Jimmi Nguyen A15976796

BIMM 143

Professor Grant

February 6th, 2023

Find a Gene Assignment Part 1

[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: Tumor protein p53

Accession: NP\_000537

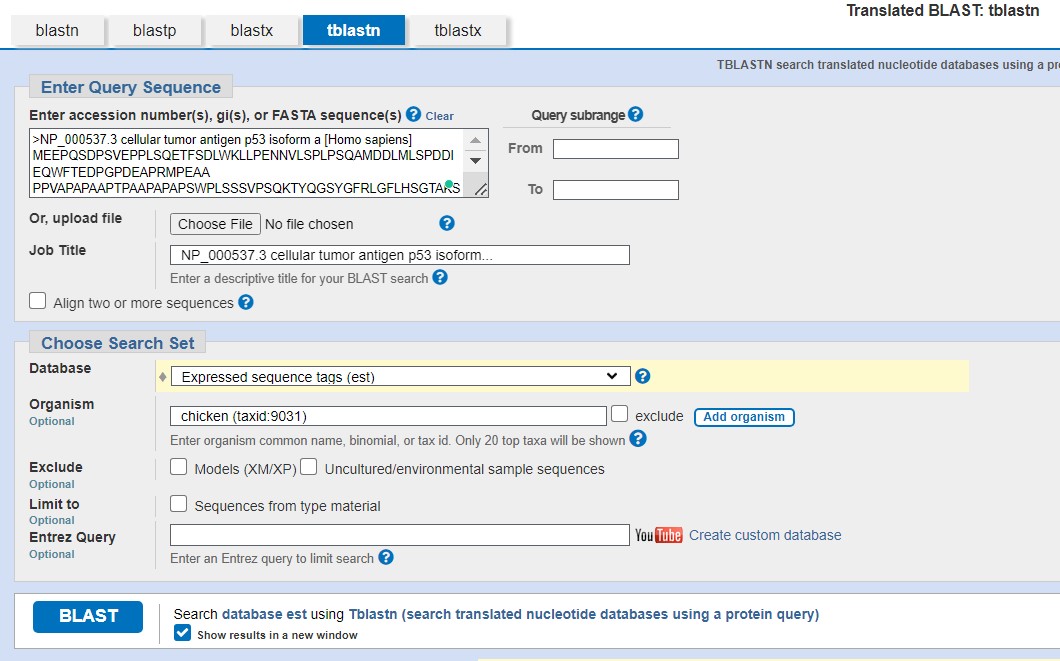
Species: Homo Sapiens

[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).

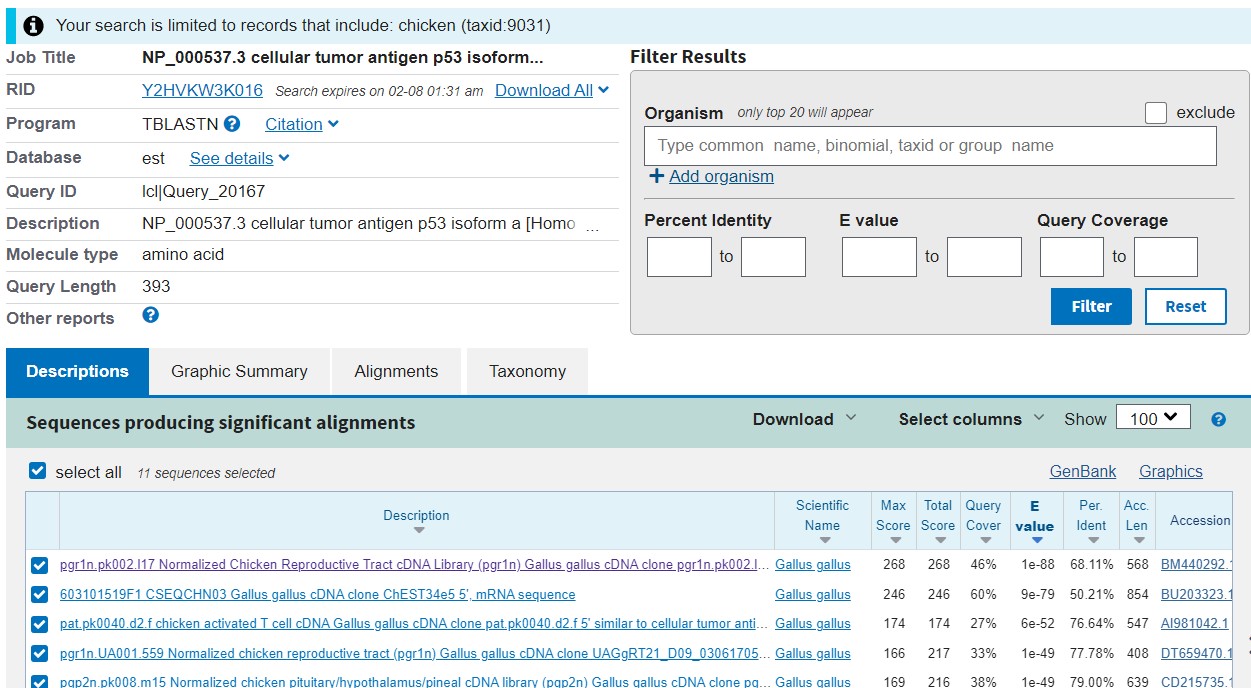
Method: TBLASTN search against chicken ESTs

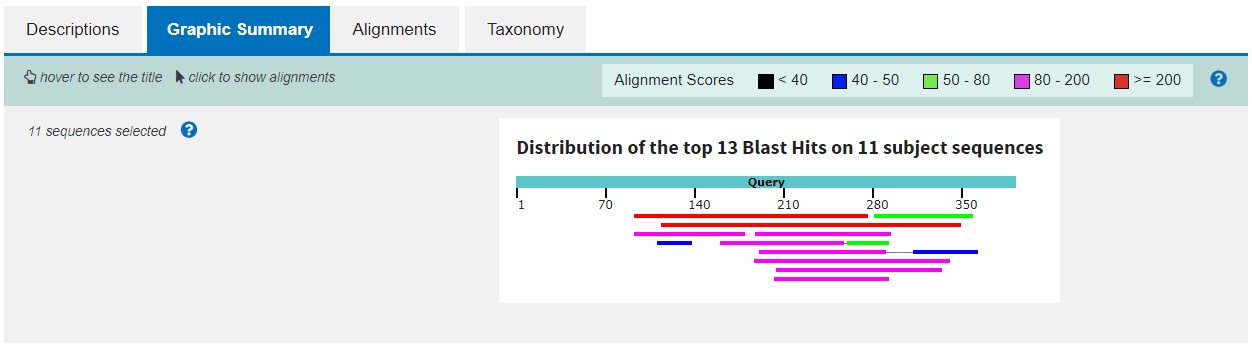
Database: Expressed Sequence Tags (est)

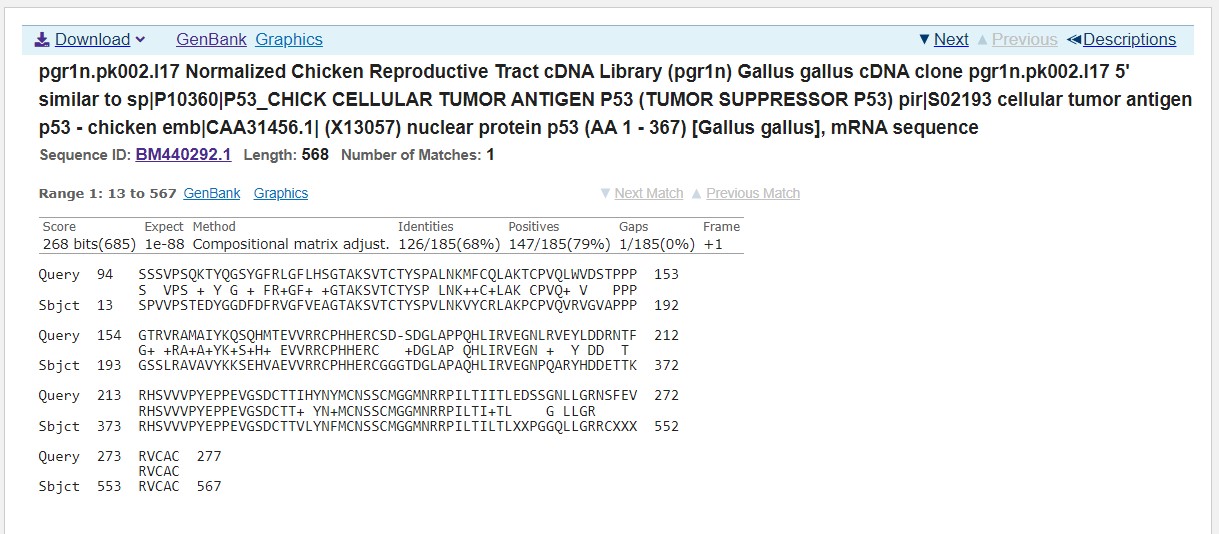
Organism: Chicken (Taxid:9031)



Chosen Match: Accession [BM440292.1](https://www.ncbi.nlm.nih.gov/nucleotide/BM440292.1?report=genbank&log$=nuclalign&blast_rank=1&RID=Y2HVKW3K016), a 568 base pair clone from Gallus gallus. See below for alignment details







Alignment Details:

pgr1n.pk002.l17 Normalized Chicken Reproductive Tract cDNA Library (pgr1n) Gallus gallus cDNA clone pgr1n.pk002.l17 5' similar to sp|P10360|P53\_CHICK CELLULAR TUMOR ANTIGEN P53 (TUMOR SUPPRESSOR P53) pir|S02193 cellular tumor antigen p53 - chicken emb|CAA31456.1| (X13057) nuclear protein p53 (AA 1 - 367) [Gallus gallus], mRNA sequence

Sequence ID: [BM440292.1](https://www.ncbi.nlm.nih.gov/nucleotide/BM440292.1?report=genbank&log$=nuclalign&blast_rank=1&RID=Y2HVKW3K016)Length: 568Number of Matches: 1

Range 1: 13 to 567[GenBank](https://www.ncbi.nlm.nih.gov/nucleotide/BM440292.1?report=genbank&log$=nuclalign&blast_rank=1&RID=Y2HVKW3K016&from=13&to=567)[Graphics](https://www.ncbi.nlm.nih.gov/nuccore/BM440292.1?report=graph&rid=Y2HVKW3K016%5bBM440292.1%5d&tracks=%5bkey:sequence_track,name:Sequence,display_name:Sequence,id:STD1,category:Sequence,annots:Sequence,ShowLabel:true%5d%5bkey:gene_model_track,CDSProductFeats:false%5d%5bkey:alignment_track,name:other%20alignments,annots:NG%20Alignments|Refseq%20Alignments|Gnomon%20Alignments|Unnamed,shown:false%5d&v=0:594&appname=ncbiblast&link_loc=fromHSP)Next MatchPrevious Match

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Alignment statistics for match #1 | | | | | | |
| Score | Expect | Method | Identities | Positives | Gaps | Frame |
| 268 bits(685) | 1e-88 | Compositional matrix adjust. | 126/185(68%) | 147/185(79%) | 1/185(0%) | +1 |

Query 94 SSSVPSQKTYQGSYGFRLGFLHSGTAKSVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPP 153

S VPS + Y G + FR+GF+ +GTAKSVTCTYSP LNK++C+LAK CPVQ+ V PPP

Sbjct 13 SPVVPSTEDYGGDFDFRVGFVEAGTAKSVTCTYSPVLNKVYCRLAKPCPVQVRVGVAPPP 192

Query 154 GTRVRAMAIYKQSQHMTEVVRRCPHHERCSD-SDGLAPPQHLIRVEGNLRVEYLDDRNTF 212

G+ +RA+A+YK+S+H+ EVVRRCPHHERC +DGLAP QHLIRVEGN + Y DD T

Sbjct 193 GSSLRAVAVYKKSEHVAEVVRRCPHHERCGGGTDGLAPAQHLIRVEGNPQARYHDDETTK 372

Query 213 RHSVVVPYEPPEVGSDCTTIHYNYMCNSSCMGGMNRRPILTIITLEDSSGNLLGRNSFEV 272

RHSVVVPYEPPEVGSDCTT+ YN+MCNSSCMGGMNRRPILTI+TL G LLGR

Sbjct 373 RHSVVVPYEPPEVGSDCTTVLYNFMCNSSCMGGMNRRPILTILTLXXPGGQLLGRRCXXX 552

Query 273 RVCAC 277

RVCAC

Sbjct 553 RVCAC 567

[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:

>BM440292.1\_1 pgr1n.pk002.l17 Normalized Chicken Reproductive Tract cDNA Library (pgr1n) Gallus gallus cDNA clone pgr1n.pk002.l17 5' similar to sp|P10360|P53\_CHICK CELLULAR TUMOR ANTIGEN P53 (TUMOR SUPPRESSOR P53) pir|S02193 cellular tumor antigen p53 - chicken emb|CAA31456.1| (X13057) nuclear protein p53 (AA 1 - 367) [Gallus gallus], mRNA sequence

RAAPSPVVPSTEDYGGDFDFRVGFVEAGTAKSVTCTYSPVLNKVYCRLAKPCPVQVRVGV

APPPGSSLRAVAVYKKSEHVAEVVRRCPHHERCGGGTDGLAPAQHLIRVEGNPQARYHDD

ETTKRHSVVVPYEPPEVGSDCTTVLYNFMCNSSCMGGMNRRPILTILTLXXPGGQLLGRR

CXXXRVCACX

Name: Cellular Tumor Antigen P53

Species: Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Archelosauria; Archosauria; Dinosauria; Saurischia; Theropoda;

Coelurosauria; Aves; Neognathae; Galloanserae; Galliformes;

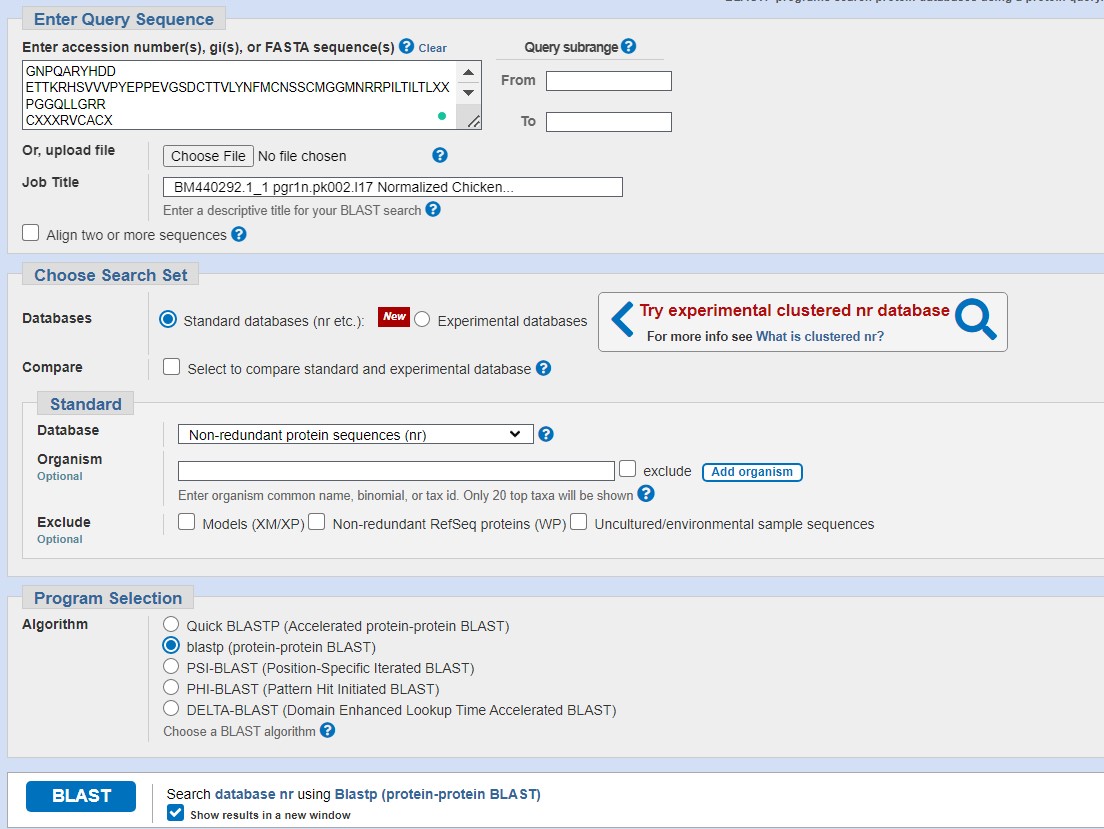
Phasianidae; Phasianinae; Gallus.

[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:

A BLASTP search against NR database yielded a top hit result to a protein from Gallus gallus. See additional screen shots below for top hits and selected alignment details:



The top result is to a protein from Gallus gallus (chicken), see second screen shot below for alignment details:

