Toxic Circumstances:

Using Bioinformatics to Understand Natural Selection

by
D. Parks Collins, Department of Biology, Mitchell Community College, Statesville, NC
Jason Macrander, Department of Biology, Florida Southern College, Lakeland, FL

Part I – One Bad Day

On September 26, 1957, Dr. Karl Schmidt woke up and reached for the thermometer. The last twelve hours had been difficult. He had experienced nausea, chills, abdominal pain, and a slight fever. The thermometer read 98.2 degrees.

He then proceeded to eat a bowl of cereal, a poached egg on toast, and applesauce. Dr. Schmidt managed to stay active most of the morning doing odd jobs around his house while enjoying a hot cup of coffee. Around 10:00 a.m., he called the museum where he was employed to let his supervisors know that he was taking the day off. He had worked long hours the previous day. Dr. Schmidt ate lunch at 12:00 p.m., but soon vomited. Shortly after, he began having trouble breathing. Around 3:00 p.m., Dr. Schmidt was pronounced dead at a nearby hospital. He was 67 years old.

The autopsy, performed the following day, showed massive amounts of internal bleeding in the small and large intestine, kidneys, lungs, eyes, heart, and brain. You can see part of the report in Figure 1.

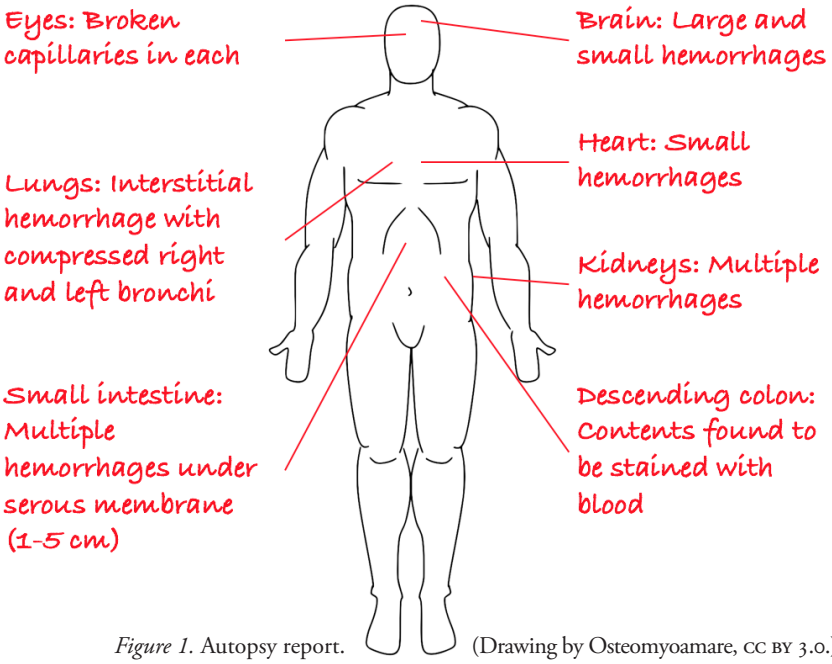

Figure 1. Autopsy report. (Drawing by Osteomyoamare, CC BY 3.0.)

Questions

1. Dr. Schmidt experienced a variety of symptoms the last 24 hours of his life. Please list and explain several potential explanations for these symptoms based on the record of events and results from the autopsy.
2. From the autopsy, what main factor seemed to have resulted in his death?
3. Dr. Schmidt's symptoms appeared over twenty-four hours. What else would you like to know to properly understand what happened to Dr. Schmidt?

Part II – The Previous Day

Born in 1890, Dr. Karl P. Schmidt was a leading herpetologist in his field. He worked at the American Museum of Natural History (New York City) and then at the Chicago Field Museum, where he became chief curator. Dr. Schmidt led expeditions all over the world, authored some 200 scientific articles and books, and described over 200 reptile species. In 1956, he was elected to the National Academy of Sciences.

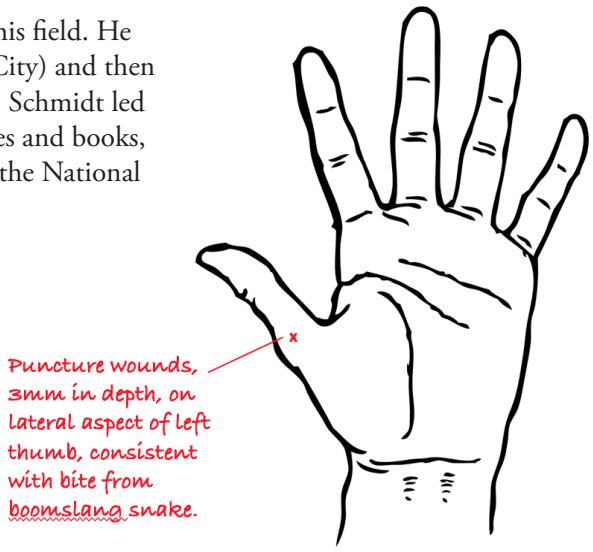

Figure 2. Puncture wounds from boomslang. (PD).

On September 25, 1957, a snake was brought in to the Chicago Field Museum from the Lincoln Park Zoo. The snake was from South Africa, and had a shape and color pattern that suggest it was a boomslang, *Dispholidus typus*. However, because it did not look quite like a normal boomslang, it was brought to Dr. Schmidt for confirmation. Robert Inger, a museum curator, brought the snake to Dr. Schmidt at 1:30 p.m. As Dr. Schmidt went to grab it, the snake turned its head and bit his left thumb (see Figure 2). No one seemed concerned about venom from the snake because the boomslang was not known to be venomous and also because this one was small and the puncture wounds were shallow. At 2:30 p.m., things seemed to change. Dr. Schmidt wrote that there was "clear evidence of envenomation" as the "area became painful to the touch."

Before you are able to characterize the venom composition, you must understand some basic genetic concepts, specifically how proteins are synthesized in the first place.

Questions

1. What is the difference between a poison and a venom? Why are some snakes considered venomous? Answers can be found either in a textbook or online.

2. Please fill out the function and location of the corresponding molecules below.

Molecule	*Function*	*Location in cell*
DNA		
mRNA		
Ribosomes		

3. When and why is DNA replicated in a particular cell?

4. What is a gene?

5. What are some differences between the processes of transcription and translation?

6. Some genes that are responsible for producing proteins required for basic cellular function are sometimes referred to as "housekeeping" genes. For example, genes that code for RNA polymerases or ribosome proteins that aid in transcription or translation are "housekeeping" genes. Can you foresee a scenario where a gene like this could change functions? How so?

7. Define gene duplication. How could the process of gene duplication make venom?

Part III – Linking Proteins to Symptoms

As a medical investigator, you will need to decide whether or not this species of snake had the proper venom composition to kill Dr. Schmidt. You have been given the responsibility of finding out what specific toxin proteins were found in the boomslang's venom and if, in fact, those compounds were likely to have caused Dr. Schmidt's immediate death.

Snake venom exhibits diversity and variation in both toxin composition and action. Therefore, these venom cocktails provide the perfect opportunity to study the relationship between natural selection and the genetic and molecular processes. When considering groups of venomous animals, some unusual things may have happened to the genes of their ancestors before they became venomous. For example, it is possible that "typical" housekeeping genes were duplicated, and the duplicate gene copies assumed toxin-like attributes. This is commonly referred to as the "birth and death" process of gene family evolution. If the duplicate copy of a gene isn't maintained, it "dies" and only the original copy of the gene remains. If the gene copy is retained and selected for potential toxic attributes, a toxin gene is born. Figure 3 shows an evolutionary timeline for the venom gene that affected Dr. Schmidt.

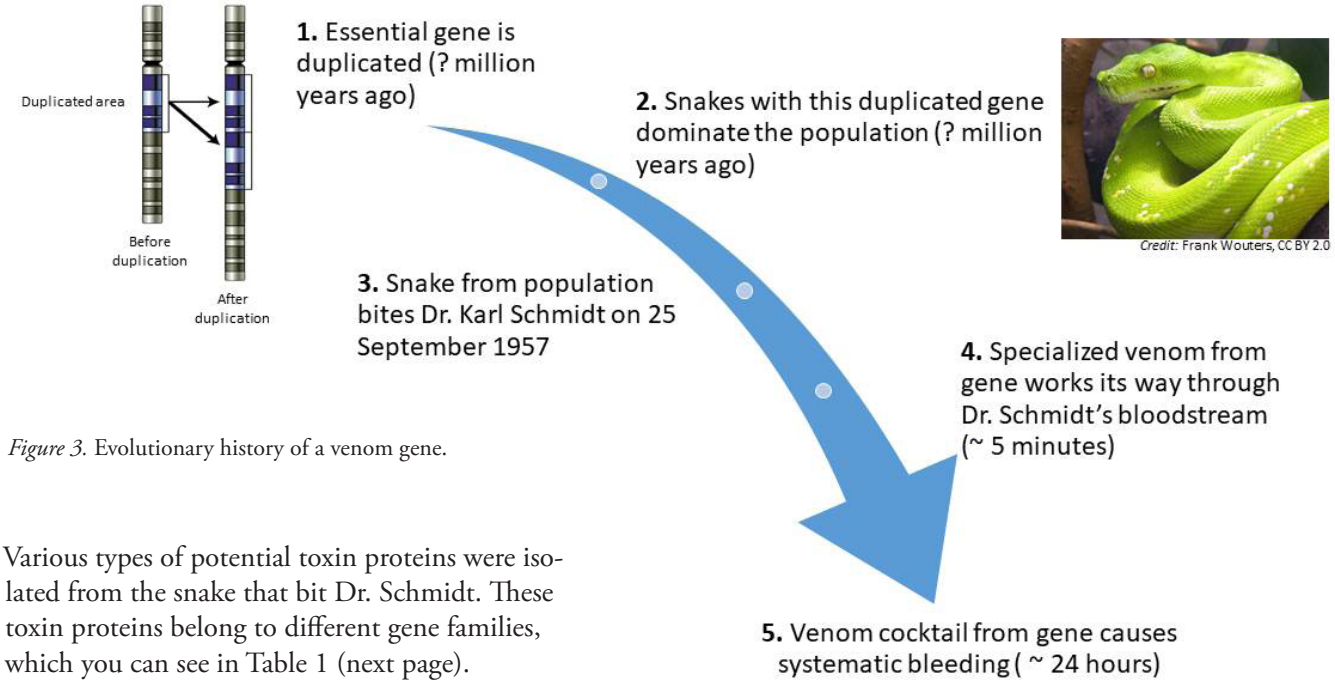

Figure 3. Evolutionary history of a venom gene.

Various types of potential toxin proteins were isolated from the snake that bit Dr. Schmidt. These toxin proteins belong to different gene families, which you can see in Table 1 (next page).

Questions

1. Not all venom genes are recruited from housekeeping gene families. Which protein families not involved in housekeeping functions could have been recruited into venom glands in the boomslang? Why do you think that? (Please refer to Table 1.)

2. Which toxin protein family from Table 1 seems to match the closest to Dr. Schmidt's symptoms? Why?

Table 1. Protein Families identified in venom from snake that bit Dr. Schmidt.

Protein Families	*Venom Target*	*Housekeeping Function*	*Proportion in the snake venom*
Three-finger toxin (3FTx)	*Nervous system:* interfere with the transmission of choline at postsynaptic sites in the peripheral and central nervous system. *Cardiovascular system:* increase or decrease heart rate, inhibit platelet aggregation, and compromise anticoagulation properties of the blood. *Neuromuscular system:* includes spontaneous contractions among muscle fibers. *Cellular homeostasis:* block calcium ion channels on the cell.	- NA -	3.8 %
Phospholipase A2 (PLA2)	*Nervous system:* presynaptic neurotoxicity. *Cardiovascular system:* intravascular hemolysis, pulmonary congestion and edema, and anticoagulation. *Neuromuscular system:* rapid necrosis of muscle fibers	Break down fatty acids.	6.4 %
Cysteine-rich secretory proteins (CRISP)	*Nervous and neuromuscular systems:* target various ion channels necessary for the transmission of neural signaling	Reproduction (mammals only).	8.2 %
Snake venom serine protease (SVSP)	*Cardiovascular system:* affect stages of blood coagulation, acting either as pro-coagulants or as anti-coagulants.	Digestion, blood clotting, fighting infections.	2.5%
C-type lectin-like (CTL)	*Cardiovascular system:* pro- or anticoagulants, or as agonists or antagonists of platelet aggregation.	Binds to mainly carbohydrates, but also proteins and lipids.	0.7 %
SV metalloproteases (SVMPs)	*Cardiovascular system:* depletes fibrinogen from the blood (preventing the coagulation of blood) and causes the rupturing of blood vessels.	Break down proteins.	77.5 %

Part IV – The Proteomics of Death

The biochemical pathways leading to venom formation in snakes as well as the mechanism of action in victims/prey are extremely complicated with many unknowns, but these systems do provide scientists the chance to study how natural selection impacts the genetic and molecular processes of venom formation (Casewell *et al.,* 2013). The majority of venoms from animals are actually made up of a cocktail of biologically active ingredients. Typically, these venoms contain a mixture of salts, neurotransmitters, amino acids, and proteins. It is usually the proteins, however, that are found in greatest abundance and are responsible for the observed symptoms.

You are a scientist exploring the cause of Dr. Karl P. Schmidt's death. You obtain some venom from the snake that bit Dr. Schmidt and send it out for protein analysis (Figure 4). Your goal is to not only identify what venom proteins there are, but also to determine which venom proteins may be responsible for the observed symptoms and ultimately Dr. Karl P. Schmidt's death.

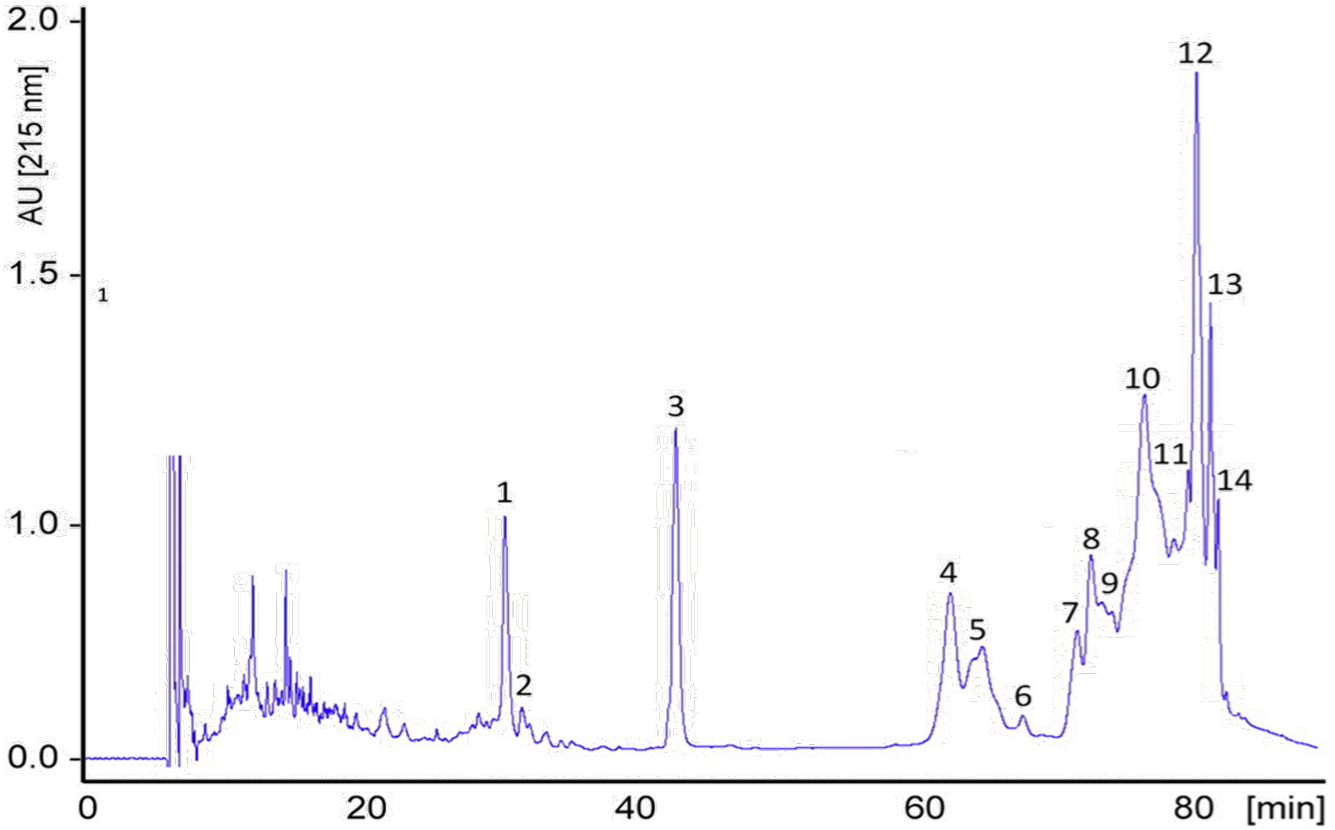

Figure 4. HPLC separation from boomslang venom (see Pla *et al.,* 2017). Numbers on each peak correspond with isolated proteins recovered from HPLC analysis. Area under the curve represents overall protein concentration.

What Is a HPLC?

High performance liquid chromatography (HPLC) is a type of chromatography that is used to separate, identify, and quantify different components of a mixture. This is done on specialized machines, which include a component used to bind different parts of a protein called a column. The HPLC machine pumps the sample mixture in a solvent (known as the mobile phase) at high pressure. Once the sample reaches the column it encounters a fixed substance (stationary phase) that sorts out pieces of the protein based on size or other characteristics determined by the column. After the protein components pass through the column they encounter a detector, which converts protein pieces into abundance and type. The more abundant proteins result in higher peaks (y-axis) and larger proteins take longer to run through (x-axis).

You will be using the website BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) to determine which gene families the venom proteins belong to. BLAST stands for basic local alignment search tool.

When your instructor indicates it is time to start the BLAST search, follow the checklist below to conduct a bioinformatic analysis on the 14 proteins that were identified by the HPLC analysis.

Be sure you start with a standard protein BLAST (blastp) when screening these toxins and check that all of the following search parameters are set:

- ☐ **Confirm default search set:** "Database: Non-redundant protein sequences (nr)." (Everything else should remain blank.)
- ☐ **Program selection:** make sure that blastp is selected.
- ☐ Be sure to enable **Show results in a new window** at the bottom of this page (Figure 5B)

Now that you are sure everything is ready you can begin by uploading the Boomslang_Protein.fasta file (provided by your instructor) using the **[Choose File]** button just below the entry box (Figure 5B). Depending on the load of the website the wait time for the program to run your search is unpredictable.

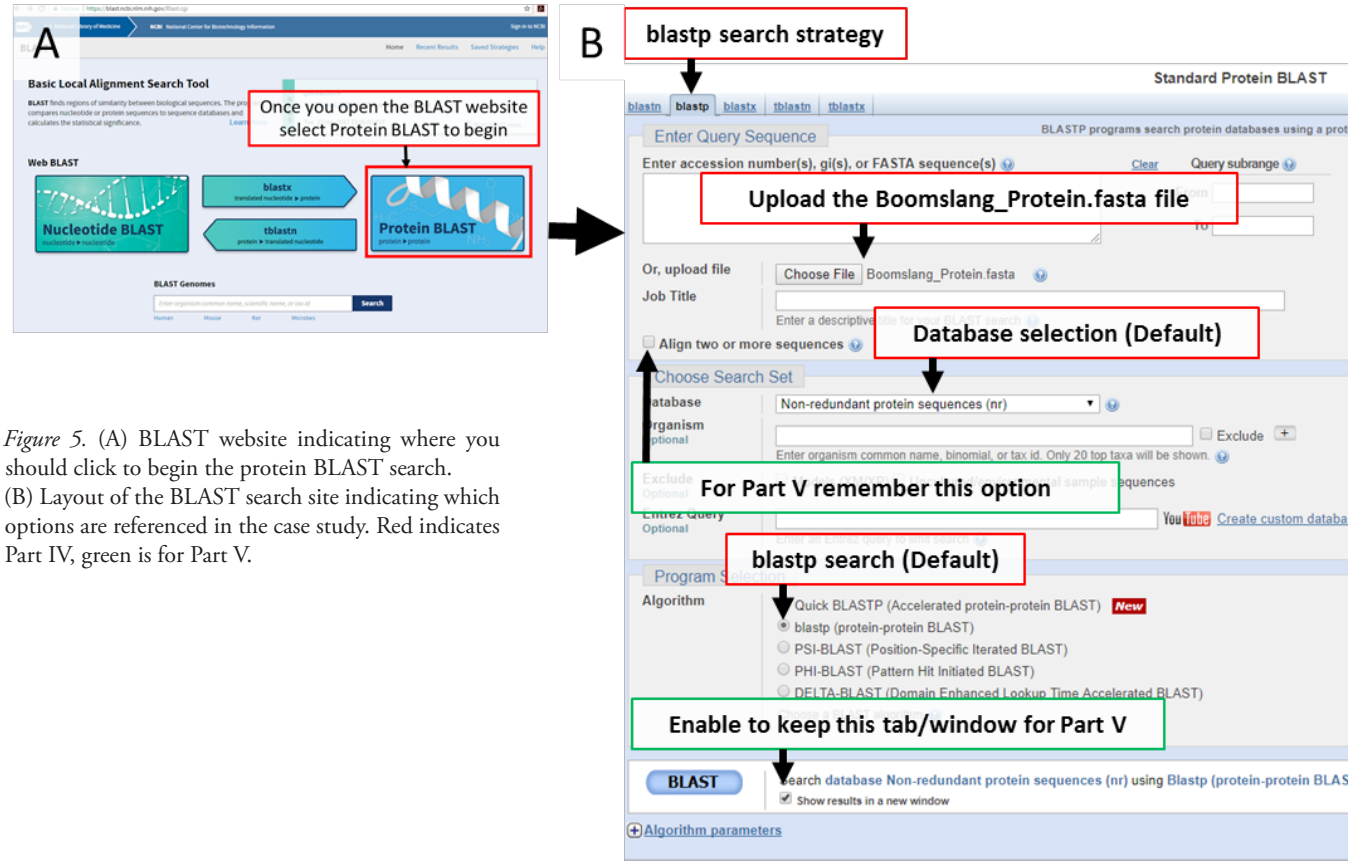

Figure 5. (A) BLAST website indicating where you should click to begin the protein BLAST search. (B) Layout of the BLAST search site indicating which options are referenced in the case study. Red indicates Part IV, green is for Part V.

The initial BLAST results will look like Figure 6 below.

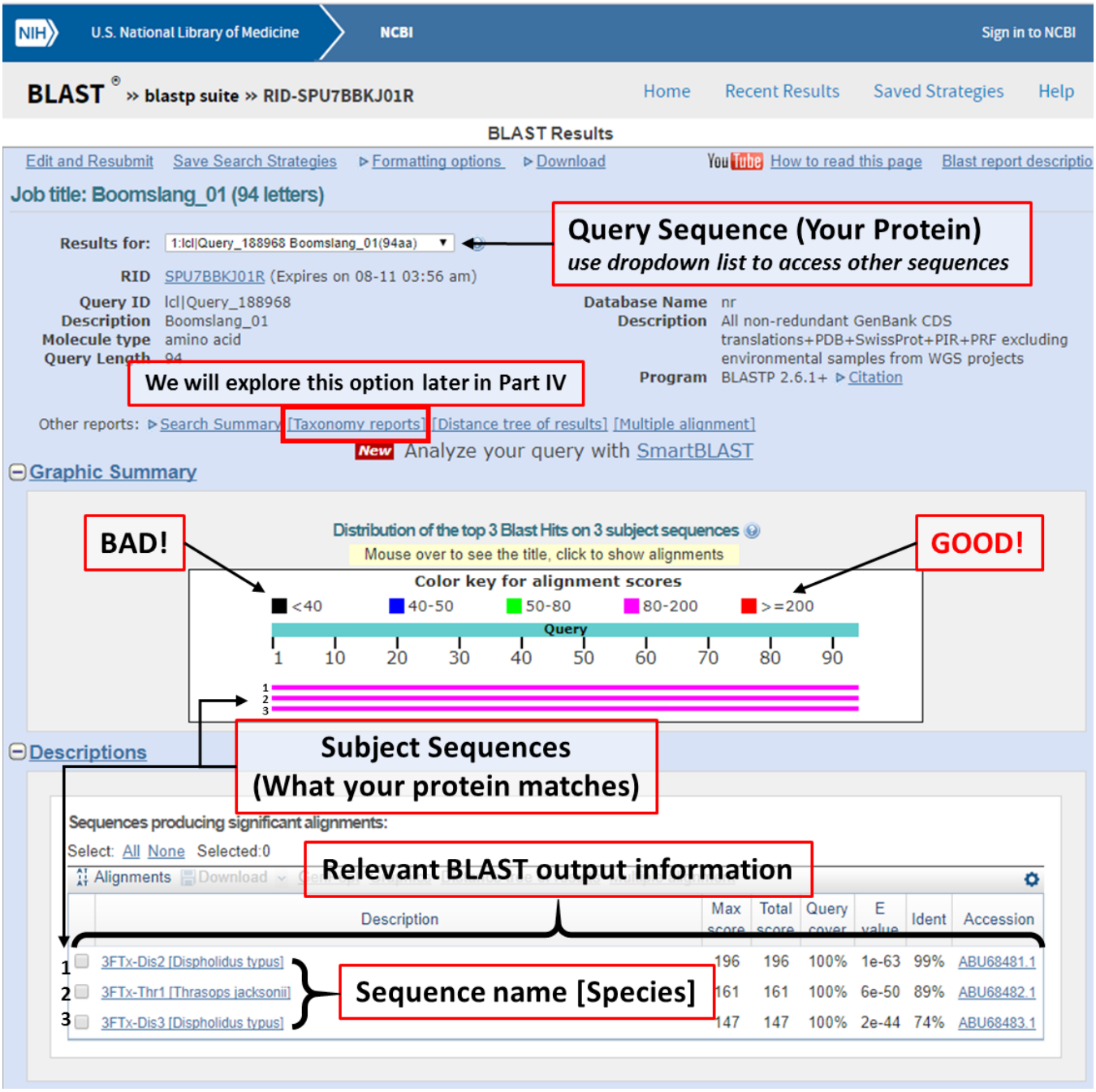

Figure 6. Screenshot of BLAST results with text highlighting key information for this case study.

Scroll down to the bottom of the BLAST webpage to visualize alignments and more info.

**** Answer the following questions with respect to the first sequence BLAST results. ****

Questions

1. How many proteins did Boomslang_01 match?

2. What species did Boomslang_01 match?

3. What venom protein family does Boomslang_01 belong to?

4. What is a relationship between the BLAST score (e.g. Max Score or Total Score) to other BLAST output information (e.g., Query Coverage, E-value, Percent Identity)?

5. What is a GenBank accession number?

**** Now look at some of the other boomslang BLAST results and answer the following questions. ****

6. Even though Boomslang_01 had a pretty good blast hit, there were proportionately less of the BLAST alignment scores colored red (≥ 200); most were magenta (80–200) or lower scores. Why do you think the BLAST alignment scores for Boomslang_01 were below other sequence alignment scores despite having nearly identical sequences?

7. Overall, what sorts of patterns are you noticing across BLAST hits?
   a. Is it better to have a high or low alignment score?
   b. Is it better to have a high or low Query Coverage?
   c. Is it better to have a high or low E-value?
   d. Is it better to have a high or low Percent Identity?

8. Any other observations you may note?

Fill out Table 2 below to characterize your boomslang proteins.

Table 2. Protein Analysis.

		Best BLAST Hit (Subject)			
Query Sequence	*Protein Family (See Table 1)*	*Sequence Name*	*Species*	*E-value*	*Accession*
Boomslang_01					
Boomslang_02					
Boomslang_03					
Boomslang_04					
Boomslang_05					
Boomslang_06					
Boomslang_07					
Boomslang_08					
Boomslang_09					
Boomslang_10					
Boomslang_11					
Boomslang_12					
Boomslang_13					
Boomslang_14					

**** Based on your protein analysis (Table 2), answer the following questions.****

9. What venom protein family was the most frequently recovered?

10. What venom protein family was the most abundant overall? (Use Figure 4 in combination with Table 2 to answer this question.)

11. Looking at the symptoms (Table 1), which venom toxin do you think ultimately lead to Dr. Karl P. Schmidt's death? Why?

12. Were there any sequences that were already on GenBank for the boomslang? (*Hint:* what's the species name for the boomslang?) Why don't you think every top hit was matched with the boomslang?

13. What seemed to be a good indicator for a low E-value/high alignment score across these sequences?

**** In any one of your BLAST hits pages click on the [Taxonomy Reports] (see Figure 6) and answer the following questions.****

14. Venom has evolved multiple times in snakes across four different families. Based on these BLAST hits, which family do you think the boomslang belongs to?

15. Which boomslang venom gene also recovered non-venomous animals? Why?

16. Why do you think multiple subject sequences from the same species are being recovered in this BLAST search?

Part V – The Transcriptome of Death

A transcriptome represents the collective messenger RNA expressed in a given organism, tissue, or cell. Although obtaining sequence data for messenger RNA can identify which genes are expressed, these genes do not always have a 1:1 ratio of the proteins identified in a given sample (as in Part IV). Now that you have characterized the protein diversity from the HPLC run, it is your job to identify the diversity of genes expressed within the transcriptome in order to get a better understanding of the processes of transcription, translation and gene family evolution that ultimately resulted in the proteins that killed Dr. Karl P. Schmidt.

In order to do this, return to the initial BLAST screen (Figure 5B), and make the following modifications. Start with the tblastn search, translated nucleotide subjects using a protein query (see Figure 7) and select the "Align two or more sequences" box within the Enter Query Sequence box.

- ☐ **Enter Query Sequence:** Click [Choose File] and select the protein sequences again (Boomslang_Proteins.fasta).
- ☐ **Enter Subject Sequence:** Click [Choose File] and select the boomslang transcriptome sequences (Boomslang_Transcriptome.fasta, supplied by your instructor).

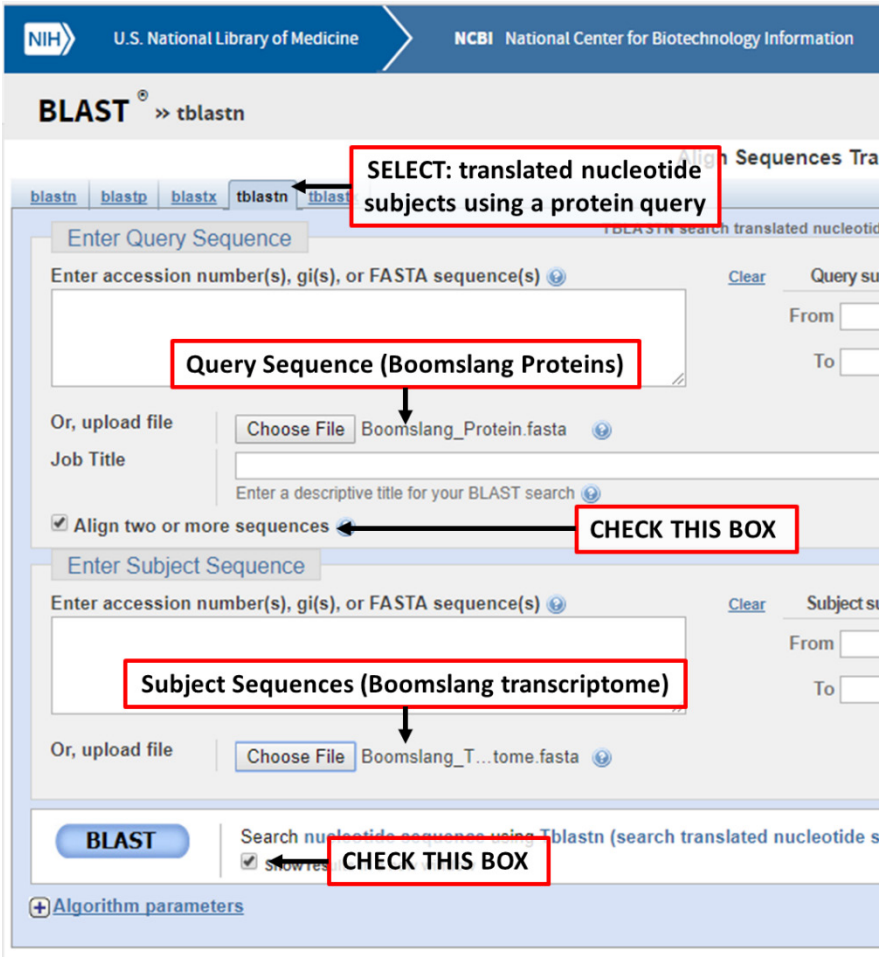

Figure 7. BLAST options relevant to Part V of this case study.

In order to evaluate which of these transcripts may (or may not) still be involved with envenomation we must first examine the sequence alignments with the transcribed messenger RNA and how it codes for the translated protein. Select Boomslang_13 (the C-type lectin-like (CTL) toxin) from the dropdown list and refer to Figure 8 to help you answer the questions below.

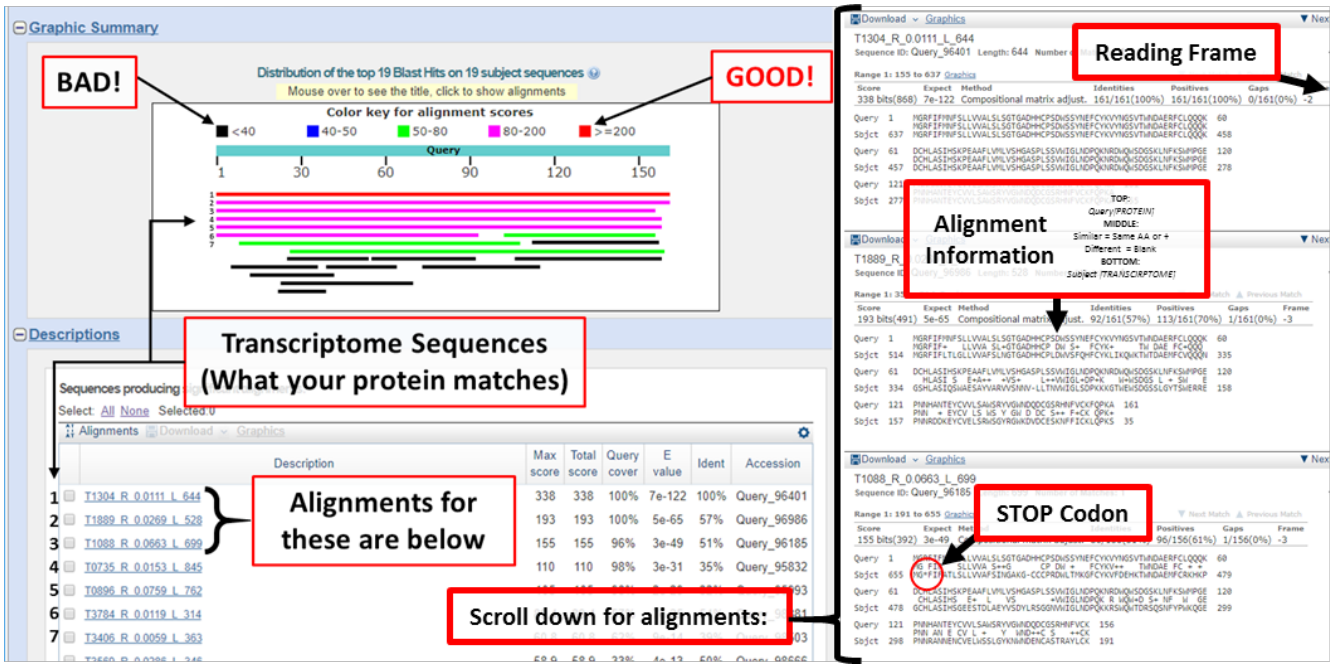

Figure 8. Summary of tblastn results for Boomslang_13.

**** Based on the tblastn results from the transcriptome survey of Boomslang_13 answer the following questions. ****

Questions

1. What sequence is most likely the messenger RNA which is coding for the protein identified in the proteomic analysis? Why?

2. According to the BLAST results what reading frame was used to translate the messenger RNA into this protein? What does this mean?

3. Regardless of alignment score (from good to bad) how many sequences were identified in this tblastn search? Do you think *all* of these are gene copies of the C-type lectin-like (CTL) toxin? Why or why not?

4. Although the alignment score for T1304_R_0.0111_L_644 was the highest, T1889_R_0.0269_L_528 and T1088_R0.0663_L_699 have good alignment scores and E-values. One major difference, however, is the presence of a stop codon (Figure 8) in T1088_R0.0663_L_699. If this is found in the messenger RNA would you expect this full protein to be found in the proteome? Why or why not?

5. Scroll through the rest of the tblastn hits for Boomslang_13. What do you notice with regard to alignment score and overall alignments as tblasn hits decrease in their alignment score and increase in their E-value?

6. If you were tasked with the job of identifying the number of venom gene copies encoded in the messenger RNA of this sample how many potential venom genes do you think would be recovered? Why? (*Hint:* there is no wrong answer.)

Fill out Table 3 below to characterize your boomslang proteins:

Table 3. Transcriptome Analysis

Query Sequence	*Protein Family (Table 2)*	*# Matches*	*Top Hit Reading Frame*	*Number of "Venom Genes"**	*Number of matches that were not Venom*
Boomslang_01					
Boomslang_03					
Boomslang_04					
Boomslang_07					
Boomslang_11					
Boomslang_13	CTL	15	-2	4+	~10
*When estimating the number of venom genes use 1, 2, 3, or 4+.					

**** Based on the subsample analysis you did for Table 3 answer the following questions:****

7. What toxin venom gene has the most gene copies (not BLAST hits)?

8. Which toxins have only one gene copy in the transcriptome?

9. How might a high number of gene copies relate to the flexibility of the overall toxin cocktail?

10. If you could isolate all the proteins for a given toxin, what sort of experiments can you do to determine if the protein is actually a toxin?

11. After taking the time to analyze the protein toxins found in boomslang venom, would you add any details to the autopsy report from Part III? If so, what details would you add?