137683 5', mRNA sequence	Bos taurus	523	523	29%	2e-176	86.24%	905	DV928965.1
✓	HX407749 full-length enriched common marmoset ES cells cDNA library Callithrix jacchus cDNA clone MES-098D...	Callithrix jacchus	521	521	25%	2e-176	98.43%	768	HX407749.1
✓	HX639419 full-length enriched common marmoset testis cDNA library Callithrix jacchus cDNA clone MTE-133O21...	Callithrix jacchus	519	519	25%	1e-175	97.25%	765	HX639419.1
✓	904613 MARC 4PIG Sus scrofa cDNA 5', mRNA sequence	Sus scrofa	513	513	25%	2e-173	97.64%	764	CK451438.1
✓	LB0284.CR_I13_GC_BGC-28 Bos taurus cDNA clone IMAGE:8284119 5', mRNA sequence	Bos taurus	511	511	25%	1e-172	96.08%	767	DV907959.1
✓	AGENCOURT_28621953 NIH_MGC_250 Rattus norvegicus cDNA clone IMAGE:7384303 5', mRNA sequence	Rattus norvegicus	509	509	26%	1e-171	94.25%	786	CQ560728.1
✓	JGI_XZG45550.fwd NIH_XGC_tropGas7 Xenopus tropicalis cDNA clone IMAGE:7560326 5', mRNA sequence	Xenopus tropicalis	508	508	28%	7e-171	82.99%	867	CX516023.2
✓	Rattus norvegicus mRNA 5-prime sequence from clone LA0ACA102YC15 (LA0ACA102YC15CM1), mRNA seque...	Rattus norvegicus	506	506	25%	9e-171	94.19%	776	FO771853.1
✓	LB02627.CR_F21_GC_BGC-26 Bos taurus cDNA clone IMAGE:8219927 5', mRNA sequence	Bos taurus	503	503	29%	6e-169	85.86%	870	DV883447.1

Chosen Match (highlighted in grey): Accession LM613445.1, a 3195 bp mRNA sequence from *Elaeis guineensis*. The E-value score is 0.0 with a percent identity of 37.69%. See alignment details below.

Elaeis guineensis EST, clone (6lib)_isotig11417, mRNA sequenceSequence ID: [LM613445.1](#) Length: 3195 Number of Matches: 1Range 1: 108 to 2999 [GenBank](#) [Graphics](#)[▼ Next Match](#) [▲ Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps	Frame
623 bits(1607)	0.0	Compositional matrix adjust.	399/1040(38%)	583/1040(56%)	125/1040(12%)	+3
Query 7		KLYRVEYAKSGRASCKKSESIPKDSLRMAIMVQSPMFDGKVPWHYHFSCFWKVGH SIRH				66
Sbjct 108		K ++ EYAKSGR+SCK C SI KD LR+ MV + FDG +P W H C +K G+ I+ KPWKA EYAKSGRSSCKSKNSIGKDELRLGKMVTATQFDGFMPMWNHAGCIFKKGNQIKS				287
Query 67		PDVEVDGFSSELRWDDQKVKKTA EAGVGTGKGQDGIGSKAEKTLGDFAAEYAKSNRSTCK				126
Sbjct 288		D +V+G LRW+DQK++K E G V+ S E T+ E +++R+ C+ VD-DVEGIDLLRWEDQKIRKYTEGGSVSTTAV----SNDECTI-----EVSQTSRAACR				437
Query 127		GCMEKIEKGQVRLSKKMDPE-KPQLGMIDRWYHPGCFVKNREELGFRPEYSASQLKGFS				185
Sbjct 438		C +KI KG VR+S K PE + G+ W+H CF++ P ++ G+ RCNQKIMKGMVRVSTK---PEGQARGLA---WHHVNCFIE-----MSPSTIIEKMSGWD				584
Query 186		LLATEDK-----EALKKQLPG-----VKSEGKR-----KGDEVDGVD E V a				220
Sbjct 585		L+ EDK A ++QL V SE K DE D + SLSPEDKVTVTALVKKDKSNKNTAQEQQLSGTKRKKVGS E DHHSKVPKSDENDSAGGAS				764
Query 221		kkkskkekdkdskle----kalkaQNDLIWNIKDELKKVCSTNDLKELLIFNKQVPSGE				276
Sbjct 765		K++ E K L+ Q+ +W IKDELKK T +L+E+L N+Q E SKENPAESGNAYSSSIELEKKLEEQSKALWEIKDELKKHVMTAELREMLEANRQDSAGSE				944
Query 277		SAILDRVADGMVFGALLPCEECGQLVFKSDAYYCTGDVTAWTKCMVKTQTPNR--KEWV				334
Sbjct 945		+ DR ADGM+FGAL C CSG L + Y C G V+AW++C T P R ++W YDLRDRCADGMLFGALGKCPICSGSLHYSGGQYRCHGYVSAWSRCSYSTTNPLRLKEKWK				1124
Query 335		TPKEFRE---ISYLKLLKVKKQDRIFppetsasvaatpppstasapaavnssasaDKPLS				391
Sbjct 1125		P+E + + K K KK +R+ P +S + + + + DK L VPEETSNRYLLKWFKSQKAKKPNRVLPSNSSKSSCSAGMQSHPSNG-----DK-LE				1277
Query 392		NMKILTLGKLSRNKDEVKAMIEKLGKLTGTANKASLCISTkkevekmnkkmeevkeANI				451
Sbjct 1278		++K+ +G+ + +E K E+ GKG+ K + C+ E+ + ++ + + I SLKVAVVGE-PQKAN EWKHRFE EAGGKI HAKIKKDTNCLVLIGEMVDKDSEIRKARRMKI				1454
Query 452		RVVSEDFLQDVSASTKSLQ-ELFLAHILSPWGaevkaepvevvapRgksgaalskkskgQ				510
Sbjct 1455		+V ED+L + + K + +L+ S PIVREDYLHECISKQKKIPFDLYKVEAAS-----E				1544
Query 511		VKEEGINKSEKRMKLT LKGGAAVDPDPSGLEHSAHVLEKGGKVFSATLGLVDIVKGTNSYY				570
Sbjct 1545		G+ + + +KG +AV SGL+ H+LE G +++ TL + D+ G NSYY TSRGGM-----VTVKVKGSRSAVHEASGLQDVGHILEDGNSIYNTTLNMSDLSTGINSYY				1706
Query 571		KLQLLEDDKENRYWIFRSWGRVGT-VIGSNKLEQMPSKEDAIEQFMKLYEKTGNWHS-				628
Sbjct 1707		LQ++++DK + ++FR WGRVG IG KLE+M SK DAI++F +L+ EKTGN W + ILQIIQEDKGSACFVFRKWGRVGNKIGGKTLEEM-SKSDAIQEFKRLFLEKTGNPWEAW				1883
Query 629		---KNFTKYPKKFYPLEIDYGQDEEAVKKLTVPNGTKSKLPKPVQDLIKMIFDVESMKKA				685
Sbjct 1884		KNF K P +F+PL+IDYG + + KK + N KS L + +L+KM+FDVE+ + A EQKKNF EKQPGRFFPLDIDYGVKQVSKKKDSAN--IKSLLAPQLIELMKMLFDVETYRAA				2057
Query 686		MVEYEIDLQKMPGLGKLSKRQIQAAYSILSEVQQAVSQGSS-----DSQILDLSNRFYTL				739
Sbjct 2058		M+E+EI++ +MPLGKLSK IQ + L+E+Q ++ + +S I+D SNRF+TL MLEFEINMSEMPGLGKLSKMNIIQKGFEALTEIQNLLNNNAKHDPVVKESLIVDASNRF+TL				2237
Query 740		IPHDFGMKKPPLNNADSVQAKVEMLDNLLDIEVAYSLLRGGSDSSKDPIDVNYEKLKT				799
Sbjct 2238		IP P ++ + D +AKV+ML+ L DIE+A L+ G D + + +DV Y+KL+ IPSIH-----PHVIRDEDDFKAKVKMLEALQDIEIASRLV--GFDSGNDES LDVYKKLQC				2399
Query 800		DIKVVDRDSEAEIIRKYVNKTHATTHSAYDLEVIDIFKIEREGECQRYKPFK-QLHNRR				858
Sbjct 2400		DI + DSE+ +++ KY+ NTHA TH + LE+ ++F +EREGE +Y P++ +L N+ DITPLPHDSEDYKLVEKYLNTHTAPHKWESLEEEV FALEREGEYDKYTPYRDKLQNM				2579
Query 859		LLWHGSRTTNFAGILSQGLRIAPPEAVPTGYMFGKGIYFADMVSKSANYHTSQGDP IGL				918
Sbjct 2580		LLWHGSR TNF GI+SQGLRIAPPEAP TGYMFGKGIYFAD+VSKSA Y + + +P+GL LLWHGSR LTNFVGII SQGLRIAPPEAPATGYMFGKGIYFADLVSKSAQYCYVDKKNPVGL				2759
Query 919		ILLGEVALGNMYELKHASHISRLPKGKHSVKGLGKTTDPDSANISL-DGVDVPLGTGISS				977
Sbjct 2760		+LL EVALG +YELK A+++ + PKGKHS KGLGKT P S + D V VP G + S MLLSEVALGEIYELKKATYMEKPPKGKHS TKGLGKTVPLESDFVKWQDQVVVPCGKPVPS				2939
Query 978		GVIDTSLLYNEYIVYDIAQV 997				
Sbjct 2940		+ + LLYNEYIVYD AQV 2999 SIRASELLYNEYIVYDIAQV				

Alignment Details -

Score	Expect	Method	Identities	Positives	Gaps	Frame
623 bits(1607)	0.0	Compositional matrix adjust.	399/1040(38%)	583/1040(56%)	125/1040(12%)	+3
Query 7		KLYRVEYAKSGRASCKKCSESI PKDSLRLMAIMVQSPMFDGKVPWHYHFSCFWKVGHSIRH				66
Sbjct 108		K ++ EYAKSGR+SCK C SI KD LR+ MV + FDG +P W H C +K G+ I+				287
Query 67		PDVEVDGFSELRWDDQQKVKKTAEAGGVTGKGQDGIGSKAEKTLGDFAAEYAKSNRSTCK				126
Sbjct 288		D +V+G LRW+DQQK++K E G V+ S E T+ E ++++R+ C+				437
Query 127		GCMEKIEKGQVRLSKKMVDPE-KPQLGMIDRWYHPGCFVKNREELGFRPEYSASQLKGFS				185
Sbjct 438		C +KI KG VR+S K PE + G+ W+H CF++ P ++ G+				584
Query 186		LLATEDK-----EALKKQLPG-----VKSEGKR---KGDEV DGVDEVa				220
Sbjct 585		L+ EDK A ++QL V SE K DE D +				764
Query 221		kkkskkekdkdskle---kalkAQNDLIWNIKDELKKVCSTNDLKELLIFNKQQVPSGE				276
Sbjct 765		K++ E K L+ Q+ +W IKDELKK T +L+E+L N+Q E				944
Query 277		SAILDRVADGMVFGALLPCEECGQLVFKSDAYYCTGDVTAWTKCMVKTQTPNR--KEWV				334
Sbjct 945		+ DR ADGM+FGAL C CSG L + Y C G V+AW++C T P R ++W				1124
Query 335		TPKEFRE---ISYLKKLKVKKQDRIFppetsasvaatpppstasapaavnssasADKPLS				391
Sbjct 1125		P+E + + K K KK +R+ P +S + + + + DK L				1277
Query 392		NMKILTLGKLSRNKDEVKAMIEKLGKLTGTANKASLCISTkkevekmnkkmeevkeANI				451
Sbjct 1278		++K+ +G+ + +E K E+ GGK+ K + C+ E+ + ++ + + I				1454
Query 452		RVVSEDFLQDVSASTKSLQ-ELFLAHILSPWGaevkaepvevvapRgksgaalskkskgQ				510
Sbjct 1455		+V ED+L + + K + +L+ S +				1544
Query 511		VKEEGINKSEKRMKLTCLKGGAAVDPDSGLEHSAHVLEKGGKVFSA TLGLVDIVKGTNSYY				570
Sbjct 1545		G+ + + +KG +AV SGL+ H+LE G +++ TL + D+ G NSYY				1706
Query 571		KLQLEDDKENRYWIFRSWGRVGT-VIGSNKLEQMPSKEDAIEQFMKLYEEKTGNAWHS-				628
Sbjct 1707		LQ++++DK + ++FR WGRVG IG KLE+M SK DAI++F +L+ EKTGN W +				1883
Query 629		ILQIIQEDKGSACFVFRKWGRVGN DKIGGTKLEEM-SKSDAIQE FKR L FLEKTGNPWEAW				
Query 629		---KNFTKYPKKFYPLEIDYGQDEEAVKLT VNPGT SKLPKPVQDLIKMIFDVESMKKA				685
Sbjct 1884		KNF K P +F+PL+IDYG + + KK + N KS L + +L+KM+FDVE+ + A				2057
Query 686		EQKKNFEKQPGRRFFPLDIDYGVKQVSKKKDSAN--IKSLLAPQLIELMKMLFDVETYRAA				
Query 686		MVEYEIDLQKMPLGKLSKRQIQAAYSILSEVQQAVSQGSS-----DSQILDLSNRFYTL				739
Sbjct 2058		M+E+EI++ +MPLGKLSK IQ + L+E+Q ++ + +S I+D SNRF+TL				2237
Query 740		MLEFEINMSEMP LGKLSKMNIQKGFEALTEIQNLLNNNAKHDPVVKESLIVDASNRFCTL				
Query 740		IPHDFGMKKPPLLNNADSVQAKVEMLDNLLDIEVAYSLLRGGSDSSKDPIDVNYEKLKT				799
Sbjct 2238		IP P ++ + D +AKV+ML+ L DIE+A L+ G D + + +DV Y+KL+				2399
Query 800		IPSIH----PHVIRDEDDFKAKVKMLEALQDIEIASRLV--GFDSGNDESLDVKYKKLQC				
Query 800		DIKVVDRDSEEA EIIRKYVKNTHATTHSAYDLEVIDIFKIEREGECQRYKPKF-QLHNRR				858
Sbjct 2400		DI + DSE+ +++ KY+ NTHA TH + LE+ ++F +EREGE +Y P++ +L N+				2579
		DITPLPHDSEDYKLVEKYL NTHAPTHKEWSLEEEVFALEREGEYDKYTPYRDKLQNKM				

Query	859	LLWHGSRTTNFAGILSQGLRIAPPEAPVTGYMFGKGIYFADMVSKSANYYHTSQGDPIGL	918
		LLWHGSR TNF GI+SQGLRIAPPEAP TGYMFGKGIYFAD+VSKSA Y + + +P+GL	
Sbjct	2580	LLWHGSRLTNFVGIIISQGLRIAPPEAPATGYMFGKGIYFADLVSKSAQYCYVDKKNPVGL	2759
Query	919	ILLGEVALGNMYELKHASHISRLPKGKHSVKGLGKTTTPDPSANISL-DGVDVPLGTGISS	977
		+LL EVALG +YELK A+++ + PKGKHS KGLGKT P S + D V VP G + S	
Sbjct	2760	MLLSEVALGEIYELKKATYMEKPPKGKHSTKGLGKTVPLESDFVKWQDQVVVPCGKPVPS	2939
Query	978	GVIDTSLLYNEYIVYDIAQV	997
		+ + LLYNEYIVYD AQV	
Sbjct	2940	SIRASELLYNEYIVYDTAQV	2999

Q3) The protein sequence translated from EMBOSS Transeq at the EBI is.

```
>LM613445.1_1 Elaeis guineensis EST, clone (6lib)_isotig11417, protein
QPFSLSFARSRARSRLAEIILLDLGRPSNSSAMATPPKPWKAHEYAKSGRSSCKSCKNSIGKD
ELRLGKMVTATQFDGFMPMWNHAGCIFKKGNQIKSVDDVEGIDLLRWEDQQKIRKYTEGG
SVSTTAVSNDECTIEVSQTSRAACRRCNQKIMKGMVRVSTKPEGQGARGLAWHHVNCFIE
MSPSTIIIEKMSGWDSLSPEDKVTVTALVKKDKSNKNTAQEQQLSKGTRKRVKVGSEDHHSK
VPKSDENDSAGGASSKENPAESGNAYSSSIELEKKLEEQSKALWEIKDELKKHVMTAELR
EMLEANRQDSAGSEYDLRDRCADGMLFGALGKCPICSGSLHYSGGQYRCHGYVSAWSRCS
YSTTNPLRLKEKWVPEETSNRYLLKWFKSQKAKKPNRVLPSPSSNKSSCSAGMQSHPSN
GDKLESCLKVAVVGEPQKANEWKHRFEEAGGKIHAKIKKDTNCLVLIGEMVDKDSEIRKAR
RMKIPIVREDYLHECISKQKIPFDLYKVEAASETSRGGMVTVKVKGSAVHEASGLQDV
GHILEDGNSIYNTTLNMSDLSTGINSYYILQIIQEDKGSACFVFRKWGRVGNKIGGTKL
EEMSKSDAIQEFKRLFLEKTGNPWEAWEQKKNFEKQPGRFFPLDIDYGVKQVSKKKDSAN
IKSLLAPQLIELMKMLFDVETYRAAMLEFEINMSEMPLGKLSKMNIQKGFALTEIQNLL
NNNAKHDPVVKESLIVDASNRFFTLIPSIHPHVIRDEDDFKAKVKMLEALQDIEIASRLV
GFDSGNDESLDVYKKLQCDITPLPHDSEDYKLVEKYLNTHAPTHKEWSLELEEVFAL
REGKYDKYTPYRDKLQNKMLLWHGSRLTNFVGIIISQGLRIAPPEAPATGYMFGKGIYFAD
LVSKSAQYCYVDKKNPVGLMLLSEVALGEIYELKKATYMEKPPKGKHSTKGLGKTVPLES
DFVKWQDQVVVPCGKPVPSIRASELLYNEYIVYDTAQVKLQFLKVRFHHRK*KVLDVL
QMIFLRMLGVCQ*AAR*NG*MCS*CWCFLGCFKYQPFATILGSX
```

Name: Elaeis guineensis EST, clone (6lib)_isotig11417

Species: Elaeis guineensis

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; Liliopsida; Arecaceae; Arecoideae;
Cocoseae; Elaeidinae; Elaeis

Q4) A blastp search against the non-redundant protein sequences (NR) database yielded results all with less than 100% identity. A screenshot of the output is attached below along with the setup screenshot.

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) ? Clear

RNLSLSRSLGRALARRRSS*TSVGPQIHRQWRPRRNHGRSTPSRGGLRAS
RARTPSAR
TSSASARWLPPPSSTDSCPCGIMLGASSRKETRLNLLMTLKA*TCFVGRISRR*
GSTQRV

Query subrange ?
From
To

Or, upload file

Choose File

No file chosen ?

Job Title

Enter a descriptive title for your BLAST search ?

☐ Align two or more sequences ?

Choose Search Set

Databases

☒ Standard databases (nr etc.): ☐ Experimental databases

Compare

☐ Select to compare standard and experimental database ?

Standard

Database

Non-redundant protein sequences (nr) ?

Organism
Optional

Enter organism name or id--completions will be suggested

☐ exclude

Add organism

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown ?

Exclude
Optional

☐ 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)

☐ Show results in a new window

Job Title

Protein Sequence

RID

[TY9GH2PM013](#)
Search expires on 02-03 15:00 pm
[Download All](#)

Program

BLASTP [Citation](#)

Database

nr [See details](#)

Query ID

lcl|Query_4160333

Description

unnamed protein product

Molecule type

amino acid

Query Length

1065

Other reports

[Distance tree of results](#)
[Multiple alignment](#)
[MSA viewer](#)

Filter Results

Organism

only top 20 will appear

☐ exclude

[+ Add organism](#)

Percent Identity

to

E value

to

Query Coverage

to

Filter

Reset

Compare these results against the new Clustered nr database

BLAST

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download

Select columns

Show 100

☒ select all
 6 sequences selected

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

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	hypothetical protein lba_chr07aCG1000 [Ipomoea batatas]	Ipomoea batatas	91.7	91.7	10%	2e-17	50.86%	120	GMD12175.1
<input checked="" type="checkbox"/>	hypothetical protein lba_chr07bCG1530 [Ipomoea batatas]	Ipomoea batatas	89.0	89.0	10%	2e-16	50.00%	120	GMD13932.1
<input checked="" type="checkbox"/>	hypothetical protein HYC85_031358 [Camellia sinensis]	Camellia sinensis	77.8	77.8	8%	1e-12	54.02%	105	KAF5930485.1
<input checked="" type="checkbox"/>	Poly [ADP-ribose] polymerase 1 [Camellia lanceoleosa]	Camellia lanceoleosa	77.0	77.0	8%	1e-11	52.87%	163	KAI8028661.1
<input checked="" type="checkbox"/>	Poly [ADP-ribose] polymerase 1 [Platanthera quangdongensis]	Platanthera quangdongensis	70.9	70.9	10%	3e-10	43.93%	107	KAK8955608.1
<input checked="" type="checkbox"/>	unnamed protein product [Camellia sinensis]	Camellia sinensis	75.1	75.1	8%	3e-09	50.56%	757	CAL5427461.1

This is the alignment detail for the top hit (Ipomoea batatas) with 51% identity.

Download

GenPept Graphics

Next Previous Descriptions

hypothetical protein lba_chr07aCG1000 [Ipomoea batatas]

Sequence ID: [GMD12175.1](#) Length: 120 Number of Matches: 1

[See 4 more title\(s\)](#)
[See all Identical Proteins \(IPG\)](#)

Range 1: 6 to 118 GenPept Graphics

Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
91.7 bits(226)	2e-17	Compositional matrix adjust.	59/116(51%)	69/116(59%)	8/116(6%)

Query 312

QDQSMIYEIVVLMGCSLEHWSARSVLVPCITLEASIVAMAMQYHGVGLTPQPILYA*R
 371

Sbjct 6

QDQNLIIYEIDVLMGCFLEHSLYVHFVLVTFATQKASINAMAIYRHGASVHTPPLSLNVLM
 65

Query 372

KNGKSLKKQVTGIF*SGSSSRKRQRNLTEFFL-----LHLPINLLVVLVCSHIHQMV
 422

Sbjct 66

GSGKSLKRQGMSIYSSGISHKRQRNQEQEYCLHLPAPHLPAKLLTSLVS---HQQV
 118

Related Information

[Identical Proteins](#) - Identical proteins to GMD12175.1

It is also to be noted that there are hits that match to the original query of PARP1 (specifically hits 4 and 5) with less than 100% identity which indicates that the novel gene is homologous to the original query.

Q5) Generate a multiple sequence alignment with your novel protein, your original query protein, and a group of other members of this family from different species.

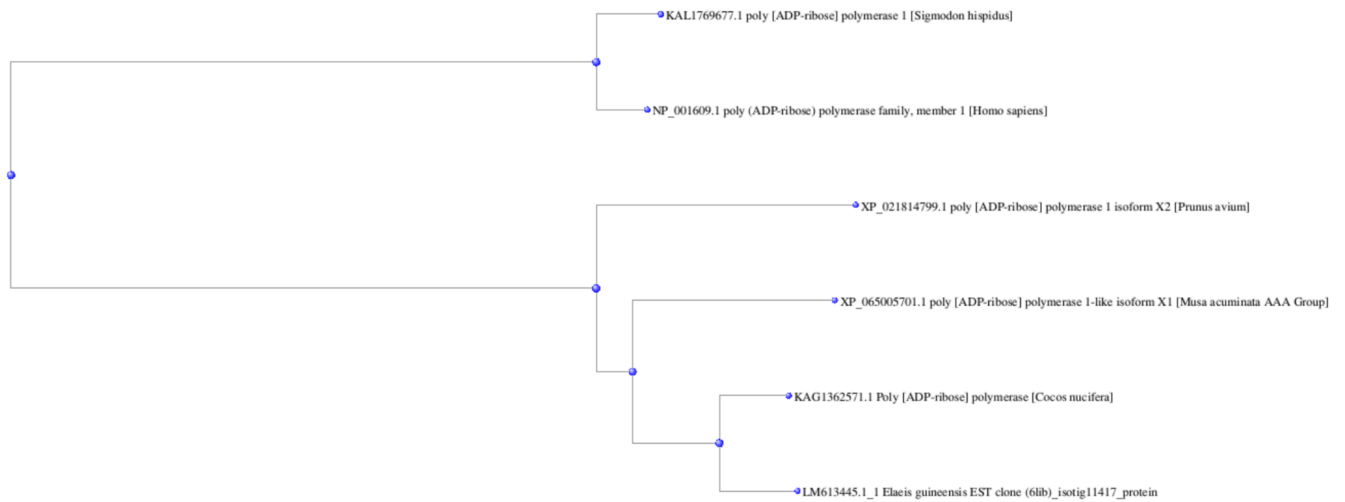
Homo_sapiens	-MAESSDKLYRVEYAKSGRASCKKCSSEIPKDSLRMAIMVQSPMFDGKVPWHYHFSCFWK	59
Elaeis_guineensis	-AMATPPKPWKAHEYAKSGRSSCKSKNSIGKDELRLGKMVTATQFDGFMPMWNHAGCIFK	59
Sigmodon_hispidus	-MAEASERLYRVEYAKSGRASCKKCSSEIPKDSLRMAIMVQSPMFDGKVPWHYHFTCFWK	59
Prunus_avium	MANPQPPKPWKVEYAKSSRSSCKTCKSPIEKEKLRLGKMVTATQFDGFMPMWNHADCIMK	60
Musa_acuminata	--MANPPKPWKAHEYAKSGRSSCKTCKSPIDRDQLRLGKMVAATQFDGYMPMWNHAGCIFK	58
Cocos_nucifera	--MANPPKPWKAHEYAKSGRSSCKSKNSIDKDQLRLGKMVAATQFDGFMPMWNHAGCIFK	58
	: :.:*****.*:***.*.. * :.:**:. ** : *** : * * * *: *	
Homo_sapiens	VGHSIRHPDVEVDGFSELRWDDQQKVKKTAEAGGVTKGQDYGIGSKAEKTLGDFAAEYAK	119
Elaeis_guineensis	KGNQIKSVD-DVEGIDLLRWEDQQKIRKYTEGGSVSTT-----AVSNDECTIEVSQ	109
Sigmodon_hispidus	VGH CIRQPDTEVDGFSELRWDDQQKVKKTAEAGGVTKGQNGSGGKAEKTLGDFAAEYAK	119
Prunus_avium	KAKQIKSTD-DVEGLELLRWEDQKEIRNYVQSGGPPDTITT----ATTSKVSSGIEVSP	115
Musa_acuminata	KQNQIKSLD-DVEGIDLLRWEDQKSIRKYVEDGSSTST-----TVANSECAIEVSQ	108
Cocos_nucifera	KGNQIKSVD-DVEGIDLLRWEDQQKIRKYVEGGSVSTT-----AVSNDGCAMEVSQ	108
	: *: * :*:*. :*:***:~::~ :. *. . : *	
Homo_sapiens	SNRSTCKGCMKIEKGQVRLSKKMVDPEKPQLGMIDRWYHPGCFVKNREELGFRPEYSAS	179
Elaeis_guineensis	TSRAACRRCNQKIMKGMVRVSTKPEGQGA---RGLAWHHVNCFIEMS-----PSTIIE	159
Sigmodon_hispidus	SNRSTCKGCMKIEKGQMRSLSKMMLDPEKPQLGMIDRWYHPACFVNNREELGFRPEYSAS	179
Prunus_avium	TSRATCKSCSQKILKAEVRISTKPEGQGP---RGLAWHHANCFMELS-----PSTEVE	165
Musa_acuminata	TSRATCRHCSQKITKTGTVRVSTKAEGQGA---RGISWHHVNCFTTMS-----PSTSLE	158
Cocos_nucifera	TSRAACRRCNQKIMKGMVRVSTKAEGQGT---RGLAWHHVNCFIDMS-----PSTVLE	158
	:.*:~::~ * :** *. :*:~::~ . *:~::~ ** *. .	
Homo_sapiens	QLKGFSLLATEDKEALKKQLPGVKSEGKRKGDE-----VDGVDE-	218
Elaeis_guineensis	KMSGWDSLSPEDKVTVTALVKKDKSNKNTAQE-----QQLSKGTRKKVGSSEDH	208
Sigmodon_hispidus	QLKGFSLLSAEDKEALKKQLPGVKSEGKRKGDE-----VDGADE-	218
Prunus_avium	KLSGWETLPVADQAAVRALVKKVPSNARGKKTEEQEDKEFLQQATSNTSTKRRKDSGGDQ	225
Musa_acuminata	KISGWDSLSPODKESLSAFSRKDTSKKTE--D-----QVTSRSAKRKAVGSDEQ	205
Cocos_nucifera	KISGWDNLSPEDKATITALGKK-----AVAQE-----KHATKGTTRKKVGSSEDH	202
	:~::~.* ~::~ * ~::~ : . :	
Homo_sapiens	---VAK-----KKSKEKDKDSKLEKALKAQNDLIWNIDELKKVCSTN	259
Elaeis_guineensis	HSKVPKSDENDSAGGASSKENPAESGNAYSSSIELEKKLEEQSKALWEIKDELKKHVMTA	268
Sigmodon_hispidus	---VAR-----KKSKEKDKDSKLEKALKAQNDLIWNIDELKKACSTN	259
Prunus_avium	KSKVARSEGDVSTSG-----DVSVRDATDLESKLEAQSKELWALKDDLKKHVTTA	275
Musa_acuminata	KTKVSKSEKRN SAGKSSTNGSKDEPNHGDFSTIGLEKKLEEQSKLLWDIKDLKIHVTTA	265
Cocos_nucifera	NSKVPKSNENDSAGGALGKGSAAESGNANSSSMELEKNLEKQSKALWDIKDELKKHVTTA	262
	* : . **.*:~::~.*:~::~***~::~ *	

Homo_sapiens	DLKELLIFNKQQVPSGESAILDR-----VADGMVFGALL	293
Elaeis_guineensis	ELREMLEANRQDSAGSEYDLRDR-----CADGMLFGALG	302
Sigmodon_hispidus	DLKELLIFNQQQVPSGESAILDR-----VADGMAFGALL	293
Prunus_avium	ELREMLEANVQDSTGSELDRER-----CADGMMFGALS	309
Musa_acuminata	ELREMLEANGQDSTGSEYDLRDRWLLAACTLLGLRIRPLRSLSLSPHGADGMLFGALG	325
Cocos_nucifera	ELREMLEANGQDSAGSEYDLRDR-----CADGMLFGALG :*:*:* * *: ..* : :*	296
Homo_sapiens	PCEECSGQLVFKSDAYYCTGDVTAWTKCMVKQTTPNRKE--WVTPKEFREI---SYLKKL	348
Elaeis_guineensis	KCPICSGSLHYSGGQYRCHGYVSAWSRCSYSTTNPLRLKEKWVPPEETSNRYLLKWFKSQ	362
Sigmodon_hispidus	PCKECSGQLVFKSDAYYCTGDVTAWTKCMVKQTTPSRKE--WVTPKEFREI---SYLKKL	348
Prunus_avium	RCPICSGFLRYSGGMYRCHGYISEWSKCSYSTEEPRLKWKWKVPEDTDNLYLNKWFKSQ	369
Musa_acuminata	TCPICSGSLCYSGGQYRCHGYLSAWSKCSYTTEPVLRLAKWKIPKETSNGYLIKWFKSQ	385
Cocos_nucifera	KCPICSGSLHYSGGQYRCHGYVSAWSKCSYSTDPVRLKEKWKIPGETSNKYLLKWFKSQ * *** * :... * * * ::*:~* .~* * ~* : * * : : ~::~.	356
Homo_sapiens	KVKKQDRIFPPETSASVAATPPPSTASAPAAVNSSASADKPLSNMKILTGLKLSRNKDEV	408
Elaeis_guineensis	KAKKPNNRVLPSPSSNKSSCS-----AG-MQSHPSNGDKLESKVAVVGEPQKA-NEW	412
Sigmodon_hispidus	KVKKQDRIFPPETSAPAPLAPPSPSVTSAPAAVKASAPADKPLSNMKILTGLKLSQNKDEA	408
Prunus_avium	RVEKPMRIILPSPTPNKPSPGS-----QAFNGQSQSSNSASLADLKVAFRGLPKESVEEW	422
Musa_acuminata	KANKPGRVLPSPSTSKSSGR-----HAT-NLSQPSENDEKLENLKVAIAGGSADFADL	437
Cocos_nucifera	KAKKPNNRVLPNNK---SSSS-----TL-VQSQPSNGDNLENLRVAIVGESRKAIEDW ::~* ~::~* . : . * ~::: * . :	404
Homo_sapiens	KAMIEKLGGKLTGTANKASLCISTKKEVEKMNMKKMEEVKEANIRVVSEDFLQDVSASTKS	468
Elaeis_guineensis	KHRFEEAGGKIHAHIKKDTNCLVLIGEMVDKDSEIRKARRMKIPIVREDYLHECISKQKK	472
Sigmodon_hispidus	KATIEKLGGKLTGSASKASLCISTKKEVEKMSSKKMEEVRAANVRVVCEDFLQDVSASTKS	468
Prunus_avium	SRKIEGVAGLVHSKIKKDTNCLLVSGALDDKDAEMRKARRMKLPIVREDYLVDCFVKQKK	482
Musa_acuminata	KTKLEAAGVKFHMKIAKDTSCLIWVGVDNDNSEMRKARRMKLPIVRVDYLQEEMRKQKK	497
Cocos_nucifera	KHKFEEAGGKIHAHIKKDTNCLVLIGEMDADDSEIKKARRMKIPIMREDYLHECIRKQKK . ~* . . . * ~* : : . ~:::~: ~: ~: ~*~* : . *	464
Homo_sapiens	LQELFLAHILSPWGAEVKAEPVEVVAPRGKSGAALSKKSKGQVKEEGINKSEKRMTLTK	528
Elaeis_guineensis	IPFDLY-----KVEAA-----SETSRGGMVTVKVK	497
Sigmodon_hispidus	LQELLSAHSLSWWGAEVKVEPVETVPKGKSAA-PSKSKGPVKEEGVNKSEKRMTLTK	527
Prunus_avium	LPFDLY-----KVEEV-----GQT--SSMVTVKVK	505
Musa_acuminata	LPFDLY-----KIENF-----AETSRSGIVTVRVK	522
Cocos_nucifera	IPFDLY-----KVEAA-----FETSRSDMVTVKVK : : :* .:~:~*	489

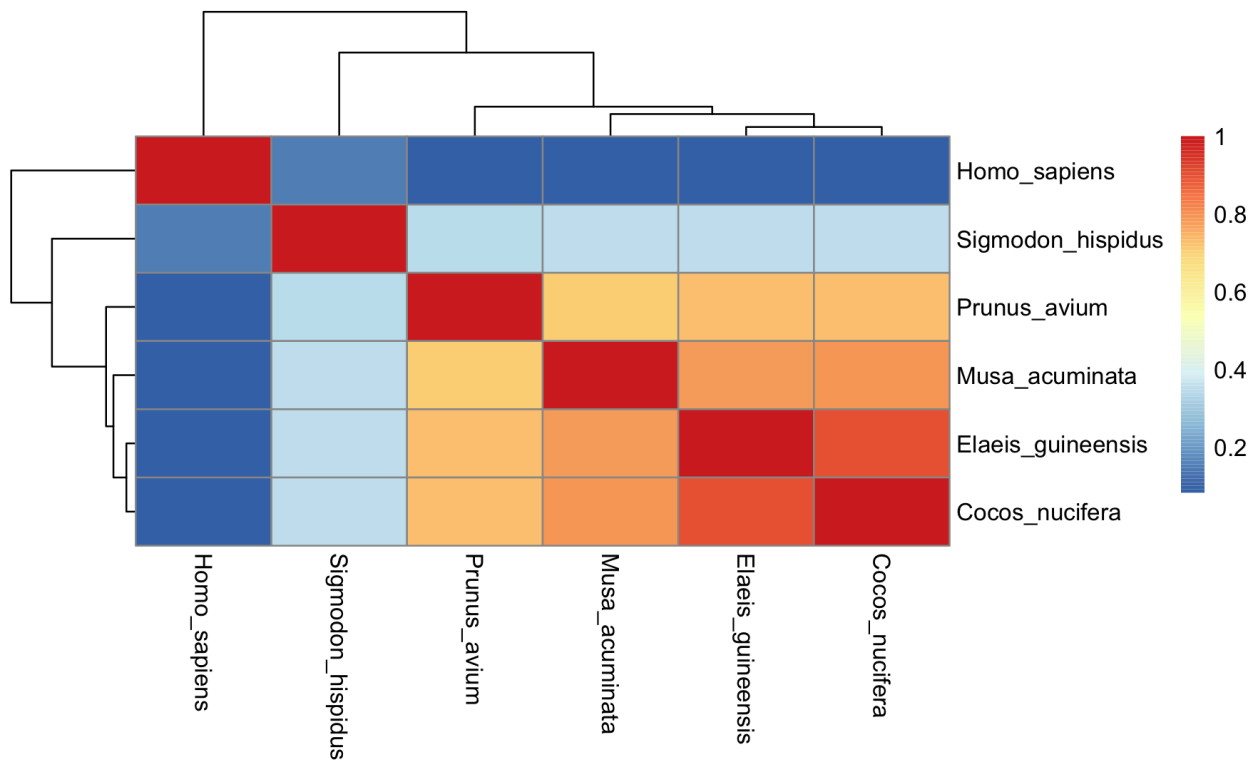
Homo_sapiens	GGAAVDPDSGLEHSAHVLEKGGKVFSATLGLVDIVKGTNSYYKLQLEDDKENRYWIFRS	588
Elaeis_guineensis	GRSAVHEASGLQDVGHILEDGNSIYNTTLNMSDLSTGINSYYILQIIQEDKGSACFVFRK	557
Sigmodon_hispidus	GGAAVDPDSGLEHSAHVLEKGGKVFSATLGLVDIVKGTNSYYKLQLEDDKESRYWIFRS	587
Prunus_avium	GRSAVHESSGLQDTHILEDGKSIYNTTLSMSDLSTGVNSYYILQIIQDDKSSDCYVFRK	565
Musa_acuminata	GRSAVHEASGLQDTGHILEDGKSIYNTTLNMSDLSTGINSYYILQIIQEDKSGCYVFRK	582
Cocos_nucifera	GRSAVHEASGLQDAGHILEDGSSIYNTTLNMSDLSTGINSYYILQIIQEDKGSDCYVFRK	549
	* :**. ***. *:**. * .:..:*. : * :. * **** *:..:*. . :***.	
Homo_sapiens	WGRVGTV-IGSNKLEQMPSKEDAIEQFMKLYEEKTGNAWH----SKNFTKYPKKFYPLEI	643
Elaeis_guineensis	WGRVGNDKIGGTKLEEM-SKSDAIQEFKRLFLEKTGNPWEAWEQKKNFEKQPGRRFFPLDI	616
Sigmodon_hispidus	WGRVGTV-IGSNKLEQMPSKEDAIEHFMKLYEEKTGNAWH----SKNFTKYPKKFYPLEI	642
Prunus_avium	WGRVGNDKIGGNKLEEM-SKSDAICEFKRLFLEKTGNSWEAWEQKQNFQKQPGRRFFPLDI	624
Musa_acuminata	WGRVGNNKIGGTKLDGM-SKSDAIQEFKRLFLEKTGNPWEAWEQKRNFEKQPGRRFFPLDI	641
Cocos_nucifera	WGRVGNDKIGGTKLEEM-SKSDAIQEFKRLFLEKTGNQWEAWEQKKNFEKQPGRRFFPLDI	608
	****. **..*: * **.*** . *: : ***** *. .:*** * *:***:*	
Homo_sapiens	DYGQDEEAVKKLTVNPGTKSKLPPVQDLIKMIFDVESMKKAMVEYEIDLQKMPLGKLSK	703
Elaeis_guineensis	DYGVKQVSKKKDS--ANIKSLLAPQLIELMKMLFDVETYRAAMLEFEINMSEMPGKLSK	674
Sigmodon_hispidus	DYGQDEEAVKKLTVKPGTKSKLPPVQELVGMVFDVESMKKALVEYEIDLQKMPLGKLSK	702
Prunus_avium	DYGVNKQVSKKNQ--NNAASKLAPPLAELMKMLFNVETYRAAMMEFEINMSEMPGKLSK	682
Musa_acuminata	DYGIKQVPKKKDL--TNKKSQ LAPQLMDLMKMLFNVETYRAAMLEFEINMSEMPGKLT	699
Cocos_nucifera	DYGVKQVSKKVES--ANTKSLLAPQLIELMKMLFDVETYRAAMLEFEINMSEMPGKLSK	666
	*** .: * . * * : :*: *:***: : *:***:..:*****:*	
Homo_sapiens	RQIQAAYSILSEVQQAVSQG-----SSDSQILDLSNRFYTLIPHDFGMKKPPLNNA	757
Elaeis_guineensis	MNIQKGFEALTEIQNLNNNAKHDPVVKESLIVDASNRFFTLIPSI---HPHVIRDEDD	730
Sigmodon_hispidus	RQIQAAYSILSEVQQAVSQG-----SSDSQILDLSNRFYTLIPHDFGMKKPPLNNA	756
Prunus_avium	RNIQKGFEALTEIQNLNLS-NGHAPSMKESLIVDASNRFFTVIPSI---HPRVIRDEDD	737
Musa_acuminata	KNIQKGFEALTEIQNLVCN-SDYDPAIKESLIIDASNRFFTLIPSI---HPHVIRHEDD	754
Cocos_nucifera	TNIQKGFEALTEIQNLNNNAHDPVVKESLIVDASNRFFTLIPSI---HPHVIRDEDD	722
	:** .. *:***: : . .: *:* *****:*** :* :... *	
Homo_sapiens	VQAKVEMLDNLLDIEVAYSLLRGSDSSKDPIDVNYEKLKTDIKVVDRDSEEAIEIRKY	817
Elaeis_guineensis	FKAKVKMLEALQDIEIASRLVG--FDSGNDESLDVYKKLQCDITPLPHDSEDYQLVEKY	788
Sigmodon_hispidus	VQAKVEMLDNLLDIEVAYSLLRGSDSSKDPIDVNYEKLKTDIKVVDRDSEEAIEIRKY	816
Prunus_avium	FKSKVKMLEALQDIEIASRLVG--FDADTDDSLDEKYRKLRCIDIPHDSDEFQLIKKY	795
Musa_acuminata	VKAKVKMLEALQDIEIASRLVC--FDGDDDESLLDKYKKLRCIDITPLLHDSSEDYQLVEKY	812
Cocos_nucifera	FKAKVKMLEALQDIEIASRLVG--FDSGNDESLDEKYKKLQCDITPLPNDSDEFKLVEKY	780
	.:***:***: * ***:* *: * . .: :* :*.***: ** :.***: :...**	

Homo_sapiens	VKNTHATTHSAYDLEVIDIFKIEREGECQRYKPFKQ-LHNRRLLWHGSRTTNFAGILSQG	876
Elaeis_guineensis	LLNTHAPTHKEWSLEEEVFALEREGEYDKYTPYRDKLQNKMLLWHGSRLTNFVGIIISQG	848
Sigmodon_hispidus	VKNTHATTHNAYDLEVIDIFKIEREGESQRYKPFKQ-LHNRRLLWHGSRTTNFAGILSQG	875
Prunus_avium	LLTHAPHTDWSLEEEVFALEREGEFDKFAPYRKKLNNRMLLWHGSRTTNFVGILSQG	855
Musa_acuminata	LLNTHAPTHKDWTLLEEEVFALEREGEFDKFAPRRDTLQNKMLLWHGSRLTNFVGILSQG	872
Cocos_nucifera	LLNTHAPTHKDWSELEEEVFALEREGEYDKYAPYRDKLYNKMLLWHGSRLTNFVGIIISQG	840
	: .*** **. : **: :.* :***** :: * :. * *: ***** **.**:**	
Homo_sapiens	LRIAPPEAPVTGYMFGKGIYFADMVSKSANYYHTSQGDPIGLILLGEVALGNMYELKHAS	936
Elaeis_guineensis	LRIAPPEAPATGYMFGKGIYFADLVSKSAQYCYVDKKNPVGLMMLSEVALGEIYELKKAT	908
Sigmodon_hispidus	LRIAPPEAPVTGYMFGKGIYFADMVSKSANYCHTSQGDPIGLILLGEVALGNMYELKHAS	935
Prunus_avium	LRIAPPEAPATGYMFGKGIYFADLVSKSAQYCYTDKKNPVGLMMLSEVALGEVHELKKAT	915
Musa_acuminata	LRIAPPEAPATGYMFGKGIYFADLVSKSAQYCYVDKKDPVGLMMLSEVALGEIYELKKAT	932
Cocos_nucifera	LRIAPPEAPATGYMFGKGIYFADLVSKSAQYCYADKKNPVGLMMLSEVALGEVYELKKAT	900
	*****.******:*****:* :..: :*:*.**.******:.**:**:	
Homo_sapiens	HISRLPKGKHSVKGLGKTTPDPSANISL-DGVDVPLGTGISSGVIDTSLLYNEYIVYDIA	995
Elaeis_guineensis	YMEKPPKGKHSTKGLGKTVPLESDFKVQDQVVVPCGKVPVSSIRASELLYNEYIVYDTA	968
Sigmodon_hispidus	HISKLPKGKHSVKGLGKTTPDPSASITL-EGVEVPLGTGVP SGVNDTCLLYNEYIVYDIA	994
Prunus_avium	YMDKPPKGKHSTKGLGKKIPQSEYVWKDDVIVPCGKVPVPSNIKASELMYNEYIVYDKA	975
Musa_acuminata	YMDKPPKGKLSTKGLGKTVPLESEHVWKDEVVVPCGRVPVSSVRASELLYNEYIVYDTA	992
Cocos_nucifera	YMEKPPKGKHSTKGLGKTVPLESEFVKWRDQVII PCGKVPVSSIRASELLYNEYIVYNTA	960
	:..: **** *.*****. * * :. : * :* * :*.: :*:*:*****: *	
Homo_sapiens	QVNLKYLLKLKFNFKTSLW 1014	
Elaeis_guineensis	QVKLQFLLKVRFHHR*KV 986	
Sigmodon_hispidus	QVNLKYLLKLKFNFKTSLW 1013	
Prunus_avium	QVKMQFLLKVRFHHR--- 991	
Musa_acuminata	QVKMQFLLKVRFHHR--- 1008	
Cocos_nucifera	QVKLQFLLKVRFHHR--- 976	
	:::*:*.*. *	

Q6) Create a phylogenetic tree, using either a parsimony or distance-based approach.



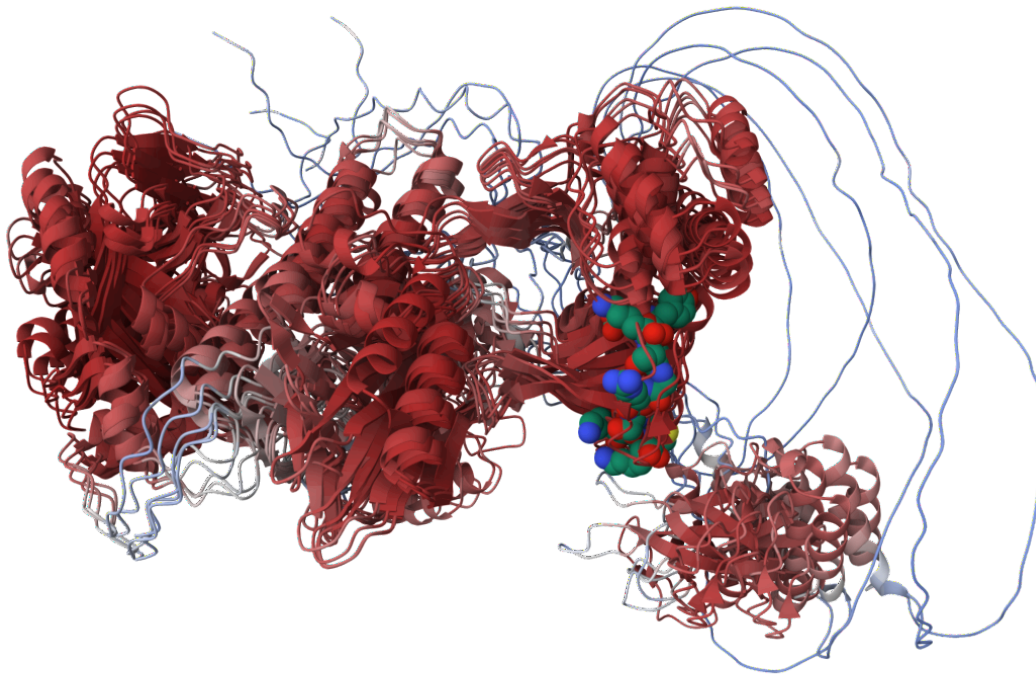
Q7) Generate a sequence identity based heatmap of your aligned sequences using R.



Q8) Using R/Bio3D (or an online blast server if you prefer), search the main protein structure database for the most similar atomic resolution structures to your aligned sequences.

ID	Experimental Technique	Resolution	Source	E-value	Sequence Identity
4OPX	X-ray	3.314	Homo Sapiens	4.25e-100	46.721
1EFY	X-ray	2.2	Gallus gallus	5.38e-76	48.857
4HHX	X-ray	1.88	Vibrio cholerae	9.42e-57	45.882

Q9) Using AlphaFold notebook, generate a structural model using the default parameters for your novel protein sequence.



Q10) Perform a “Target” search of ChEMBL (<https://www.ebi.ac.uk/chembl/>) with your novel sequence. Are there any Target Associated Assays and ligand efficiency data reported that may be useful starting points for exploring potential inhibition of your novel protein? If there are no assays listed here simply list “non available as of [date]”.

- ChEMBL lists 9 binding assays that are detailed below.
 - [CHEMBL5039976](#): Protac activity at VHL/PARP-1 in human NCI-H1299 cells assessed as induction of EGFR degradation at 0.47 uM incubated for 36 hrs by Western blot analysis relative to control
 - [CHEMBL4702093](#): Protac activity at VHL/PARP1 in human SW620 cells assessed as reduction in PARP1 protein level at 0.01 to 10 uM incubated for 12 hrs by western blot analysis
 - [CHEMBL5039975](#): Protac activity at VHL/PARP-1 in human NCI-H1299 cells assessed as induction of PARP degradation at up to 15 uM incubated for 36 hrs by Western blot analysis
 - [CHEMBL4702091](#): Protac activity at VHL/PARP1 in human MDA-MB-436 cells assessed as reduction in PARP1 protein level at 1 uM incubated for 24 hrs by western blot analysis
 - [CHEMBL4702092](#): Protac activity at VHL/PARP1 in human MDA-MB-436 cells assessed as reduction in PARP1 protein level at 0.01 to 10 uM incubated for 12 hrs by western blot analysis
 - [CHEMBL4702094](#): Protac activity at VHL/PARP1 in human MDA-MB-436 cells assessed as PARP1 protein degradation at 1 to 10 nM incubated for 24 hrs by western blot analysis
 - [CHEMBL4702095](#): Protac activity at VHL/PARP1 in human Capan1 cells assessed as PARP1 protein degradation at 1 to 10 nM incubated for 24 hrs by western blot analysis
 - [CHEMBL5039981](#): Protac activity at VHL/PARP-1 in human NCI-H1299 cells assessed as induction of PARP degradation at 4 uM incubated for 6 hrs by Western blot analysis
 - [CHEMBL5039978](#): Protac activity at VHL/PARP-1 in human A-431 cells assessed as induction of PARP-1 degradation at 0.8 to 20 uM incubated for 36 hrs by Western blot analysis
- There is no ligand efficiency data available as of March 8, 2025.