TP53 RETROGENE FOR CANCER CONTROL: THE GENE MOST COMMON MUTATED (TP53) AND EVOLUTIONRY PATTERNS

By

Rama Mohammad Rabih Ghannam

A capstone project submitted for Graduation with University Honors

May 8, 2022

APPROVED

Dr. Leonard Nunney

Department of Evolution, Ecology, & Organismal Biology

Dr. Richard Cardullo, Howard H Hays Jr. Chair

University Honors

Abstract

The primary cause of cancer is a change in the cell caused by the DNA sequence. In the United States, cancer is the second leading cause of death. In 2007, approximately 1.45 million new cases were diagnosed (Truskey, 2004). This gene will cause cancer when it's missing or damaged. The development of p53 dysfunction is the hallmark of infiltrating cancers (Shackney 2003). How species may have evolved to avoid cancer by having an extra copy of TP53 called retrogenes. Increasing the number of cells and cell divisions increases increase the risk of developing cancer (Peto 1977). All retrogenes of microbat duplicate in three main events and one even were for rat and mouse retrogenes. Because of the presence of stop codons throughout the sequence, these TP53 retrogenes are not functional. One of the rat retrogenes (chromosome 18) does not encode a protein.

Acknowledgements

This research would not have been possible without the expertise of my faculty adviser Dr. Leonard Nunney. I would also like to thank the University Honor program for providing me with the opportunity to conduct this research, as well as Dr. Richard Cardullo, Howard H Hays Jr. Chair.

Special appreciation to my family, starting with my mom (Ms. Abu Hantash) and my dad (Mr. Ghannam). I'd also like to appreciate my siblings for sticking with me, and helping me through tough times, thank you Ahmad, Sireen, Reem, and Lana. This year was almost impossible without them. Lastly thank you to my Fiancé, Ahmad Sameh.

Background

According to a study conducted between 2013 and 2017, 442.4 men and women per 100,000 were diagnosed with cancer each year. By 2040, there will be 29.5 million new cancer cases and 16.4 million cancer deaths worldwide (Armstrong, 2014). According to The Veterinary Cancer Society, cancer kills 47 percent of dogs and 32 percent of cats. Humans and dogs have the same cancer risk, and they share symptoms such as weight loss without effort, skin changes, and abnormal swelling.

Cancer is a term that refers to 200 different diseases that stem from a single cell (What is cancer, 2020).The study of the genome's sequence aids in determining whether or not there are mutations in the genes that cause and contribute to the progression of cancer (Cancer genomics, 2017). Cancer develops as a result of somatic mutations affecting the genetic control of cell growth. It is expected that large long-lived animals would have additional cancer-suppressing mechanisms. Comparing the body size and longevity in humans and mice, a vastly greater incidence of cancer is expected in humans (Peto 1977). Understanding why this does not happen in large and/or long-lived animals may be beneficial to humans as well.

DNA contains many genes, which are then transcribed into RNA and used as a template before being translated into protein. DNA is composed of four chemical bases: adenine (A), thymine (T), cytosine (C), and guanine (G), and DNA sequencing determines the arrangement of these four bases in any given region of the genome (DNA sequencing fact sheet, 2020). In the human body, the DNA is about 3 billion base pairs long. Genome sequencing is a list that reveals the order of bases in an organism's entire genome (Saraswathy 2011).

Because the gene TP53 which produces protein that in the nucleus is the most frequently mutated (or lost) in cancer cells (Levine, 1991), any change in this gene will cause/promote cancer. Infiltrating cancers are often distinguished by the development of p53 dysfunction (Shackney, 2003). Mutations in the TP53 gene cause the gene protein to change if the mutation changes one or more amino acids, resulting in a new version (altered) of the protein that does not function as well as the original copy of the protein, which can be the cause of a cancer.

Studying evolution is the main concept of the phylogenetic tree. Evolution is defined as the change in a species' characteristics over time and across generations. Natural selection is the process of adaptive evolution. Because evolution can take many different paths, we will focus on whether species such as elephants, rats, mice, bats, and others may have evolved additional copies of TP53 to avoid cancer.

Introduction

Maximum body size and longevity are important life-history characteristics. Mammal longevity ranges, for example, from 6 to 12 months in the muller's giant Sunda rat (Venner 2018) to 211 years in the bowhead whale (Sulak et al 2016), and body size ranges from 2 g in the bumblebee bat (Hance, 2020) to 163193 kg in blue whales (Smith et al 2011). Lifespan has a positive correlation with body size, with larger species usually living longer than smaller species.. Increasing the number of cells (body size) and the number of cell divisions (lifespan) increases the likelihood of accumulating oncogenic mutations that promote malignancy and increase the risk of developing cancer (Peto, 1977). However, there is no cross-species relationship between body size and cancer risk, or between lifespan and cancer risk (Vazquez, 2021). This failure of expectation

has been dubbed Peto's paradox, and it was predicted that large and long-lived animals would have additional cancer-suppressing mechanisms (Nunney, 1999).

Mechanisms for cancer resistance that prevent the accumulation of genetic damage in small long-lived mammals have been identified. The naked mole rat has a very long lifespan, with a maximum lifespan exceeding 30 years (Tian et al 2013) and a minimum lifespan of seven years (Pappas, 2018). The blind mole rat has a lifespan of 20 years (Gorbunova et al 2014). The cancer rates in these two mole rats are significantly different. Whereas anticancer mechanisms in the blind mole rat evolved an amino acid change in p53 and diverged from those in the naked mole rat. The naked mole-rat is mediated by their cells' extreme sensitivity to cell contact, which aids in cancer resistance (Gorbunova et al 2014). Naked mole-rat fibroblasts secrete extremely high-molecular-mass hyaluronan (HA), which is five times larger than human or mouse HA (Tian et al 2013).

However, the mechanism for why large-bodied animals develops more cancer resistance remains unknown. One such mechanism could be the evolution of an increase in tumor suppressor gene copy number.

TP53 (encoding the protein p53 [RefSeq NM 000546]) is a critical tumor suppressor gene that is mutated in the majority of human cancers. It is also known as the genome's guardian (Sigal 2000). The p53 protein plays critical roles in the cellular response to a variety of stresses, and it also protects the genomic integrity (Aubrey 2006). When Tp53 is functioning properly, it will either prevent the proliferation of damaged cells or induce apoptosis (cell death), removing these cells from the body (Caspari 2000). Any p53 inactivation will result in three cancer cell characteristics: apoptosis suppression, increased proliferation, and genomic instability. Because the TP53 gene is only found in one copy of the human haploid genome, an individual has two

copies that aid in the prevention of cancer development. Li-Fraumeni syndrome (LFS) is caused by the absence or malfunction of one of these copies (Varley 1999). LFS is a type of inherited cancer proclivity. Patients with Li-Fraumeni syndrome (LFS) are more likely to develop cancer.

Given the importance of TP53 in cancer control, it was significant that researchers discovered not only a single TP53 gene, but also 19 TP53 retrogenes in the genome of the large, long-lived African elephant. A retrogene is a new sequence that appears after the original gene has been duplicated; this copy is known as the processed copy. There are no introns in these copies. It is produced by reverse-transcription of a gene's messenger RNA (Bai, 2008). Many of these retrogenes, according to the Sulak et al paper, are capable of producing functional proteins. The elephant's 19 TP53 retrogenes all originated from a single event, followed by duplication of the original retrogene.

Several of these duplicates have been transcribed and translated into elephant tissues (Sulak et al 2016). When compared to other large body size animals (such as the American mastodon, woolly mammoth, and Columbian mammoth), the copy number increased relatively quickly, coinciding with the evolution of large body sizes in the Proboscidean lineage. Further investigation into the function of these extra copies revealed that elephant cells have an enhanced response to DNA damage, which may be mediated by a hyperactive TP53 signaling pathway, and that this augmented TP53 signaling may be dependent on at least one of the TP53 retrogenes (Sulak et al 2016).

TP53 retrogenes have been found in several other mammals. The Rat has five retrogenes, the Microbat has seven, and the Wallaby has two, according to the Sulak et al paper. The mouse, opossum, Tasmanian devil, rabbit, squirrel, kangaroo rat, tenrec, rock, and manatee all share one

retrogene. Many Megabat species do not have retrogenes. The main question here is whether or not these retrogenes have any function.

The evidence that the 19 TP53 retrogenes copies may have lost their function would be the presence of stop codons in the sequence. This can help identify if these retrogenes have been active as extra TP53 copies. Have the extra copies of this protein lost their functionality? The production of p53 by these retrogenes will prevent cancer, but new stop codons act to shorten the protein produced. When did stop codons appear during the TP53 retrogene expansion?

Method

Find The TP53 Sequences

The first step in locating the tp53 for our species was to go to https://www.ncbi.nlm.nih.gov/nuccore/. We enter the name of the species and TP53 into the search box; for our Rat sequence, the species name was **Rattus norvegicus**. In most cases, the gene sequence will appear at the top, as illustrated in Fig 1. Click on the page, then on the top right of the page, click on **Download Datasets**, and then on **Gene Sequences (FASTA)**. The file will be downloaded as a zip file; open the file normally named **Tp53 datasets**, then navigate to the **nbci_dataset** folder, then **data**, and then open the.fna file **gene.fna**. That will be the entire gene with introns; in order to get only the TP53 sequence, we must delete these sequences. To remove the introns, locate the CDS section this will be the TP53 gene.

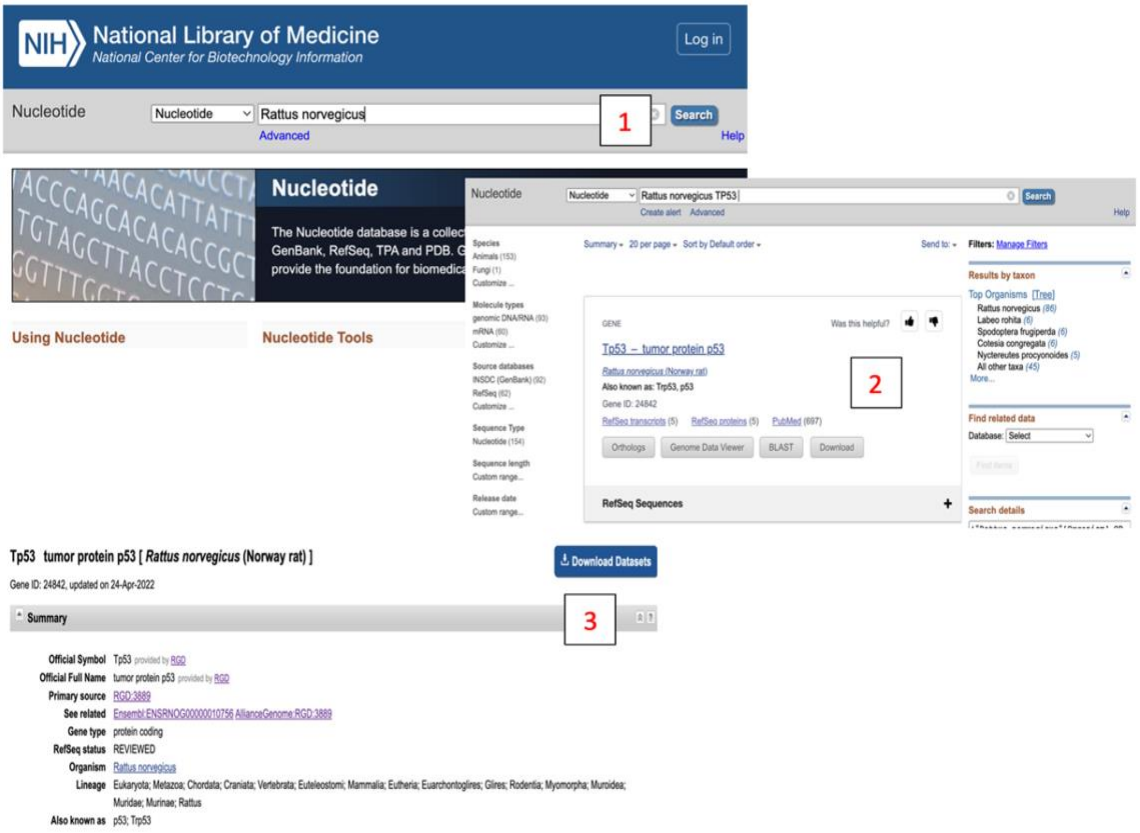

Figure 1: How to Locate TP53 Sequences

To Find The TP53RTG

Visit https://www.ncbi.nlm.nih.gov/genome for find the TP53RTG. Enter the specie **Rattus norvegicus** and hit search; the Rat page will appear, as shown in Fig 2. Under **Tools**, select **BLAST Genome**. As shown in Fig 3, a blue page will appear; enter the TP53 sequence that you just discovered in the first big white box. Then select **Somewhat similar sequences (blastn)** from the **Program Selection** menu. Then, under the **BLAST** button, check the box that says, **Show results in a new window**.

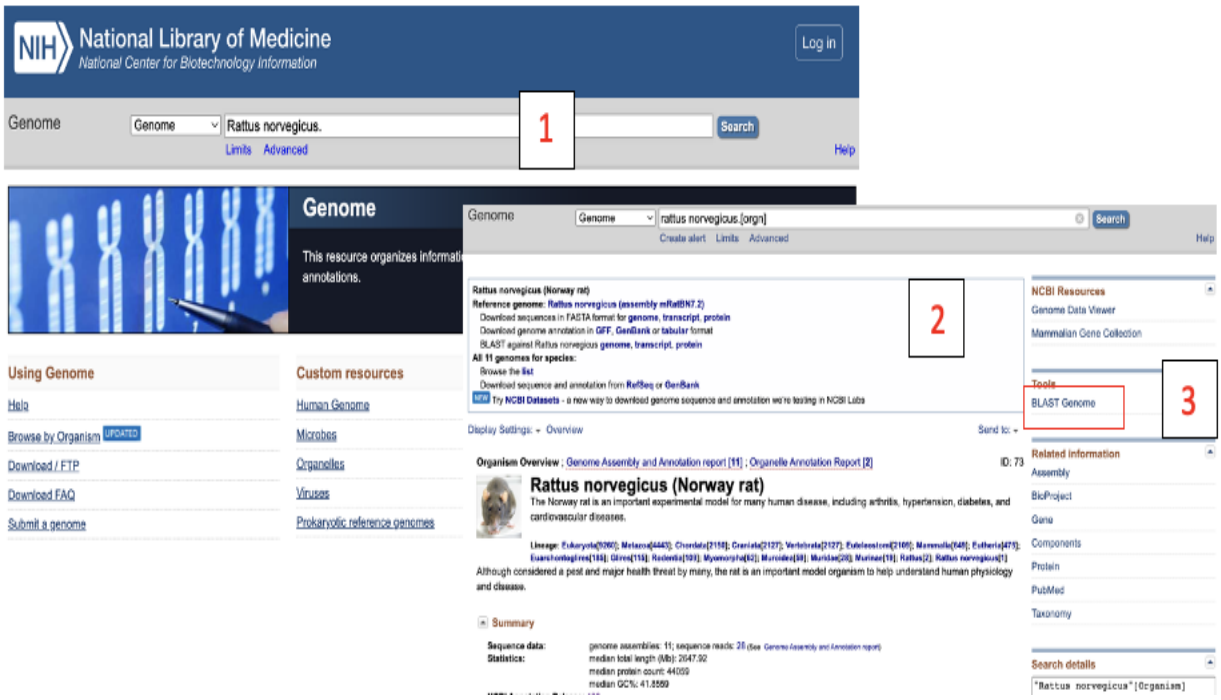

Figure 2: How to Locate the TP53RTG

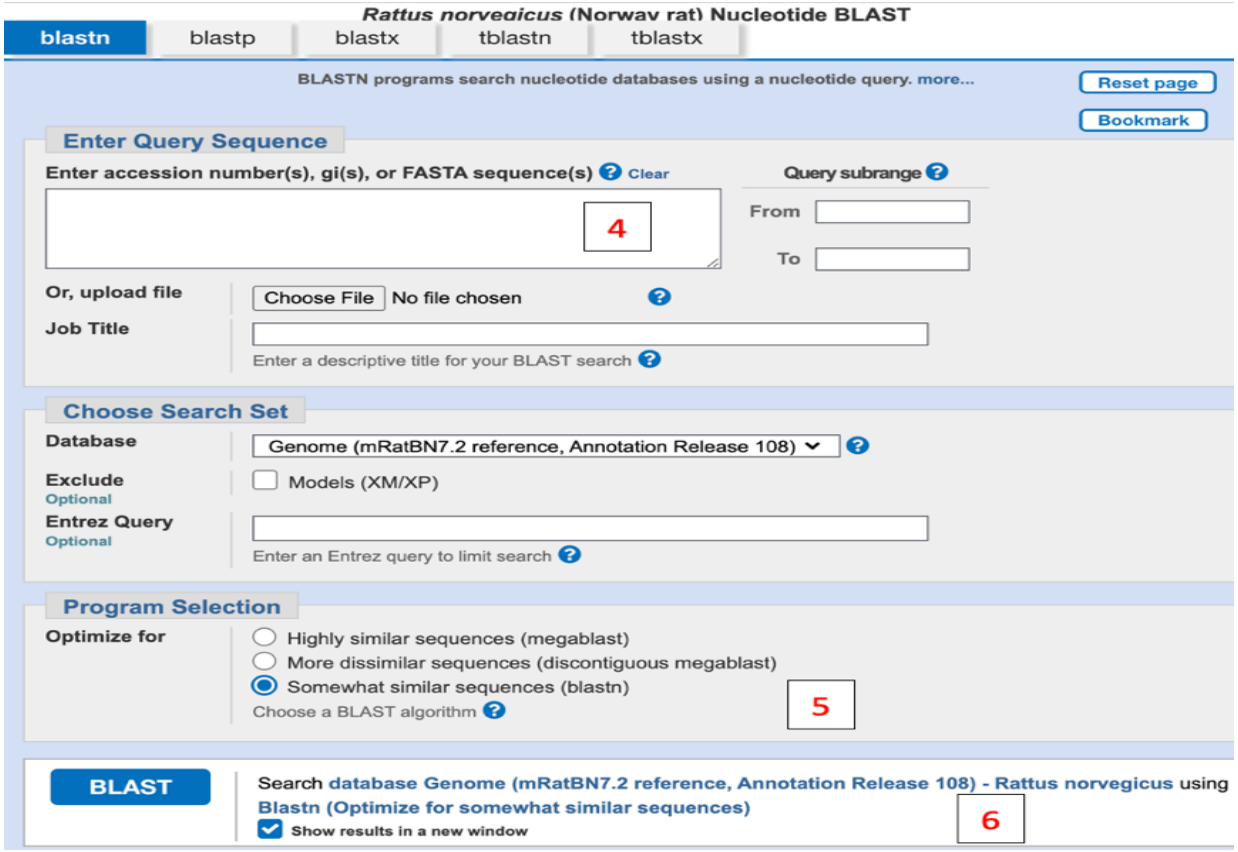

Figure 3: BLAST screen

The screen shown in Fig 4 will appear after you click the **BLAST** button. Go to the **Graphic Summary** tab to see the best match sequence of the one that was blasted. We see five retrogenes for TP53 of Rat sequence, as shown on the graphic summary tap. On the **Alignments** tab, click the **GenBank** button next to the Range 1 numbers, then click the **FASTA** button on the left side of the next screen. The sequences can then be copied and pasted into another file.

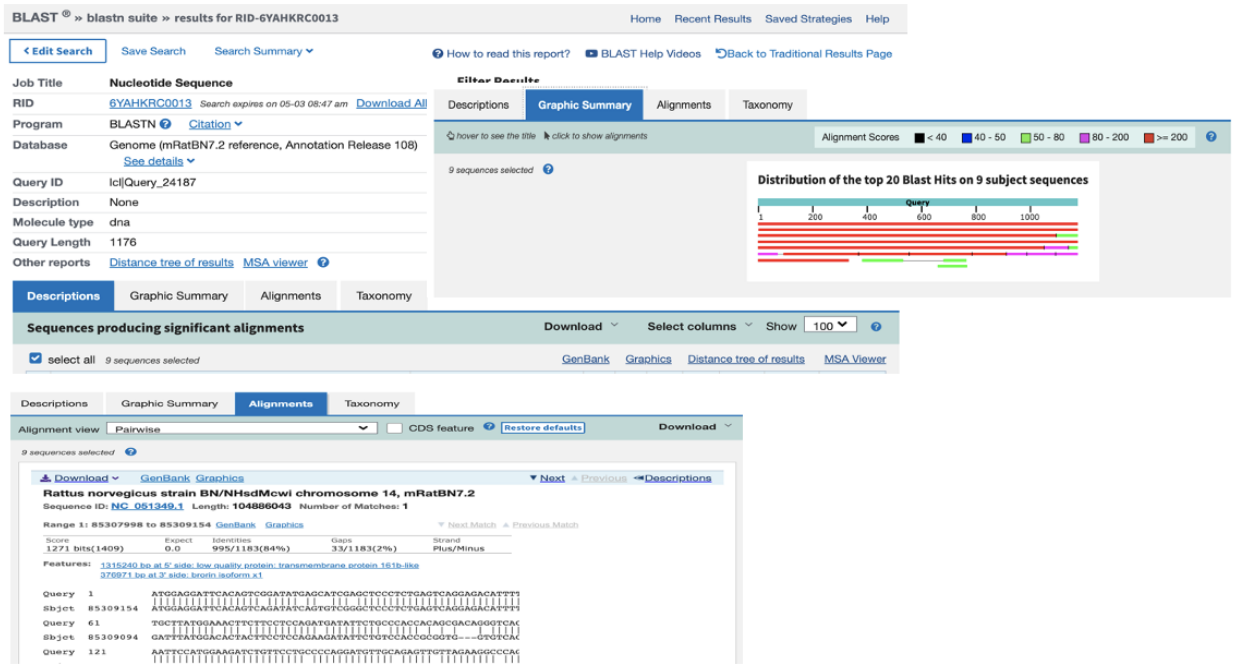

Figure 4: The graphic summary and the Alignments screen

To Create A Phylogenetic Tree

Launch the **Unipro UGENE** app and select the **Create Sequence** option. As shown in Fig 5, a Create Sequence screen will appear. Copy and paste the sequence you've been collecting into the white box, and then **Save** the sequence to a file, then press the **Create** button. Then, open the file into which we want to upload this sequence, and right-click the mouse. An **add** button will appear, allowing us to add the sequence to the folder. To start a tree on UGENE, click the **Tree** icon at the

top of the screen. We can then edit the tree and click **Build** to have it built, and you now have a phylogenetic tree.

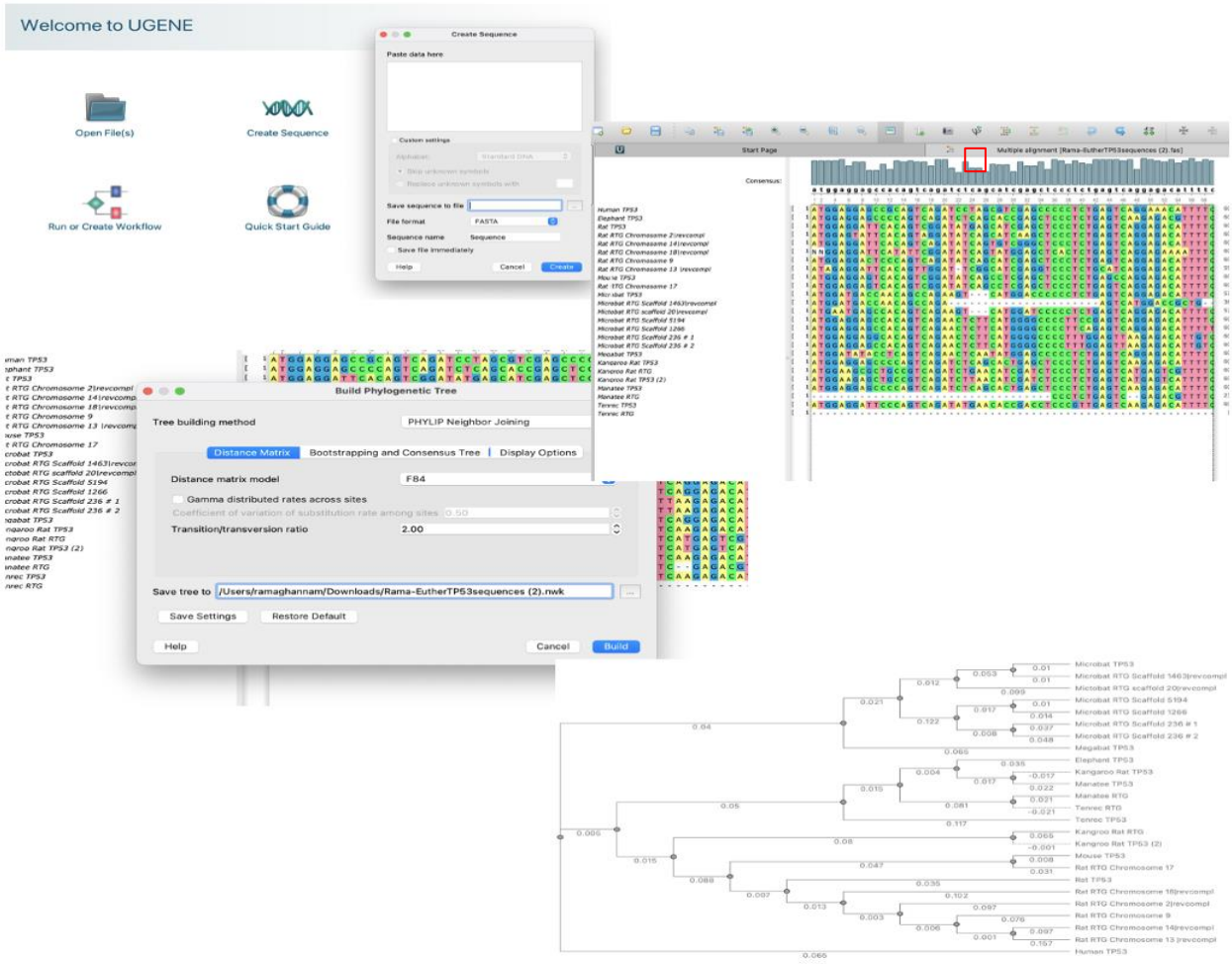

Figure 5: How to create a phylogenetic tree

Results

The number of species with TP53 retrogenes listed in Sulak's paper (plus human) is shown in Table 1. We began by confirming the presence of the retrogenes in the species under investigation. The only difference in retrogene numbers was for the Tasmanian devil, which had three copies of retrogenes rather than one as documented in Sulak's paper.

Table 1: Identification of functional TP53 gene

Name	Specie Name	Good TP53	Number of TP53 retrogenes (Sulak)	Number of TP53 retrogenes (Found)	Number of Stop Codons	Comment
Human TP53	Homo sapiens	YES	0	0	1	
Rat TP53	Rattus norvegicus	YES	5	5	1	
Mouse TP53	Mus musculus	YES	1	1	1	
Microbat TP53	Myotis lucifugus	YES	6	6	1	
Megabat TP53	Pteropus vampyrus	YES	1	1	0	There is no stop codon at the end
Elephant TP53	Loxodonta africana	YES	19	19	1	
Kangaroo Rat TP53	Dipodomys ordii	YES	1	1	1	A significant portion is missing
Kangaroo Rat TP53 (2)	Dipodomys spectabilis	NO	1	1	11	
Tenrec Tp53	Echinops telfairi	YES	1	1	1	
Manatee TP53	Trichechus manatus latirostris	YES	1	1	1	
Tasmanian devil TP53	Sarcophilus harrisii	NO	1	3	13	On the TP53 gene, there are 13 stop codons
Wallaby TP53	Macropus eugenii	YES	1	1	1	
Opossum TP53	Monodelphis domestica	NO	1	1	7	The first 221 sequences are missing

- The single stop codon in the normal (i.e. non-retrogene) TP53 sequences listed in Table 1 indicates the presence of a stop codon at the end of the sequence.

Microbat and Megabat

Searches for the TP53 gene sequence and the retrogenes for the Microbat (Myotis lucifugus) and Megabat (Pteropus vampyrus), showed that there are no retrogenes (TP53RTG) in the Megabat genome as also indicated by a previous study (Sulak et al 2016). The TP53 gene for the Megabat does not show any stop codons. There are seven sequences for the Microbat six of them are TP53RTG.

The phylogenetic tree (with the dog TP53 gene sequence as outgroup) in Fig 6 shows that TP53 gene for Microbat and the Scaffold 1463|revcompl RTG are sister taxa because these group with an immediate common ancestor. A shared derived character was found in these two sequences and Scaffold 20|revcompl RTG because they are sharing the same clade. A shared ancestral character was a found between all the Microbat sequences and the TP53 of the Megabat. Clarified that all the TP53RTG generate after the split of the two types of bat. Scaffold 1463 were the most recent sequences with two stop codons on position that shown on the table 2. Scaffold 20 has fourteen stop codons, Scaffold 5194 thirteen stop codons and Scaffold 1266 fifteen stop codons. There are two copy of scaffold 236, the first copy has eleven stop codons, and the second copy has eighteen stop codons.

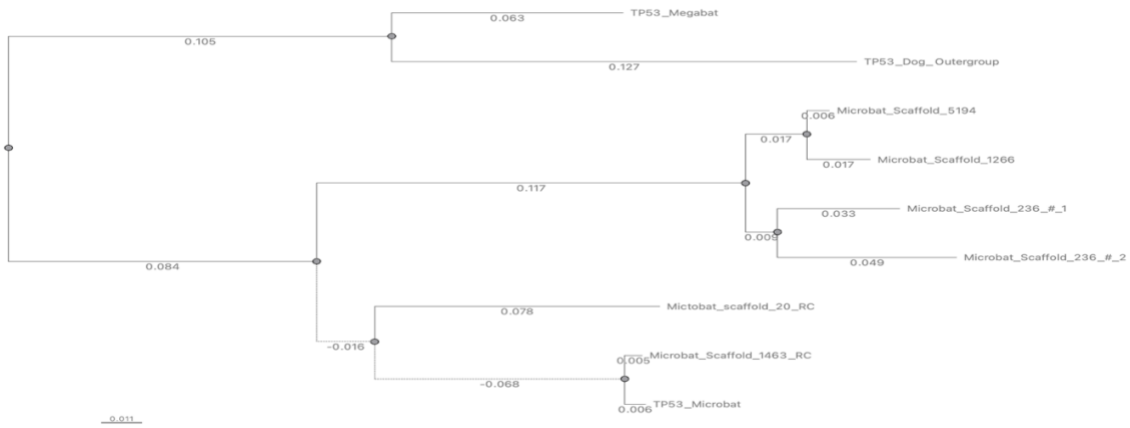

Figure 6: Phylogenetic tree for Microbat, Megabat, and dog (outgroup)

Rat and Mouse

In agreement with the Sulak et al paper, there are five retrogenes for Rat (Rattus norvegicus) plus the original TP53, i.e. six sequences in the Rat. For the Mouse (Mus musculus), there is one retrogene and in total there are two sequences for the Mouse. An elephant Tp53 gene sequence was included solely to serve as an outgroup. Using the method described above to find the TP53RTG, keep an eye on the sequence's order because some sequences must be reverse complements, which can be accomplished using UGENE. The sequences for rat chromosomes 2, 14, 18, and 13 are reverse complements.

The phylogenetic tree shows that, as seen in the elephant, all TP53RTG for Rat happen in a single event and then they all duplicate as shown in Fig 9. It's very similar to what happened in the elephant files where all TP53RTG duplicated from each other. A shared derived character was found in chromosome 13 Rat with Chromosome 14 Rat because they are sharing the same clade. A shared ancestral character was found between all the Rat retrogress and the original TP53 gene for the Rat. Because all rat and mouse TP53 retrogenes begin with ATG, they are classified as protein-coding genes. Except for chromosome 18, which lacks the first two bases.

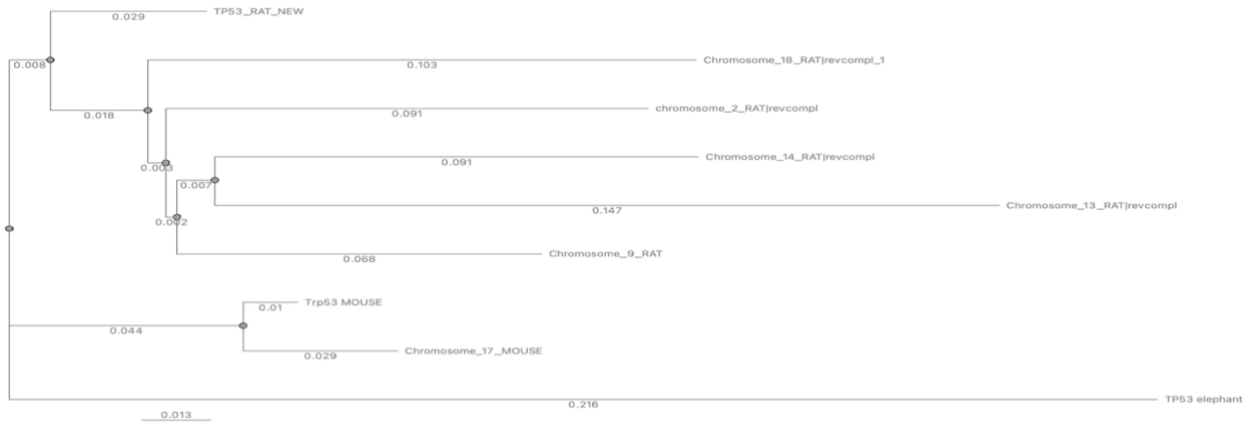

Figure 8: Phylogenetic tree for Rat, Mouse, and Elephant (outgroup)

Discussion

Microbat and Megabat

The position of the stop codons was discovered to be related to the Microbat TP53, and this position is shown in table 2. Figure 7 depicts the position of each stope codon on these sequences, whether it occurs after or before the split, and how many events there are. The phylogenetic tree reveals three major events that occurred in the microbat sequence, which are depicted in different colors in Fig 7. Unlike the elephant RTG sequences, where all retrogenes are derived from a single event by later duplication, the retrogenes for the microbat[HOW MANY RTG ORIGINS?] and the TP53 begin with ATG, so there are organized as protein-coding genes.

Table 2: Position of stop codons for Microbat

	Number of stop codons	Position of stop codons (bp) related the Microbat Tp53*	Mutations position (bp)
Scaffold 1463|revcompl	2	902 and 926	21 (22), 58 (92), 152 (3), 167 (3), 183 (6), 223 (24), 305 (5), 331 (3), 404 (1), 453 (3), 489 (1), 601 (1), 608 (14), 626 (3), 685 (3), 708 (3), 729 (7), 745 (1), 754 (1), 783 (3), 801 (11), 838 (1), 859 (6), 871 (1), 886 (37), 930 (86), 1078 (3), 1095 (7), 1110 (5), 1120 (1), 1143 (3), 1255 (6), 1286 (3), 1307 (1), 1317 (5), 1334 (1),
Scaffold 20|revcompl	14	205, 298, 307, 331, 455, 464, 586, 755, 779, 902, 926, 1022, 1055, 1130	24 (3), 120 (18), 167 (3), 183 (6), 218 (31), 305 (5), 331 (3), 404 (1), 455 (1), 489 (1), 601 (1), 608 (14), 626 (3), 685 (3), 729 (7), 745 (1), 783 (3), 800 (12), 838 (1), 859 (6), 869 (3), 896 (1), 952 (1), 962 (14), 978 (1), 995 (21), 1023 (3), 1033 (3), 1044 (3), 1054 (1), 1110 (5), 1143 (3), 1183 (1), 1208 (3), 1255 (6), 1286 (3), 1317 (5), 1334 (1),
Scaffold 5194	13	98, 119, 173, 278, 402, 629, 638, 650, 684, 876, 906, 1088	71 (2), 120 (18), 152 (3), 167 (3), 183 (6), 223 (24), 305 (5), 331 (3), 386 (1), 404 (1), 453 (3), 489 (1), 608 (14), 626 (3), 685 (3), 729 (7), 745 (1), 783 (3), 801 (11), 820 (10), 838 (1), 859 (6), 869 (3), 896 (1), 952

			(1), 962 (14), 978 (1), 995 (21), 1023 (3), 1033 (3), 1044 (3), 1054 (1), 1078 (3), 1095 (7), 1110 (5), 1143 (3), 1155 (6), 1208 (3), 1214 (9), 1254 (8), 1267 (1), 1286 (3), 1307 (4), 1317 (1), 1334 (1)
Scaffold 1266	15	63, 98, 119, 173, 278, 336, 611, 614, 629, 638, 650, 684, 876, 906, , 1088	83 (1), 100 (1), 120 (18), 152 (3), 167 (3), 183 (6), 223 (24), 305 (5), 386 (1), 404 (1), 453 (3), 489 (1), 608 (14), 626 (3), 685 (3), 729 (7), 745 (1), 783 (3), 801 (11), 820 (10), 838 (1), 859 (6), 869 (3), 896 (1), 952 (1), 962 (14), 978 (1), 995 (21), 1023 (3), 1033 (3), 1044 (3), 1054 (1), 1078 (3), 1095 (7), 1110 (5), 1120 (1), 1143 (3), 1155 (6), 1208 (3), 1214 (9), 1254 (8), 1267 (1), 1286 (3), 1307 (4), 1334 (1)
Scaffold 236 #1	11	45, 331, 376, 472, 484, 706, 779, 992, 1022, 1103, 1127	120 (18), 152 (3), 167 (3), 183 (6), 223 (24), 331 (3), 386 (1), 404 (1), 453 (3), 489 (1), 601 (1), 608 (14), 626 (3), 685 (3), 729 (7), 745 (1), 783 (3), 801 (11), 820 (10), 838 (1), 859 (6), 869 (3), 896 (1), 952 (1), 962 (14), 978 (1), 995 (21), 1023 (3), 1033 (3), 1044 (3), 1054 (1), 1078 (3), 1095 (7), 1110 (5), 1120 (1), 1143 (3), 1155 (6), 1208 (3), 1214 (9), 1254 (8), 1267 (1), 1286 (3), 1307 (4), 1317 (2), 1334 (1)
Scaffold 236 #2	18	45, 331, 376, 472, 484, 611, 614, 638, 697, 706, 739, 901, 955, 1018, 1030, 1036, 1073, 1096	120 (18), 152 (3), 167 (3), 183 (6), 223 (24), 305 (5), 331 (3), 386 (1), 404 (1), 453 (3), 489 (1), 601 (1), 608 (14), 626 (3), 687 (1), 729 (7), 743 (3), 783 (3), 801 (11), 838 (1), 859 (6), 869 (3), 896 (1), 952 (1), 962 (14), 978 (1), 995 (21), 1023 (3), 1033 (3), 1044 (3), 1054 (1), 1078 (3), 1095 (7), 1110 (5), 1120 (1), 1143 (3), 1155 (6), 1208 (3), 1216 (9), 1254 (8), 1267 (1), 1286 (3), 1307 (4), 1334 (1)

* The position of the last element of the amino acid

|revcompl : Reverse complement

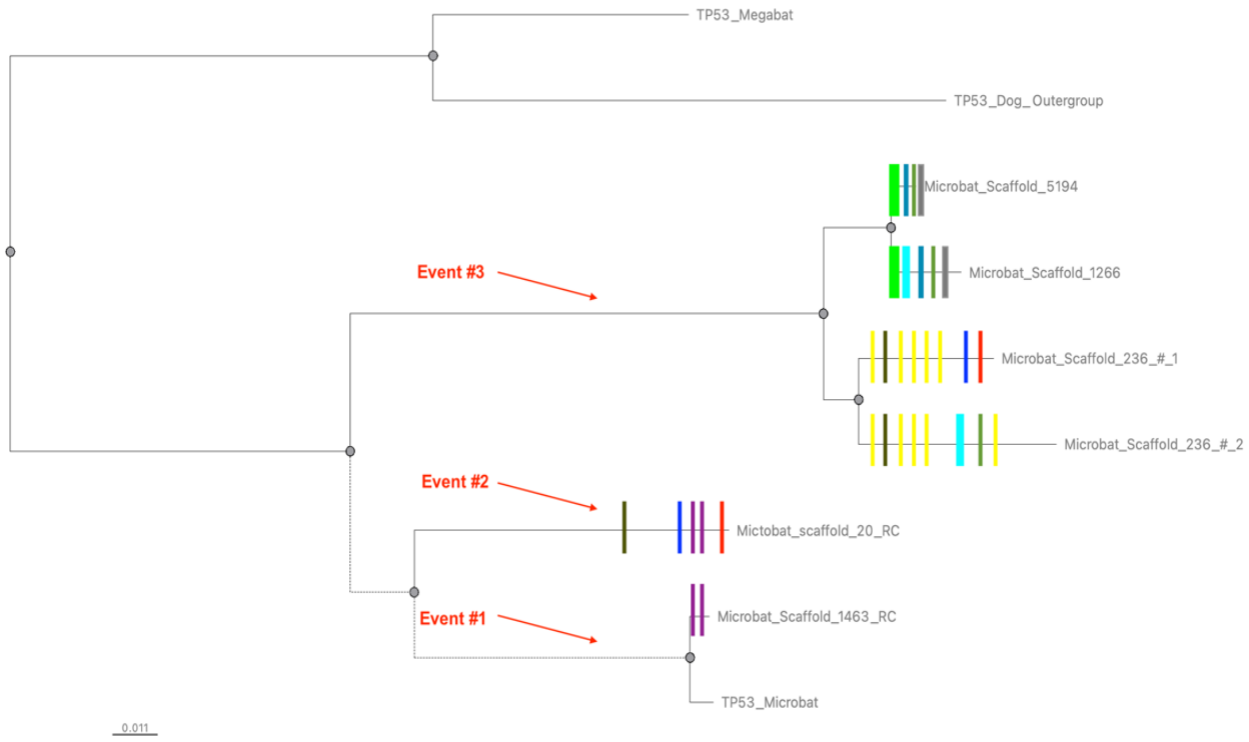

Figure 7: the stop codons tree and the position of spilt for Microbat

Rat and Mouse

The position of the stop codons was discovered to be related to the original, as shown in table 3. Figure 9 depicts the location of each stope codon on these sequences, whether it occurs after or before the split, and the number of events. The phylogenetic tree reveals one major event in the rat sequence, which they duplicated after the original rat TP53. And, as shown, the only mouse retrogene duplicated well after the rat/mouse split and quite recently.

Table 3: Position of stop codons for Rat and Mouse

	Number of stop codons	Position of stop codons (bp) related there Original Tp53*	Mutations position (bp)
Chromosome 2 RAT|revcompl	19	18, 66, 69, 192, 346, 409, 506, 533, 545, 551, 554, 720, 733, 908, 1020, 1045, 1102, 1166.	104 (3), 129 (9), 152 (3), 167 (3), 183 (6), 191 (3), 223 (48), 283 (8), 330 (3), 337 (3), 403 (1), 452 (3), 488 (1), 516 (1), 585 (3), 600 (1), 625 (3), 684

			(3), 728 (7), 753 (1), 782 (3), 800 (11), 858 (6), 868 (3), 961 (14), 994 (21), 1022 (3), 1032 (3), 1043 (3), 1077 (24), 1142 (3), 1189 (13), 1207 (3), 1254 (6), 1285 (3), 1306 (1), 1316 (12),
Chromosome 14 RAT |revcompl	15	213, 424, 502, 514, 605, 656, 665, 677, GAP, 733, 806, 922, 1027, 1045, 1102	104 (3), 129 (9), 152 (3), 167 (3), 183 (6), 191 (3), 223 (3), 305 (5), 331 (2), 371 (10), 403 (1), 452 (3), 488 (1), 585 (3), 600 (1), 607 (14), 627 (1), 684 (3), 728 (7), 753 (1), 782 (3), 800 (11), 858 (6), 868 (3), 890 (6), 961 (14), 994 (21), 1022 (3), 1032 (3), 1043 (3), 1077 (10), 1094 (8), 1142 (3), 1207 (3), 1254 (6), 1285 (3), 1316 (9),
Chromosome 18 RAT|revcompl Red	18	3, 69, 169, 256, 506, 533, 651, 605, 656, 665, 677, 710, 713, 806, 908, 1027, 1094, 1136.	104 (3), 129 (9) , 152 (3), 167 (3), 183 (6), 191 (3), 223 (3), 295 (1), 305 (5), 403 (1), 452 (3), 488 (1), 585 (3), 600 (1), 607 (14), 625 (3), 684 (3), 728 (7), 753 (1), 782 (3), 800 (11), 858 (6), 868 (3), 961 (14), 994 (21), 1022 (3), 1032 (3), 1043 (3), 1077 (3), 1094 (3), 1142 (3), 1194 (46), 1254 (6), 1285 (3), 1306 (4), 1316 (5),
Chromosome 9 RAT	17	69, 122, 158, 221, 227, 409, 514, 646, 733, 806, 875, 908, 965, 986, 1001, 1031, 1166.	105 (2), 129 (9) , 152 (3), 167 (3), 183 (6), 191 (3), 212 (3), 223 (18), 305 (5), 318 (15), 403 (5), 452 (3), 488 (1), 585 (3), 600 (1), 607 (14), 625 (3), 680 (15), 728 (7), 753 (1), 782 (3), 800 (11), 858 (6), 868 (3), 961 (14), 994 (21), 1022 (3), 1032 (3), 1043 (3), 1077 (3), 1094 (7), 1142 (3), 1166 (9), 1207 (3), 1254 (3),
Chromosome 13 RAT|revcompl	14	169, 223, 256, 389, 455, 506, 533, 733, 908, 986, 995, 1010, 1031, 1040.	104 (3), 129 (9) , 152 (3), 167 (3), 183 (6), 191 (3), 223 (18), 305 (5), 330 (3), 403 (1), 452 (3), 488 (1), 574 (16), 600 (1), 607 (14), 625 (3), 684 (3), 728 (7), 753 (1), 782 (3), 800 (20), 858 (6), 868 (3), 1022 (3), 1032 (3), 1043 (3), 1077 (3), 1094 (6), 1142 (3), 1207 (3),
Chromosome 17 Mouse	14	377, 521, 539, 665, 698, 701, 794, 875, 890,	100 (9), 129 (9) , 152 (3), 167 (3), 183 (6), 191 (3), 305 (5), 331 (2), 397 (5), 403 (1), 452 (3), 488 (1), 585 (3), 600

		1015, **1027,** 1033, 1090, 1164.	(1), 607 (14), 625 (3), 684 (3), 728 (7), 753 (1), 782 (3), 800 (11), 858 (6), 868 (3), 961 (14), 994 (21), 1022 (3), 1032 (3), 1043 (3), 1077 (3), 1142 (3), 1207 (3), 1254 (6), 1285 (3), 1306 (3), 1316 (5),

* The position of the last element of the amino acid

|revcompl : Reverse complement

_ missing sequence so it appears as stop codon

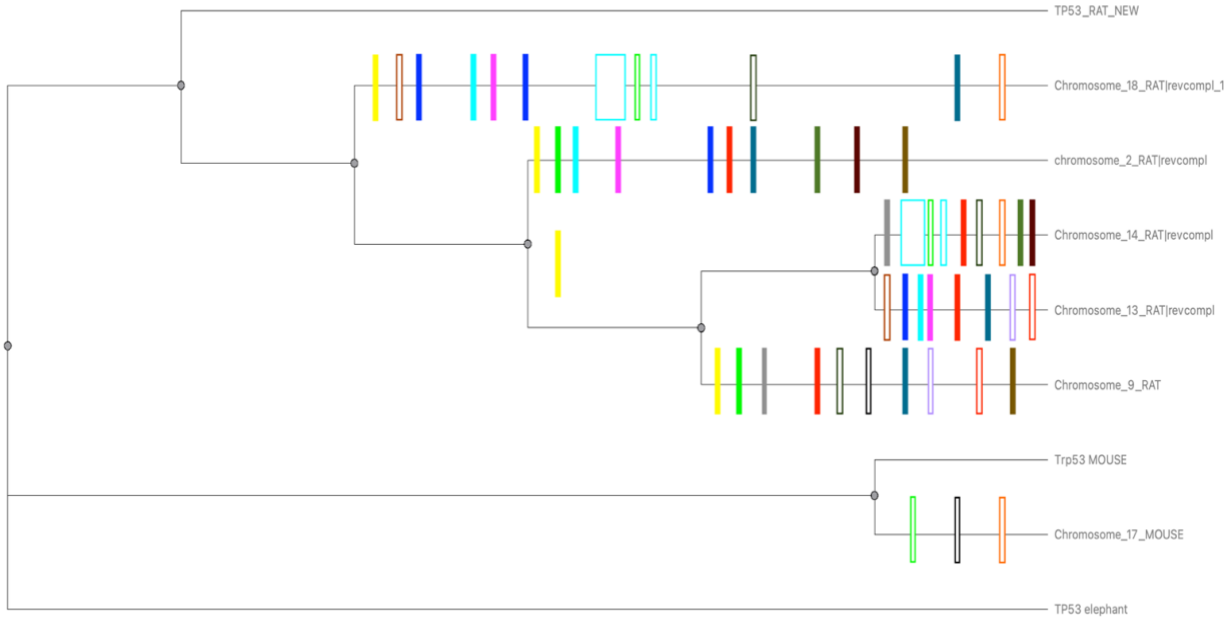

Figure 9: the stop codons tree and the position of spilt for Rat and Mouse

Conclusion

According to the information obtained from the TP53 of the Tenrec (Echinops telfairi) and Manatee (Trichechus manatus latirostris), each of these spices has one retrogene (TP53RTG). Tenrec retrogenes have seven stop codons on the sequences, while Manatee retrogenes have ten stop codons on the sequences. The retrogenes for Tenrec are missing the first 266 bases of the sequences, as shown in the UGENE file. We also see something for Manatee retrogenes where 35 bases of the sequences are missing.

All the research done agrees on a number of retrogenes for the species mentioned in the Sulak et al paper, but not for the Tasmanian devil. All the retrogenes in rats and mouse form in a single event, while microbats form in three events. Test our hypothesis to show that these retrogenes do not help in cancer because of the number of the stop codons that this retrogene has. These retrogenes are not fully functional, and they do not serve as a duplicate of the protein's lost function. Some of these retrogenes are not thought to be protein-coding genes.

Reference

Abegglen, L. M., Caulin, A. F., Chan, A., Lee, K., Robinson, R., Campbell, M. S., ... & Schiffman, J. D. (2015). Potential mechanisms for cancer resistance in elephants and comparative cellular response to DNA damage in humans. Jama, 314(17), 1850-1860.

Armstrong, S. (2014). P53: the gene that cracked the cancer code. Bloomsbury Publishing.

Cancer statistics. September 25, 2020, from https://www.cancer.gov/abouthttps://www.cancer.gov/about-cancer/understanding/statisticscancer/understanding/statistics.

Aubrey, B. J., Strasser, A., & Kelly, G. L. (2016). Tumor-suppressor functions of the TP53 pathway. Cold Spring Harbor perspectives in medicine, 6(5), a026062.

Bai, Yongsheng, Claudio Casola, and Esther Betrán. "Evolutionary origin of regulatory regions of retrogenes in Drosophila." *BMC genomics* 9.1 (2008): 1-9

Cancer genomics. National Cancer Institute. (2017). from https://www.cancer.gov/research/areas/genomics.

Caspari, T. (2000). Checkpoints: how to activate p53. Current biology, 10(8), R315-R317.

DNA sequencing fact sheet. Aug 16, 2020. DNA sequencing determines the order of the four chemical building blocks - called "bases" - that make up the DNA molecule. from https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Fact-Sheet#:~:text=Sequencing%20DNA%20means%20determining%20the,in%20a%20particular%20DNA%20segment.

Gorbunova, V., Seluanov, A., Zhang, Z., Gladyshev, V. N., & Vijg, J. (2014). Comparative genetics of longevity and cancer: insights from long-lived rodents. *Nature Reviews Genetics*, *15*(8), 531-540.

Hance, J. (2020, July 28). Vanishing point: Bumblebee bat is World's smallest; it's also at risk. Mongabay Environmental News. Retrieved May 13, 2022, from https://news.mongabay.com/2016/12/vanishing-point-bumblebee-bat-is-worlds-smallest-its-also-at risk/#:~:text=Informally%20known%20as%20the%20bumblebee,the%20weight%20of%20two%20Skittles.

Levine, A. J., Momand, J., & Finlay, C. A. (1991). The p53 tumour suppressor gene. Nature, 351(6326), 453-456.

Nunney, L. (1999). Lineage selection and the evolution of multistage carcinogenesis. Proceedings of the Royal Society of London. Series B: Biological Sciences, 266(1418), 493-498.

Pappas, S. (2018, January 30). *Weird: Naked mole rats don't die of old age*. LiveScience. from https://www.livescience.com/61568-naked-mole-rats-no-aging.html.

Peto, R. 1977 Epidemiology, multistage models, and short-term mutagenicity tests. In The origins of human cancer (ed. H. H. Hiatt, J. D. Watson & J. A. Winsten), pp. 1403^1428. Cold Spring Harbor Conferences on Cell Proliferation, vol. 4. NY: Cold Spring Harbor Laboratory Press.

Saraswathy, N., & Ramalingam, P. (2011). “Genome Sequencing methods”. Concepts and techniques in genomics and proteomics. Elsevier.

Shackney, S. E., & Silverman, J. F. (2003). Molecular evolutionary patterns in breast cancer. Advances in anatomic pathology, 10(5), 278-290.

Sigal, A., & Rotter, V. (2000). Oncogenic mutations of the p53 tumor suppressor: the demons of the guardian of the genome. Cancer research, 60(24), 6788-6793.

Smith, F. A., & Lyons, S. K. (2011). How big should a mammal be? A macroecological look at mammalian body size over space and time. Philosophical Transactions of the Royal Society B: Biological Sciences, 366(1576), 2364-2378.

Sulak, M., Fong, L., Mika, K., Chigurupati, S., Yon, L., Mongan, N. P., ... & Lynch, V. J. (2016). TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. elife, 5, e11994.

Tian, X., Azpurua, J., Hine, C., Vaidya, A., Myakishev-Rempel, M., Ablaeva, J., ... & Seluanov, A. (2013). High-molecular-mass hyaluronan mediates the cancer resistance of the naked mole rat. *Nature*, *499*(7458), 346-349.

Truskey, G. A., Yuan, F., & Katz, D. F. (2004). Transport phenomena in biological systems, 43-44.

Varley, J. M., McGown, G., Thorncroft, M., James, L. A., Margison, G. P., Forster, G., ... & Birch, J. M. (1999). Are there low-penetrance TP53 alleles? Evidence from childhood adrenocortical tumors. The American Journal of Human Genetics, 65(4), 995-1006.

Vazquez, J. M., & Lynch, V. J. (2021). Pervasive duplication of tumor suppressors in Afrotherians during the evolution of large bodies and reduced cancer risk. *Elife*, *10*, e65041.

Venner, James. (2018). Longevity: a look at the shortest and longest lives in the animal kingdom. Zooportraits. https://www.zooportraits.com/animal-longevity/#:~:text=According%20to%20the%20longevity%20records,found%20across%20South%20East%20Asia.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews genetics, 10(1), 57-63.

What is cancer? American Association for Cancer Research (AACR). (2020, October 15). from https://www.aacr.org/patients-caregivers/about-cancer/what-is-cancer/