

Exploring Sensory Evolution Using NCBI BLAST

Summary: Genome sequences and computational tools have revolutionized how scientists study evolution. It is now possible to determine evolutionary relationships by comparing gene sequences between species. In Part I of this exercise you will compare protein sequences to create a phylogenetic tree for a group of animals. In Part II, you will use BLAST, a popular online sequence database search tool, to determine when a gene family arose during animal evolution.

Evolutionary biology in the genomic age

The universal biochemistry of life suggests that all living organisms on Earth evolved from a common ancestor. Reconstructing the family tree of life is one of the most exciting problems in modern biology. It is also one of the most challenging, since the events of evolution took place in the distant past. Traditionally, scientists have pieced together the puzzle of evolution by comparing the physical characteristics of living organisms and fossilized organisms.

In recent decades, sequencing of genes and genomes from all branches of the tree of life has opened new opportunities to study evolution on the molecular level. Comparing DNA and protein sequences can refine hypotheses about evolution, correct misconceptions, and reveal the molecular mechanisms that generate evolutionary novelty. Using vast sequence databases and powerful supercomputers, evolutionary biology has become a science of “big data” where efficient algorithms and database searches are as important as fossils and comparative anatomy.

Part I: Reconstructing evolutionary relationships from sequence similarity data

How do we use DNA and protein sequences to determine the evolutionary relationships between species? When one species diverges into two, initially their genome sequences will be very similar as are the sequences of proteins encoded by their genes. Over time, the sequence of each species’ genome will change as they are exposed to different selective pressures and random mutations accumulate. A note of caution- in these exercises, we will focus on the sequence of one gene, but be aware that scientists use many different genes and proteins or the whole genomes to reconstruct evolutionary relationships. One sequence never tells the whole story.

Consider a simple example: we want to determine the evolutionary relationships between three species of small mammals. We know that all three of these mammals have a protein P, but the amino acid sequence of P in each species is slightly different. Genes in different species that evolved from a common ancestor gene are called **orthologs**. We might hypothesize that orthologous sequences are more similar in species that are closely related than in species that are more distantly related. The amino acid sequence of P from each of these animals is given below.

Sequences of Protein P:

Mouse: MGAPTAVHARFKVLIHRPAHV
 Rat: MGAATAVHARFKVLIHKPAHL
 Rabbit: MGMPTAVHAKFRVLVHRAAHL

Analyze the differences as follows: compare the sequences from mouse and rat position by position, count the number of amino acids that are mismatched, and record the number in the table below. Now compare the mouse sequence with the rabbit sequence, and finally compare the rat with the rabbit sequence. Zeros have been entered in the cells on the diagonal because there are zero differences between a sequence and itself. You will only need to fill out half the table because the second half would contain duplicate information. The cell with the highest number corresponds to the greatest genetic difference and the greatest evolutionary distance. The cell with the lowest number corresponds to the smallest genetic difference and the smallest evolutionary distance.

Table 1. Number of amino acid mismatches in alignment of protein P.

	Mouse	Rat	Rabbit
Mouse	0		
Rat	-----	0	
Rabbit	-----	-----	0

Which two animals are most closely related?
 Which two animals are most distantly related?

In the space next to the table sketch a phylogenetic tree representing the evolutionary relationships between these animals. Recall that the difference between a phylogenetic tree and a cladogram is that in a phylogeny, the lengths of the branches indicate the relative evolutionary distance between lineages. A cladogram just shows the pattern of evolutionary relationships and all the branches are the same length.

Now we will move on to a more complicated example. We have aligned the amino acid sequences of the bitter taste receptor protein TAS2R3 from five small mammals and counted the number of mismatches. Your job is to reconstruct a phylogenetic tree from the data in the table below. Here are some tips to get started:

- Find the lowest number in the table. This corresponds to the sequences with the smallest number of differences and the closest evolutionary relationship. Place these animals close together on your phylogenetic tree.
- Find the second lowest number. Do either of these animals already belong to the first group? If so, this indicates that the animal that you have not already placed should be grouped with the first two. If neither are in the first group, they should both be placed in a separate group.
- Repeat for the third lowest number.
- When you have placed all five animals check your work by looking at the highest number in the table. Do these animals represent the most distantly related animals on your phylogeny?

Animal species:

1. *Sorex araneus* (common shrew)
2. *Echinops telfairi* (lesser hedgehog tenrec)
3. *Tupaia chinensis* (tree shrew)
4. *Elephantulus edwardii* (elephant shrew)
5. *Erinaceus europaeus* (European hedgehog)

Number of mismatches in the TAS2R3 amino acid sequence from five mammal species.

	<i>S. araneus</i>	<i>E. telfairi</i>	<i>T. chinensis</i>	<i>E. edwardii</i>	<i>E. europaeus</i>
<i>S. araneus</i>	0	131	123	129	110
<i>E. telfairi</i>	-----	0	146	122	146
<i>T. chinensis</i>	-----	-----	0	131	116
<i>E. edwardii</i>	-----	-----	-----	0	136
<i>E. europaeus</i>	-----	-----	-----	-----	0

Use the space below to sketch a phylogenetic tree of the organisms in the table:

Analysis Questions:

1. Do common shrews and elephant shrews share a more recent common ancestor than common shrews and European hedgehogs? Explain your reasoning.
2. Use a textbook or the internet to look up information about the organisms you have placed on the phylogeny. Describe a possible example of convergent evolution among these organisms.
3. Why do scientists typically use many sequences to reconstruct phylogenies instead of just one sequence? List two problems with the single-sequence approach used in this exercise.
4. Go back to the sequences of protein P presented at the beginning of this exercise. Put a star underneath each amino acid that is the same between all three sequences. Which parts of the protein do you think are most important for its function?

Part II: Exploring evolutionary relationships with online sequence databases

In Part I, we presented amino acid sequences of one protein encoded by the genome of three animals. That protein is shared by these three animals because they share a common evolutionary ancestor. Since these proteins have similar sequences, they probably have similar functions.

How do we determine the degree of similarity between two sequences? We can quantify the difference between two words by first **aligning** them to maximize the number of matches, then comparing them letter by letter and counting the number of letters that are the same and the number that are different. For example, when aligned, “genetics” and “genomics” have six letters that are the same and two that are different. Similarly, two DNA or protein sequences can be aligned, and their similarity scored as in Part I.

Aligning two sequences is called a **pairwise alignment**. Pairwise alignments are useful for scoring similarity between two sequences and locating the positions where the two sequences differ. Aligning three or more sequences is called **multiple alignment**. Multiple alignments can be useful for studying the evolutionary relationships within a group of organisms.

Sequence alignments can also be used to find related sequences in a database. You are probably familiar with how Google searches the Web for a keyword and returns websites that contain a perfect match for the keyword or a word that differs by only a couple letters. An input, or query, sequence is aligned with all sequences in a database (subject sequences) and the subject sequences with the top alignment scores are reported to the user. **BLAST** (Basic Local Alignment Search Tool) is a popular program from the National Center for Biotechnology Information (NCBI) that uses sequence alignment to search a vast database of genes, genomes, or proteins to find similar sequences. In this exercise you will use BLAST to explore the evolution of type 2 taste receptor (TAS2R) proteins.

In this activity you will work with the sequence for TAS2R38 protein from human. TAS2R38 is a membrane-bound receptor protein that binds PTC, a bitter tasting compound, and triggers a cellular response that leads to the perception of bitter taste. Orthologs of this sequence are found in many animal species, as you will see when you search the NCBI database.

Your search will also return many other sequences with names like TAS2R7 and TAS2R3. These sequences are other members of the TAS2R family. There may be multiple representatives of this family in the genomes of various animal species. Gene families are believed to have evolved through duplication of whole segments of the genome. Genes within the same genome that arose through duplication are called **paralogs**. Paralogs often evolve new, but related, functions. For example, a duplicated taste receptor gene might, through a few point mutations, acquire the ability to bind to a different bitter tasting compound. Duplication followed by sequence divergence is thought to be an important evolutionary mechanism for the generation of new genes with novel functions. Paralogs and orthologs are types of **homolog**, a term that refers to genes related by descent from an ancestral sequence.

The duplication leading to the TAS2R family of taste receptors occurred in one particular branch of the animal phylogenetic tree. This means that some animals have this gene and others do not. By searching the genomes of animals from different lineages we will be able to determine which have the TAS2R taste receptor genes and which do not. From the pattern you uncover you will be able to determine at which step of animal evolution this gene family appeared.

If you have not already downloaded the human TAS2R38 protein sequence in a previous lesson, you can access it here: https://www.ncbi.nlm.nih.gov/protein/NP_789787.4?report=fasta

As you progress through this lesson, record your results on the attached Cladogram Worksheet. The instructions will lead you through the first BLAST search then leave you to complete the remaining searches yourself. When you are done, complete the analysis questions based on your results.

Instructions:

1. Navigate to the NCBI BLAST homepage: <https://blast.ncbi.nlm.nih.gov/BlastAlign.cgi>
2. Note that BLAST can accept both protein and nucleotide sequences as inputs and search both protein and nucleotide sequence databases. In this exercise we are going to search the protein database using a protein query. Click on **Protein BLAST**.

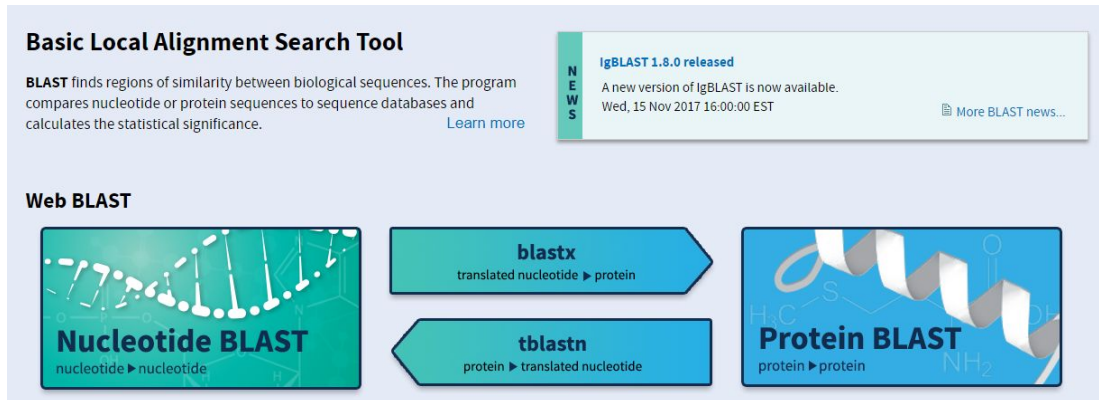
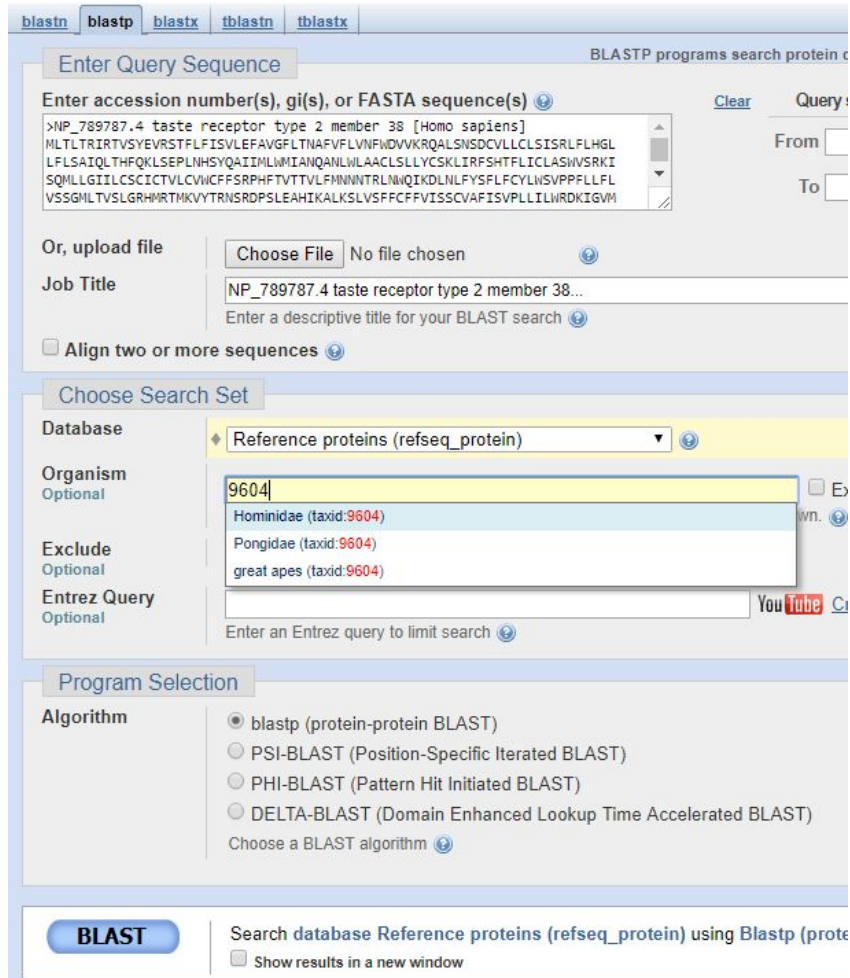


Figure 3. BLAST homepage, select protein BLAST.

3. Figure 4 shows the parameters we will configure on the BLAST search page. We have the option of limiting our search in several ways depending on what questions we are trying to answer. In our first BLAST search we are going to find orthologues of the human TAS2R38 gene in the genomes of other great apes. Copy the human TAS2R38 protein sequence into the query box, including the first line starting with a ">NP...", and paste it into the query box. The job title will be automatically filled in.
4. From the **Database** drop down menu, select **Reference proteins**. This limits our search to only high quality, non-redundant sequences.
5. In the **Organism** bar, type the **9604**, the Tax ID (taxonomic identifier) for our closest animal relatives, the Great Apes. The Tax ID is a unique number that is assigned to every species and higher group of organisms. This will allow us to search a subset of the database. You will notice that there may be multiple options with the same Tax ID. This is because one taxonomic group may be referred to by several names in database, both common and scientific. For this exercise you can select any entry corresponding to the Tax ID on the Cladogram worksheet.
6. Click **BLAST** to search the database. Be patient, sometimes this will take a few seconds.



The screenshot shows the BLAST web interface. At the top, there are tabs for 'blastn', 'blastp', 'blastx', 'tblastn', and 'tblastx'. The 'blastp' tab is selected. Below the tabs is the 'Enter Query Sequence' section, which contains a text area with a protein sequence: >NP_789787.4 taste receptor type 2 member 38 [Homo sapiens] followed by the sequence: MLTLTRIRTVSYEVRSTFLFISVLEFVAVGFLTNAFVFLVNFVNDVVKRQALSNSDCVLLCLLSISRLFLHGL LFLSAIQLTHFQKLEPLNHSYQAIIMLWMIANQANLWLAACLSSLLYCSKLIKIRFSHTFLICLASIWSRKI SQMLLGIILCSCICTVLCVWCFFSRPHFTVTTLFPHNWNTRLNQIKDLNLFYSFLFCYLWSVPPFLFL VSSGMLTVSLGRHRTMKVYTRNSRDPSPLEAHIKALKSLVSFFCFVIVSSCVAFISVPLLILWRDKIGVM. Below the text area are fields for 'Or, upload file' (Choose File), 'Job Title' (NP_789787.4 taste receptor type 2 member 38...), and a checkbox for 'Align two or more sequences'. The 'Choose Search Set' section includes a 'Database' dropdown set to 'Reference proteins (refseq_protein)', an 'Organism' dropdown with '9604' selected, and a list of organisms: Hominidae (taxid:9604), Pongidae (taxid:9604), and great apes (taxid:9604). There are also checkboxes for 'Exclude' and 'Entrez Query'. The 'Program Selection' section has radio buttons for 'blastp (protein-protein BLAST)', 'PSI-BLAST (Position-Specific Iterated BLAST)', 'PHI-BLAST (Pattern Hit Initiated BLAST)', and 'DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)'. At the bottom, there is a 'BLAST' button and a checkbox for 'Show results in a new window'.

7. When your search is complete you will see the BLAST search results window. Take a moment to scroll down the page and note the features shown in Figure 5. There are three main sections on the page:
 - **Graphic Summary**- shows visual representation of subject sequences aligning with your query sequence, color coded by their similarity score
 - **Descriptions**- lists details of each hit, including organism, name of sequence, and alignment statistics.
 - i. **Score**- a mathematical value representing the closeness of the alignment, taking into account both the number of matches and the length of the sequence. Max score and total score will only be different when there are distinct regions of alignment.
 - ii. **E Value**- Estimate of the number of the frequency of observing this score for two unrelated sequences. An E-value close to zero means a high probability the two sequences are related.
 - iii. **Identity**- percentage of amino acids that are identical between the query and the subject sequences
 - iv. **Query coverage**- percentage of the query sequence that is aligned with the subject sequence. This is visualized as the length of the colored box in the graphic summary.
 - **Alignments**- Shows pairwise alignments of query (top line) and subject (bottom line) sequences. The middle line shows the amino acids that are the same in both sequences. "+" indicates amino acids that are not identical, but that share chemical characteristics.

BLAST Results

ed to records that include: [great apes \(taxid:9604\)](#) > [Full Entrez Query](#)

[View Search Strategies](#) > [Formatting options](#) > [Download](#) [YouTube](#) [How to read this page](#) [Blast report description](#)

Job title: NP_789787.4 taste receptor type 2 member 38...

RID: 2P9SZYFC015 (Expires on 12-10 06:58 am)

Query ID: Icl|Query_189206

Description: NP_789787.4 taste receptor type 2 member 38 [Homo sapiens]

Molecule type: amino acid

Query Length: 333

Database Name: refseq_protein

Description: NCBI Protein Reference Sequences

Program: BLASTP 2.7.1+ > [Citation](#)

Other reports: > [Search Summary](#) [[Taxonomy reports](#)] [[Distance tree of results](#)] [[Multiple alignment](#)] [[MSA viewer](#)]

Graphic Summary **New** Analyze your query with [SmartBLAST](#)

Show Conserved Domains

Putative conserved domains have been detected, click on the image below for detailed results.

Distribution of the top 100 Blast-Hits on 100 subject sequences

Mouse over to see the title, click to show alignments

Color key for alignment scores

- <40
- 40-50
- 50-80
- 80-200
- >=200

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected: 0

[Alignments](#) [Download](#) [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#)

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	taste receptor type 2 member 38 [Homo sapiens]	667	667	100%	0.0	100%	NP_789787.4
<input type="checkbox"/>	taste receptor type 2 member 38 [Pan troglodytes]	657	657	100%	0.0	98%	XP_001009139.1
<input type="checkbox"/>	PREDICTED: taste receptor type 2 member 38 [Pan troglodytes]	655	655	99%	0.0	98%	XP_016800945.1
<input type="checkbox"/>	PREDICTED: taste receptor type 2 member 38 [Pan paniscus]	654	654	100%	0.0	98%	XP_003813418.1
<input type="checkbox"/>	PREDICTED: taste receptor type 2 member 38 [Gorilla gorilla gorilla]	653	653	100%	0.0	98%	XP_004046392.2
<input type="checkbox"/>	PREDICTED: LOW QUALITY PROTEIN: taste receptor type 2 member 38 [Pongo abelii]	627	627	100%	0.0	94%	XP_00377202.2
<input type="checkbox"/>	PREDICTED: taste receptor type 2 member 7 [Pan troglodytes]	134	134	95%	8e-36	32%	XP_016780285.1
<input type="checkbox"/>	taste receptor type 2 member 7 [Homo sapiens]	134	134	94%	2e-35	33%	NP_076408.1
<input type="checkbox"/>	PREDICTED: taste receptor type 2 member 7 [Pan paniscus]	134	134	95%	2e-35	32%	XP_014202577.1
<input type="checkbox"/>	PREDICTED: taste receptor type 2 member 7 [Gorilla gorilla gorilla]	133	133	94%	3e-35	32%	XP_004052767.1

[Download](#) [GenPept](#) [Graphics](#)

PREDICTED: taste receptor type 2 member 7 [Pan troglodytes]

Sequence ID: [XP_016780285.1](#) Length: 325 Number of Matches: 1

Range 1: 4 to 320 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
134 bits(338)	8e-36	Compositional matrix adjust.	104/327(32%)	162/327(49%)	19/327(5%)
Query 13	EVRSTFLFISVLEFVAVGFLTNAFVFLVNFVWVKRQALSNSDCVLLCLLSIRLFLHGLLF				72
Sbjct 4	+V++T LF++V EF+VG L NAF+ LVN D VK++ ++ D +L L+ISR+ L ++				63
Query 73	LSAIQLTHFQKLEPLNHSYQAIIMLWMIANQANLWLAACLSLLYCSKLIKIRFSHTFLICL				132
Sbjct 64	L L + + + I W + N ++W A CLS+ Y ++ F H + +				122
Query 133	ASWVSRKISQMLGIIILCSICICTVLCVNCFFSRPHFTVTVLF-----MWNTRLN--Q				185
Sbjct 123	+ R IS +LLG C VL V F S P F F T L W +				173
Query 186	IKDNLNFYSFLCYLWVPPFLFLVSSGMLTVSLGRHMRTHKVTYTRNSRDPSSLEAHIKA				243
Sbjct 174	+ LF L ++ PF + L+S +L +SL RH+R H++ RDP S EAH++A				235
Query 246	LKSLVSFFCFVVISSCVAFISVPLLLWRDKIGVMVCGIMAACPSGHAAILISGNAKLR				305
Sbjct 234	LK++SF F+ ++ + ++ V+ I PS H+ ILL GN KLR				293
Query 306	RAVMTILLWAQSSLVRA--DHKADSR 330				
Sbjct 294	A + ++ S LK R HK R				
	HASLKVIVKVMISLGRKFKQHKQIGR 320				

Figure 5. Details of BLAST results page. Scroll down through the page and identify the three major sections (graphic summary, descriptions, and alignments).

8. For simplicity, in this exercise we are going to use **% identity** as the measure of sequence similarity. To begin, let's look at the first five hits shown under **Descriptions**. You will see that the top hit is just the sequence we searched for, since it is included in the subject database. The next few entries should correspond to various species within the Great Apes. If you do not recognize the scientific names, take a moment with Google to find their common names.
9. On your Cladogram Worksheet, record the organism scientific name, protein sequence name, and % identity to the Human TAS2R38 sequence for the top hit for human and the Great Ape species with the TAS2R protein sequence most similar to the human sequence.
10. Looking at the list of sequence names under **Descriptions** you will see that some are labeled TAS2R38 while some have other sequence names. Now examine the column under **Identity**. After the first few sequences, you see an abrupt jump in the values. Why does the percent identity change so dramatically?
11. BLAST uses pairwise alignments to display results from the database search, but it can also be used to perform multiple alignments on selected sequences. Use the check boxes next to the sequence names under Descriptions to select five or six TAS2R38 sequences from different species, including human, then click the **Multiple alignment** link (Figure 6).

Use check boxes to select sequences

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input checked="" type="checkbox"/> taste receptor type 2 member 38 [Homo sapiens]	667	667	100%	0.0	100%	NP_789787.4
<input checked="" type="checkbox"/> taste receptor type 2 member 38 [Pan troglodytes]	657	657	100%	0.0	98%	NP_001009139.1
<input type="checkbox"/> PREDICTED: taste receptor type 2 member 38 [Pan troglodytes]	655	655	99%	0.0	98%	XP_016800945.1
<input checked="" type="checkbox"/> PREDICTED: taste receptor type 2 member 38 [Pan paniscus]	654	654	100%	0.0	98%	XP_003813418.1
<input checked="" type="checkbox"/> PREDICTED: taste receptor type 2 member 38 [Gorilla gorilla gorilla]	653	653	100%	0.0	98%	XP_004046392.2
<input checked="" type="checkbox"/> PREDICTED: LOW QUALITY PROTEIN: taste receptor type 2 member 38 [Pongo abelii]	627	627	100%	0.0	94%	XP_003777202.2
<input type="checkbox"/> PREDICTED: taste receptor type 2 member 7 [Pan troglodytes]	134	134	95%	8e-36	32%	XP_016780285.1
<input type="checkbox"/> taste receptor type 2 member 7 [Homo sapiens]	134	134	94%	2e-35	33%	NP_076408.1
<input type="checkbox"/> PREDICTED: taste receptor type 2 member 7 [Pan paniscus]	134	134	95%	2e-35	32%	XP_014202577.1

Figure 6. Select sequences for multiple alignment.

12. The multiple alignment page will open in a new tab. Scroll down to the **Alignments** section. From the **Conservation Settings** drop-down menu, select **Identity** (Figure 7). You will now see all identical amino acids in red and all non-identical amino acids in blue. Compare the amino acids that are in the same sequence position but vary between species. How many positions contain non-identical amino acids?

Alignments Select All Re-align Mouse over the sequence identifier for sequence title

View Format: Compact Conservation Setting: Identity

<input checked="" type="checkbox"/> NP_789787	1	MLTLTRIRTVSYEVRSTFLFISVLEFVAVGFLTNAFVFLVNFWDVVKRQALSNSDCVLLCLISIRFLHGLLFLSAIQLTH	80
<input checked="" type="checkbox"/> NP_001009139	1	MLTLTRIHTVSYEVRSTFLFISVLEFVAVGFLTNAFVFLVNFWDVVKRQPLSNSDCVLLCLISIRFLHGLLFLSAIQLTH	80
<input checked="" type="checkbox"/> XP_003813418	1	MLTLTRIHTVSYEVRSTFLFISVLEFVAVGFLTNAFVFLVNFWDVVKRQPLSNSDCVLLCLISIRFLHGLLFLSAIQLTH	80
<input checked="" type="checkbox"/> XP_004046392	1	MLTLTRIRTVSYEVRSTFLFISVLEFVAVGFLTNAFVFLVNFWDVVKRQPLSNSDCVLLCLISIRFLHGLLFLSAIQLTH	80
<input checked="" type="checkbox"/> XP_003777202	1	MLTLTRICAVSYEVRSTFLFISVLEFVAVGFLTNAFVFLVNFWDVVKRQPLSNSDCVLLCLISIRFLHGLLFLSAIQLTH	80
<input checked="" type="checkbox"/> NP_789787	81	FQKLEPLNHSYQAIIMLWMIANQANLWLAACL SLLYCSKLRFSHTFLICLASWVSRKISQMLLGIILCSCICTVLCW	160
<input checked="" type="checkbox"/> NP_001009139	81	FQKLEPLNHSYQAIIMLWMIANQANLWLAACL SLLYCSKLRFSHTFLICLASWVSRKISQMLLGIILCSCICTVLCW	160
<input checked="" type="checkbox"/> XP_003813418	81	FQKLEPLNHSYQAIIMLWMIANQANLWLAACL SLLYCSKLRFSHTFLICLASWVSRKISQMLLGIILCSCICTVLCW	160
<input checked="" type="checkbox"/> XP_004046392	81	FQKLEPLNHSYQAIIMLWMIANQANLWLAACL SLLYCSKLRFSHTFLICLASWVSRKISQMLLGIILCSCICTVLCW	160
<input checked="" type="checkbox"/> XP_003777202	81	FQKLEPLNHSYQAIIMLWMIASQANLWLAACL SLLYCSKLRFSHTFLICLASWVSRKISQMLLGIILCSCICTVLCW	160

Figure 7. Multiple alignment between TAS2R38 protein sequences from five different species. Select **Identity** from the drop-down menu.

13. Return to the BLAST results page. Un-check all check boxes except for the Human TAS2R38 sequence. Scrolling down you will notice that some of the hits are also labeled *Homo sapiens*. These correspond to other TAS2 family proteins within the human genome. Select 5-10 of these sequences and click the **Multiple alignment** link.
14. Again, select **Identity** from the drop-down menu and analyze the pattern of blue and red amino acids.
15. Now that you know how to read and interpret results from BLAST, repeat steps 5-7 & 9 substituting the Tax ID for each group of organisms on the Cladogram Worksheet, filling in the worksheet as you go. Keep all other search parameters the same. To submit another search without having to re-enter all of your parameters, click the “Edit and Resubmit” link in the upper left-hand corner of the BLAST result page (top panel, Figure 5).

Analysis Questions:

1. On the first BLAST search results page, what type of homologue were we examining when we did the multiple alignment with sequences from different species?
2. On the first BLAST search results page, what type of homologue were we examining when we did the multiple alignment with other sequences from the human genome?
3. Analyze your cladogram: which happened first in evolutionary time, the gene duplication that led to the TAS2R38 gene or the divergence of the species within the Great Apes?
4. Based on your results, which organism in the NCBI database has a TAS2R38 protein sequence that is most similar to the human sequence?
5. Do hits with the highest percent identity always have a lower E-value? Why or why not?
6. Based on your Cladogram Worksheet, on what branch of the animal phylogeny did TAS2 family proteins evolve? Justify your answer.
7. Based on your Cladogram Worksheet, which organisms have a protein sequence that is clearly identifiable as TAS2 member 38? Justify your answer.
8. Do you think sharks can taste bitter compounds? Why or why not?

9. Do you think a phylogenetic tree representing the evolution of multiple species can be reconstructed by analyzing the sequence of a single gene? Why or why not?

10. And now the burning question: do elephant shrews share a more recent common ancestor with elephants or with shrews? Use your knowledge of NCBI and BLAST, as well as the phylogeny you developed in Part I, to find the answer. Describe the reasoning you followed to answer this question.

11. Bonus: Having aligned the human TAS2R38 gene with highly dissimilar protein sequences from other organisms, why do you think we chose to do this investigation with a protein sequence instead of a nucleotide sequence?