

# Impact of Self-Induced Mutations on Skin Protein

**\* Ms. Sunanda Pandita**

School of Computer Science and Engineering, Sandip University Nashik  
\*sunanda7723@gmail.com

**Pravin Gundalwar**

School of Computer Science and Engineering, Sandip University Nashik  
gundalwarp@gmail.com

## Abstract

The human skin is the biggest organ in the body and serves as the body's first line of defence against external aggressions thanks to its strategic placement at the boundary between the interior and exterior. It is becoming more and more obvious that mutations in numerous cutaneous-associated keratin genes cause a variety of genetic skin diseases, each of which is characterised by compromised specific cell-tissue integrity, impairing the skin's capacity to function as a proper barrier and withstand repeated physical trauma. Several other concepts, including identity, similarity, and homology, are frequently (mis)used to describe sequence comparisons. Even though they are frequently used interchangeably, their meanings are very distinct. It is necessary to find the right alignment of two sequences before computing their similarity. The ability to recognise sequence similarity allows us to save countless biological years to apply knowledge of one sequence to other comparable sequences.

**Keywords:** Sequence, patterns, alignment, identity, similarity, homology

## Introduction

The structure but not usually the function of a protein can be impacted by changes in the amino acid sequence. A modification at some places, like conserved residues, however, can have an impact on the protein's structure as well as its functionality. This study illustrates how the initial amino acid sequence of a skin protein changes when mutations are introduced at various rates.

### Subjects and samples

- a) Homo Sapiens- Homo sapiens (Latin for "wise man") is the name given to the human species. Homo sapiens is the sole living member of the genus Homo, which contains multiple extinct species.
- b) Gorilla- Gorillas are giant apes that are herbivorous and spend most of their time on the ground in the tropical forests of equatorial Africa. There are two species of gorillas, the eastern and western, and up to five subspecies within the genus gorilla.
- c) Nomascus leucogenya- (Northern White Cheeked Gibbon) is a species of [primates](#) in the family [gibbons](#). They are listed as critically endangered by IUCN and in cites appendix i.

They are native to Asia. They are diurnal herbivores. Individuals can grow to 545 mm. Reproduction is viviparous. They have parental care (female provides care and paternal care).

- d) *Pongo abelii*- There are three different kinds of orangutans, and one of them is the Sumatran orangutan. Rarer than Bornean orangutans but more prevalent than the recently discovered Tapanuli orangutans, both of which may be found on Sumatra, this species is critically endangered and is located exclusively in the island's northern regions.
- e) *Hylobates moloch*- The silvery gibbon, or Javan gibbon, is a primate belonging to the genus *Hylobatis* and subfamily *Hylobatidae*. It can only be found on the Indonesian island of Java, which is in which it lives in the unspoiled rainforests at elevations of up to 2,450 meters.
- f) *Pan troglodytes*- Chimpanzees, or chimps as they are more commonly known, are a great ape species found across the tropical forests and savannahs of Africa. There are four recognized subspecies and a possible fifth.
- g) *Rhinopithecus roxallana*- Snub-nosed monkeys are a group of Old World monkeys and make up the entirety of the genus *Rhinopithecus*. The genus is rare and not fully researched. These monkeys are named for the short stump of a nose on their round faces, with nostrils arranged forward. They have relatively multicolored and long fur, particularly at the shoulders and backs.
- h) *Rhinopithecus bieti*- The black-and-white snub-nosed monkey (*Rhinopithecus bieti*), also known as the Yunnan snub-nosed monkey, is a large black and white primate that lives only in the southern Chinese province of Yunnan, where it is known to the locals as the Yunnan golden hair monkey. It is threatened by habitat loss, and is considered an endangered species. With their unique adaptations to their environment, these monkeys thrive at extreme altitudes despite the below freezing temperatures and thin air.<sup>1</sup>
- i) *Macaca fascicularis*- The crab-eating macaque is a species of cercopithecine primate found in Southeast Asia. It is additionally referred to as the long-tailed macaque or the cynomolgus monkey in research settings. The crab-eating macaque, a type of macaque, has coexisted with humans for a very long time.
- j) *Macaca mulatta*- One kind of Old World monkey is the rhesus macaque, more often known as the rhesus monkey. Six to nine recognized subspecies exist, with one group originating in China and the other in India.

## Simulation Tool

Computational biologists employ numerous techniques, including similarity searches and sequence comparison. This method is commonly used by different kinds of biologists. The most popular tool for this is BLAST (basic local alignment search technique), which compares pairs of sequences to look for local similarities. The original article that developed

BLAST has received more than 12,000 citations in the 11 years after it was published, and using BLAST has become an essential technique in biology. Understanding how it functions, what it accomplishes, how to use it effectively, and how to evaluate previously published results are essential. Among the several current iterations of the BLAST algorithm, NCBI BLAST and WU-BLAST are the most recent. The National Centre for Biotechnology Information (NCBI) offers NCBI BLAST, whereas Washington University in St. Louis offers WU-BLAST. The principles, procedures, applications, and potential concerns of the NCBI version of BLAST are all discussed.

### Performance parameter

- **Maximum Score** - The Maximum Score represents the highest possible alignment score (bit-score) that may be achieved between the query sequence & the database segments. It has a relationship with the e-value that is inversely proportional.
- **Total Score** - The term "Total Score" refers to the sum of the alignment scores for all of the sequences contained in the shared database.
- **Percent Query Coverage**- The term "Percent Query Coverage" refers to the proportion of the total length of the query that may be located in the segments that have been aligned.
- **E-value**- The e-value provides a measure of the likelihood that the similarity in sequences is not just the result of random chance.
- **Percent Identity**- The percentage of identity reveals how closely the query sequence is related to the sequences that have been aligned.

### Experimental Findings

ORIGINAL AA SEQUENCE OF HUMAN SKIN PROTEIN

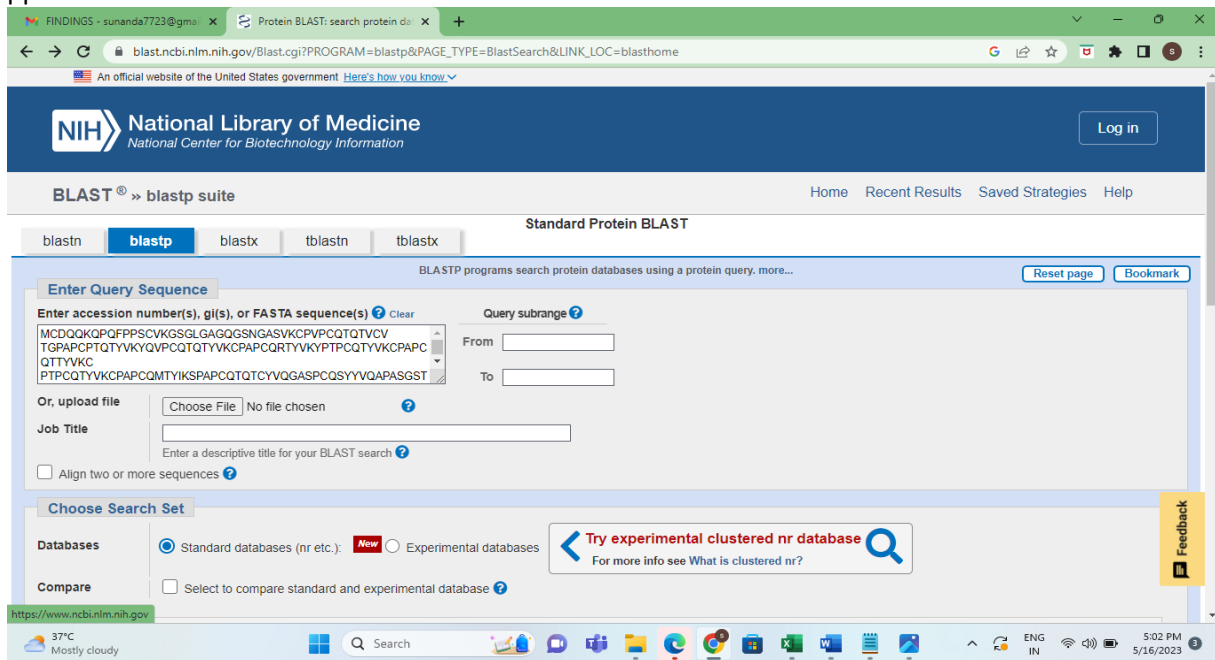
MCDQQKQPQFPPSCVKGSLGAGQGSNGASVKCPVPCQTQTVCV

TGPAPCPTQTYVKYQVPCQTQTYVKCPAPCQRTYVKYPTPCQTYVKCPAPCQTTYV  
KC

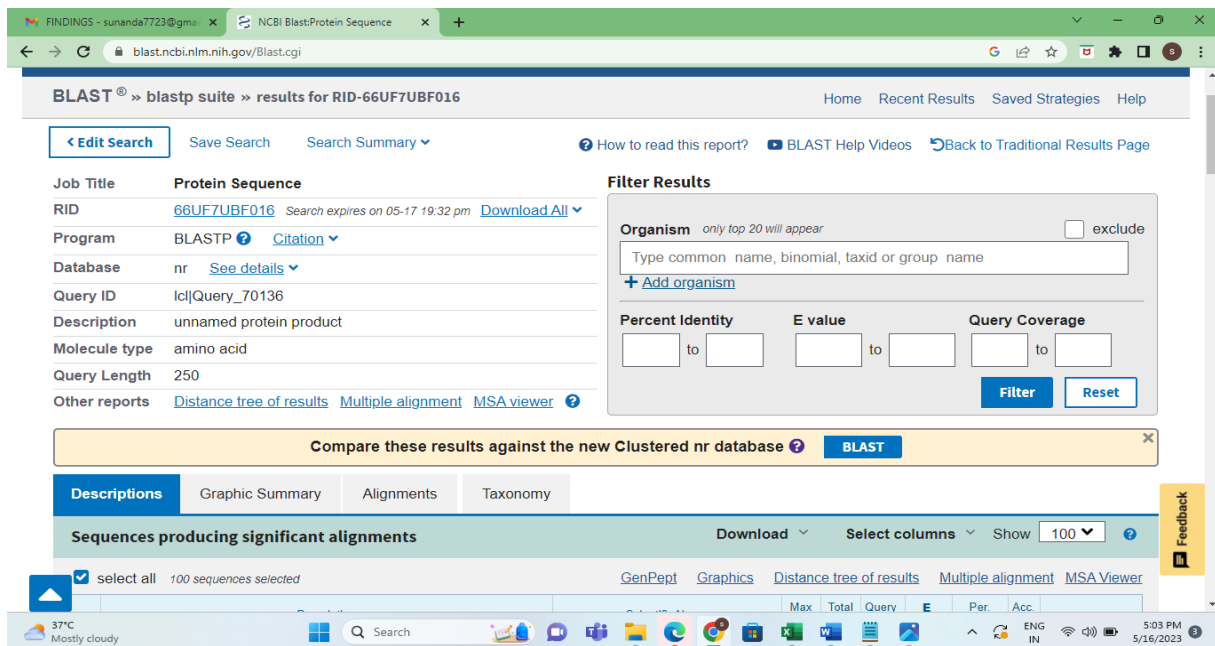
PTPCQTYVKCPAPCQMTYIKSPAPCQTQTCYVQGASPCQSYVQAPASGSTSQYCVT  
D

PCSAPCSTSYCCLAPRTFGVSPLRRWIQRPQNCNTGSSGCCENSGSSGCCGSGGCGCS

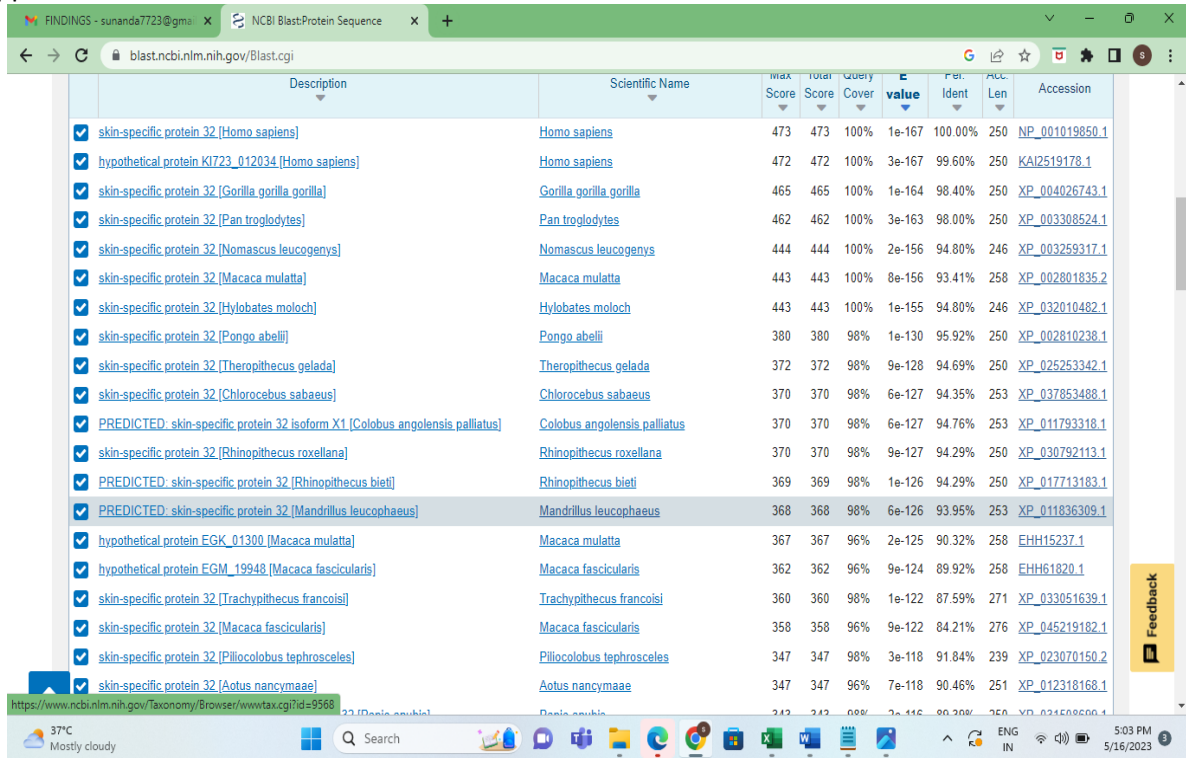
CGCGSSGCCCLGIIPMRSRGPACCDHEDDCCC



**Fig 1 BLAST Window showing Original AMINO ACID Sequence of skin protein pasted in the text box**



**Fig 2. Output 1 awaiting result**



**Fig 3. Output of original sequence**

Original

Sr. No.	Output: Scientific name	Max Score	Total Score	Query Cover	E-Value	Percentage Identity
1	Homo Sapiens	473	473	100%	1.00E-167	100%
2	Gorilla gorilla	465	465	100%	1.00E-164	99.60%
3	Pan troglotytes	462	462	100%	3.00E-163	98%
4	Nomascus Leucogenya	444	444	100%	2.00E-156	94.80%
5	Macaca mulatta	443	443	100%	8.00E-156	93.41%
6	Hylobates moloch	443	443	100%	1.00E-155	94.80%
7	Pongo abelii	380	380	98%	1.00E-130	95.92%

					130	
8	Rhinopithecus roxellana	370	370	98%	1.90E-126	94.29%
9	Rhinopithecus bieti	369	369	98%	1.00E-126	94.29%
10	Macaca fascicularis	358	358	96%	9.00E-124	89.92%

MUTATION 1 at 14 C with A at 25 G with K at 32 K with G

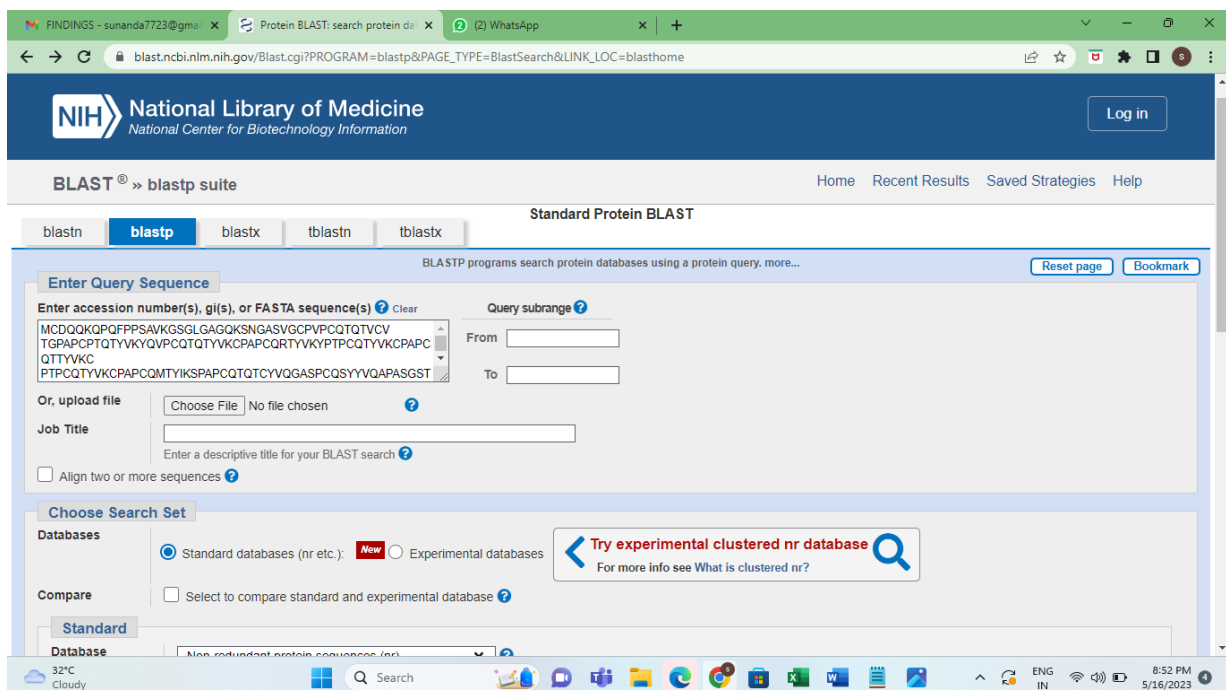
MCDQQKQPQFPPSAVKGSGGLGAGQKSNGASVGCPCVPCQTQTVCV

TGPAPCPTQTYVKYQVPCQTQTYVKCPAPCQRTYVKYPTPCQTYVKCPAPCQTTYV  
 KC

PTPCQTYVKCPAPCQMTYIKSPAPCQTQTCYVQGASPCQSYVQAPASGSTSQYCVT  
 D

PCSAPCSTSYCCLAPRTFGVSPLRRWIQRPNQNTGSSGCCENSGSSGCCGSGGCGCS

CGCGSSGCCCLGIIPMRSRGPACCDHEDDCCC



**Fig 4 MUTATION 1 BLAST QUERY SENT**



Sr. No.	Scientific name	Max Score	Total Score	Query Cover	E-Value	Percentage Identity
1	Homo Sapiens	466	466	100%	6.00E-165	98.80%
2	Gorilla gorilla	458	458	100%	8.00E-162	97.20%
3	Pan troglotytes	455	455	100%	2.00E-160	96.80%
4	Nomascus Leucogenya	372	372	96%	2.00E-127	90.42%
5	Macaca mulatta	360	360	96%	2.00E-123	89.52%
6	Hylobates moloch	370	370	96%	4.00E-127	90.42%
7	Pongo abelii	372	372	98%	1.00E-127	94.69%
8	Rhinopithecus roxellana	362	362	98%	1.00E-123	93.06%
9	Rhinopithecus bieti	362	362	98%	2.00E-123	93.06%
10	Macaca fascicularis	351	351	96%	6.00E-119	83.08%

MUTATION 2 at 5 Q with P at 10 F with T at 31 V with K

MCDQPKQPQTPPSCVKGSGLGAGQGSNGASKKCPVPCQTQTVCV

TGPAPCPTQTYVKYQVPCQTQTYVKCPAPCQRTYVKYPTPCQTYVKCPAPCQTTYV  
KC

PTPCQTYVKCPAPCQMTYIKSPAPCQTQTCYVQGASPCQSYVQAPASGSTSQYCVT  
D

PCSAPCSTSYCCLAPRTFGVSPLRRWIQRPQNCNTGSSGCCENSGSSGCCGSGGCGCS

CGCGSSGCCCLGIIPMRSRGPACCDHEDDCCC

2837



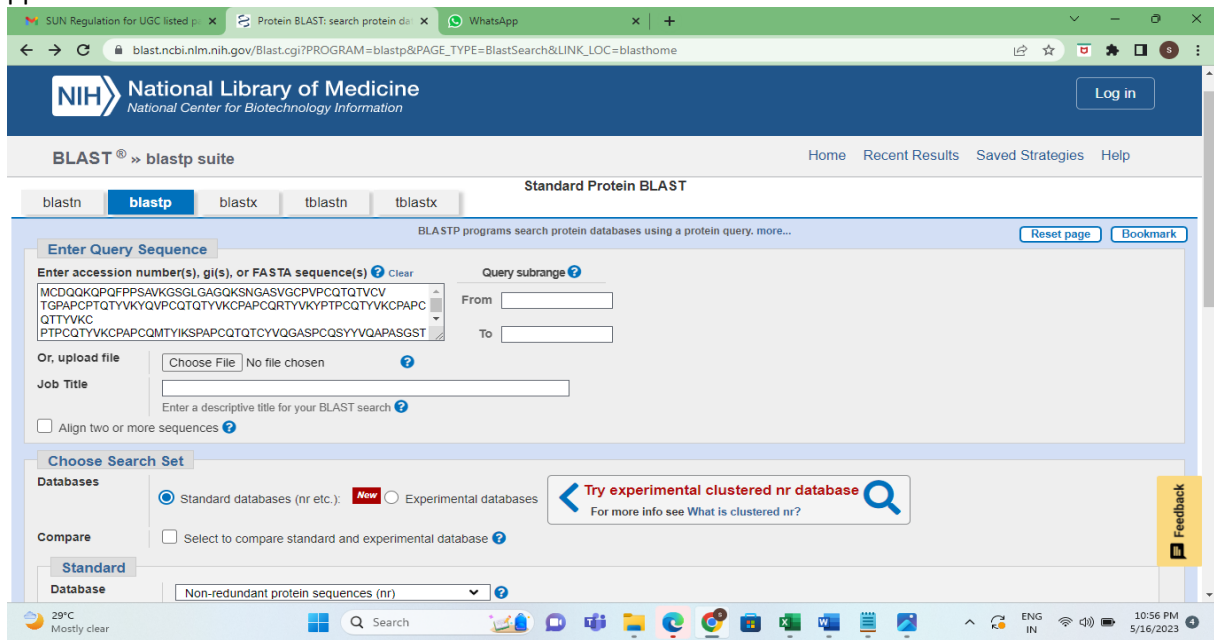


Fig 7 MUTATION 2 BLAST QUERY SENT

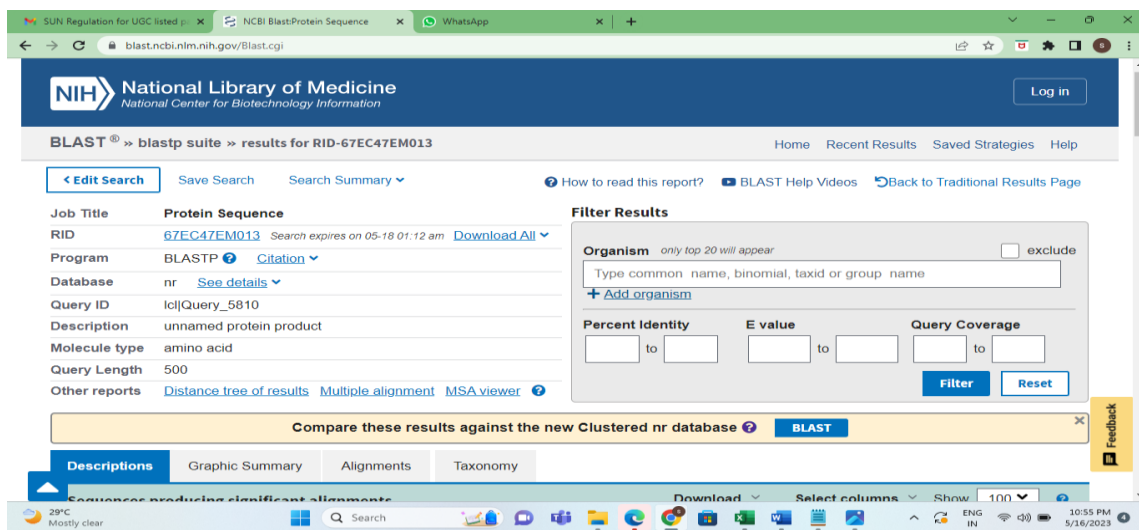


Fig 8 MUTATION 2 OUTPUT 1 RESULT AWAITED

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
skin-specific protein 32 [Homo sapiens]	Homo sapiens	465	929	100%	3e-160	98.80%	250	NP_001019880.1
hypothetical protein KI723_012034 [Homo sapiens]	Homo sapiens	464	927	100%	6e-160	98.40%	250	KA12519178.1
skin-specific protein 32 [Gorilla gorilla gorilla]	Gorilla gorilla gorilla	457	913	100%	4e-157	97.20%	250	XP_004026743.1
skin-specific protein 32 [Pan troglodytes]	Pan troglodytes	454	907	100%	6e-156	96.80%	250	XP_003308524.1
skin-specific protein 32 [Pongo abelii]	Pongo abelii	371	743	98%	2e-123	94.69%	250	XP_002810238.1
skin-specific protein 32 [Nomascus leucogenya]	Nomascus leucogenya	370	741	96%	3e-123	90.42%	246	XP_003259317.1
skin-specific protein 32 [Hylobates moloch]	Hylobates moloch	369	739	96%	9e-123	90.42%	246	XP_032010482.1
skin-specific protein 32 [Theropithecus gelada]	Theropithecus gelada	364	727	98%	2e-120	93.47%	250	XP_025253342.1
skin-specific protein 32 [Rhinopithecus roxellana]	Rhinopithecus roxellana	362	723	98%	9e-120	93.06%	250	XP_030792113.1
skin-specific protein 32 [Chlorocebus sabaeus]	Chlorocebus sabaeus	362	724	98%	1e-119	93.15%	253	XP_037853488.1
PREDICTED_skin-specific protein 32 isoform X1 [Colobus angolensis palliatus]	Colobus angolensis palliatus	362	723	98%	1e-119	93.55%	253	XP_011793318.1
PREDICTED_skin-specific protein 32 [Rhinopithecus bieti]	Rhinopithecus bieti	361	721	98%	2e-119	93.06%	250	XP_017713183.1
PREDICTED_skin-specific protein 32 [Mandrillus leucophaeus]	Mandrillus leucophaeus	360	719	98%	6e-119	92.74%	253	XP_011836309.1
skin-specific protein 32 [Macaca mulatta]	Macaca mulatta	360	719	96%	9e-119	89.52%	258	XP_002801835.2
hypothetical protein EGK_01300 [Macaca mulatta]	Macaca mulatta	358	716	96%	4e-118	89.11%	258	EHH15237.1
hypothetical protein EGM_19948 [Macaca fascicularis]	Macaca fascicularis	353	707	96%	2e-116	88.71%	258	EHH61820.1
skin-specific protein 32 [Trachypithecus francoisi]	Trachypithecus francoisi	352	704	98%	1e-115	86.47%	271	XP_033051639.1
skin-specific protein 32 [Macaca fascicularis]	Macaca fascicularis	350	699	96%	1e-114	83.08%	276	XP_045219182.1
skin-specific protein 32 [Ptilocobus tephrosceles]	Ptilocobus tephrosceles	339	678	98%	5e-111	90.61%	239	XP_023070150.2

**Fig 9 MUTATION 2 RESULT 2**

Mutation 2

Sr. No.	Scientific name	Max Score	Total Score	Query Cover	E-Value	Percentage Identity
1	Homo Sapiens	465	929	100%	3.00E-160	98.80%
2	Gorilla gorilla	457	913	100%	4.00E-157	97.20%
3	Pan troglotytes	454	907	100%	6.00E-156	96.80%
4	Nomascus Leucogenya	370	741	96%	3.00E-123	90.42%
5	Macaca mulatta	360	719	96%	9.00E-118	89.52%
6	Hylobates moloch	369	739	96%	9.00E-123	90.42%
7	Pongo abelii	371	743	98%	2.00E-123	94.69%
8	Rhinopithecus roxellana	362	723	98%	9.00E-120	93.06%
9	Rhinopithecus bieti	361	721	98%	2.00E-119	93.06%

10	Macaca fascicularis	353	707	96%	1.00E-114	88.71%
----	---------------------	-----	-----	-----	-----------	--------

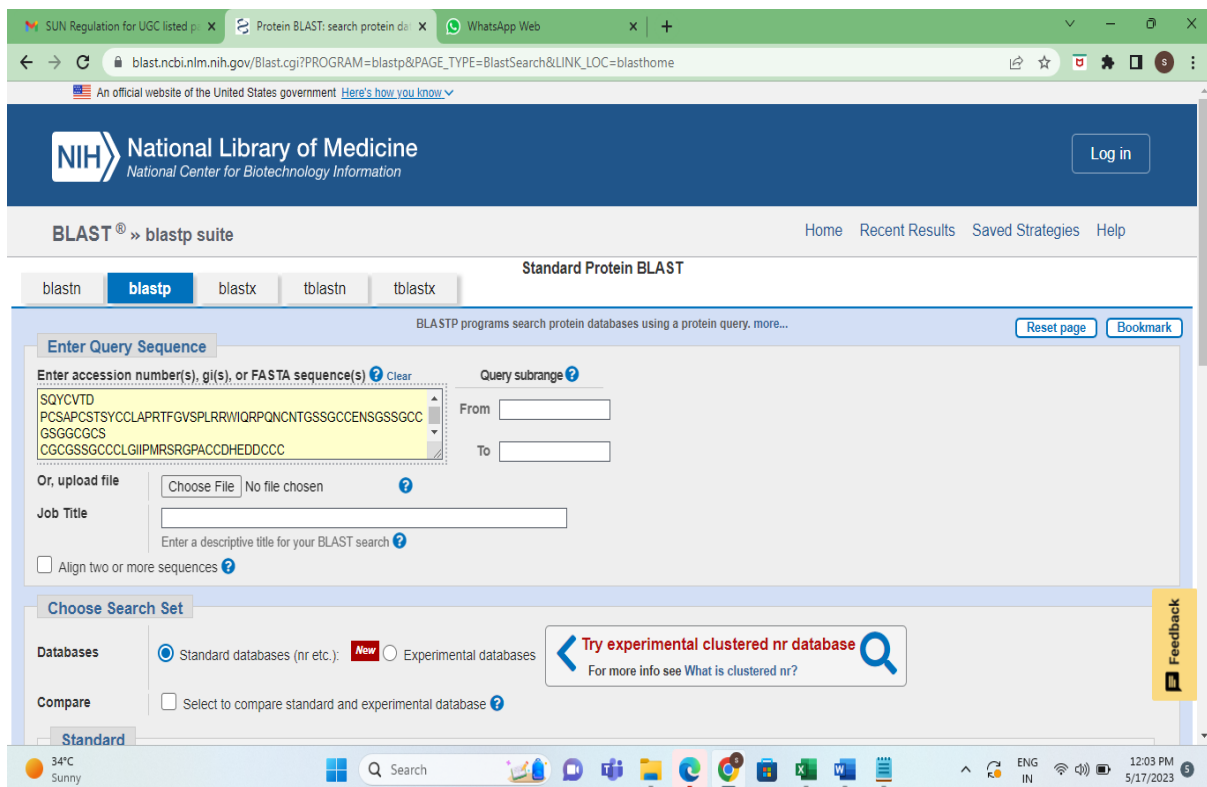
MUTATION 3 at position 2 replace C with P at position 11,12 replace P P with T T at position 43 replace C with G

MPDQQKQPQFTTSCVKGSLGAGQGSNGASVKCPVPCQTQTVGV

TGPAPCPTQTYVKYQVPCQTQTYVKCPAPCQRTYVKYPTPCQTYVKCPAPCQTTYV  
 KC

PTPCQTYVKCPAPCQMTYIKSPAPCQTQTCYVQGASPCQSYYVQAPASGSTSQYCVT  
 D

PCSAPCSTSYCCLAPRTFGVSPLRRWIQRPQNCNTGSSGCCENSGSSGCCGSGGCGCS  
 CGCGSSGCCCLGIIPMRSRGPACCDHEDDCCC



**Fig 10 MUTATION 3 BLAST QUERY SENT**

**BLAST® » blastp suite » results for RID-68XAWN5V016**

Job Title: Protein Sequence  
 RID: 68XAWN5V016  
 Program: BLASTP  
 Database: nr  
 Query ID: lcl|Query\_101480  
 Description: unnamed protein product  
 Molecule type: amino acid  
 Query Length: 250

**Filter Results**

Organism: only top 20 will appear  
 Percent Identity: [ ] to [ ]  
 E value: [ ] to [ ]  
 Query Coverage: [ ] to [ ]

**Sequences producing significant alignments**

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per Ident	Acc. Len	Accession
skin-specific protein 32 [Homo sapiens]	Homo sapiens	461	461	100%	5e-163	98.40%	250	NP_001019850.1
hypothetical protein K1723_012034 [Homo sapiens]	Homo sapiens	461	461	100%	1e-162	98.00%	250	KAJ2519178.1
skin-specific protein 32 [Gorilla gorilla gorilla]	Gorilla gorilla gorilla	454	454	100%	6e-160	96.80%	250	XP_004026743.1
skin-specific protein 32 [Pan troglodytes]	Pan troglodytes	450	450	100%	2e-158	96.40%	250	XP_003308524.1
skin-specific protein 32 [Pongo abelii]	Pongo abelii	367	367	97%	2e-125	94.65%	250	XP_002810238.1
skin-specific protein 32 [Nomascus leucogenys]	Nomascus leucogenys	366	366	96%	2e-125	90.00%	246	XP_003259317.1
skin-specific protein 32 [Hylobates moloch]	Hylobates moloch	365	365	95%	6e-125	90.34%	246	XP_032010482.1
skin-specific protein 32 [Theropithecus gelada]	Theropithecus gelada	364	364	98%	2e-124	93.47%	250	XP_025253342.1
skin-specific protein 32 [Chlorocebus sabaeus]	Chlorocebus sabaeus	358	358	98%	5e-122	92.74%	253	XP_037853488.1
PREDICTED: skin-specific protein 32 isoform X1 [Colobus angolensis palliatus]	Colobus angolensis palliatus	357	357	98%	6e-122	93.15%	253	XP_011793318.1
skin-specific protein 32 [Rhinopithecus roxellana]	Rhinopithecus roxellana	357	357	97%	9e-122	93.00%	250	XP_030792113.1
PREDICTED: skin-specific protein 32 [Rhinopithecus bieti]	Rhinopithecus bieti	357	357	97%	1e-121	93.00%	250	XP_017713183.1
skin-specific protein 32 [Macaca mulatta]	Macaca mulatta	356	356	96%	4e-121	89.11%	258	XP_002801835.2
PREDICTED: skin-specific protein 32 [Mandrillus leucophaeus]	Mandrillus leucophaeus	355	355	98%	5e-121	92.34%	253	XP_011836309.1
hypothetical protein EGK_01300 [Macaca mulatta]	Macaca mulatta	354	354	96%	1e-120	88.71%	258	EHH15237.1
hypothetical protein EGM_19948 [Macaca fascicularis]	Macaca fascicularis	350	350	96%	9e-119	88.31%	258	EHH61820.1
skin-specific protein 32 [Trachypithecus francoisi]	Trachypithecus francoisi	347	347	98%	1e-117	86.09%	271	XP_033051639.1
skin-specific protein 32 [Macaca fascicularis]	Macaca fascicularis	346	346	96%	4e-117	82.71%	276	XP_045219182.1

Fig 11. MUTATION 3 RESULT 1

**Sequences producing significant alignments**

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per Ident	Acc. Len	Accession
skin-specific protein 32 [Homo sapiens]	Homo sapiens	461	461	100%	5e-163	98.40%	250	NP_001019850.1
hypothetical protein K1723_012034 [Homo sapiens]	Homo sapiens	461	461	100%	1e-162	98.00%	250	KAJ2519178.1
skin-specific protein 32 [Gorilla gorilla gorilla]	Gorilla gorilla gorilla	454	454	100%	6e-160	96.80%	250	XP_004026743.1
skin-specific protein 32 [Pan troglodytes]	Pan troglodytes	450	450	100%	2e-158	96.40%	250	XP_003308524.1
skin-specific protein 32 [Pongo abelii]	Pongo abelii	367	367	97%	2e-125	94.65%	250	XP_002810238.1
skin-specific protein 32 [Nomascus leucogenys]	Nomascus leucogenys	366	366	96%	2e-125	90.00%	246	XP_003259317.1
skin-specific protein 32 [Hylobates moloch]	Hylobates moloch	365	365	95%	6e-125	90.34%	246	XP_032010482.1
skin-specific protein 32 [Theropithecus gelada]	Theropithecus gelada	364	364	98%	2e-124	93.47%	250	XP_025253342.1
skin-specific protein 32 [Chlorocebus sabaeus]	Chlorocebus sabaeus	358	358	98%	5e-122	92.74%	253	XP_037853488.1
PREDICTED: skin-specific protein 32 isoform X1 [Colobus angolensis palliatus]	Colobus angolensis palliatus	357	357	98%	6e-122	93.15%	253	XP_011793318.1
skin-specific protein 32 [Rhinopithecus roxellana]	Rhinopithecus roxellana	357	357	97%	9e-122	93.00%	250	XP_030792113.1
PREDICTED: skin-specific protein 32 [Rhinopithecus bieti]	Rhinopithecus bieti	357	357	97%	1e-121	93.00%	250	XP_017713183.1
skin-specific protein 32 [Macaca mulatta]	Macaca mulatta	356	356	96%	4e-121	89.11%	258	XP_002801835.2
PREDICTED: skin-specific protein 32 [Mandrillus leucophaeus]	Mandrillus leucophaeus	355	355	98%	5e-121	92.34%	253	XP_011836309.1
hypothetical protein EGK_01300 [Macaca mulatta]	Macaca mulatta	354	354	96%	1e-120	88.71%	258	EHH15237.1
hypothetical protein EGM_19948 [Macaca fascicularis]	Macaca fascicularis	350	350	96%	9e-119	88.31%	258	EHH61820.1
skin-specific protein 32 [Trachypithecus francoisi]	Trachypithecus francoisi	347	347	98%	1e-117	86.09%	271	XP_033051639.1
skin-specific protein 32 [Macaca fascicularis]	Macaca fascicularis	346	346	96%	4e-117	82.71%	276	XP_045219182.1

Fig 12 MUTATION 3 RESULT 2

Sr. No.	Scientific name	Max Score	Total Score	Query Cover	E-Value	Percentage Identity
1	Homo Sapiens	461	461	100%	5.00E-163	98.40%
2	Gorilla gorilla	454	454	100%	6.00E-160	96.80%
3	Pan troglotytes	450	450	100%	2.00E-158	96.40%
4	Nomascus Leucogenya	366	366	96%	2.00E-125	90.00%
5	Macaca mulatta	356	356	96%	4.00E-121	89.11%
6	Hylobates moloch	365	365	95%	6.00E-125	90.34%
7	Pongo abelii	367	367	97%	2.00E-125	94.65%
8	Rhinopithecus roxellana	357	357	97%	9.00E-122	93.00%
9	Rhinopithecus bieti	357	357	97%	1.00E-121	93.00%
10	Macaca fascicularis	350	350	96%	9.00E-119	88.31%

### Culminated Tabular representation of Results

Max Score

Sr. No.	Output: Scientific name	Original	Mutation1	Mutation2	Mutation3
1	Homo Sapiens	473	466	465	461
2	Gorilla gorilla	465	458	457	454
3	Pan troglotytes	462	455	454	450

4	Nomascus Leucogenya	444	372	370	366
5	Macaca mulatta	443	360	360	356
6	Hylobates moloch	443	370	369	365
7	Pongo abelii	380	372	371	367
8	Rhinopithecus roxellana	370	362	362	357
9	Rhinopithecus bieti	369	362	361	357
10	Macaca fascicularis	358	351	353	350

Total Score					
Sr. No.	Output: Scientific name	Original	Mutation1	Mutation2	Mutation3
1	Homo Sapiens	473	466	929	461
2	Gorilla gorilla	465	458	913	454
3	Pan troglotytes	462	455	907	450
4	Nomascus Leucogenya	444	372	741	366
5	Macaca mulatta	443	360	719	356
6	Hylobates moloch	443	370	739	365
7	Pongo abelii	380	372	743	367
8	Rhinopithecus roxellana	370	362	723	357
9	Rhinopithecus bieti	369	362	721	357
10	Macaca fascicularis	358	351	707	350

Query Cover
----------------

Sr. No.	Output: Scientific name	Original	Mutation1	Mutation2	Mutation3
1	Homo Sapiens	100%	100%	100%	100%
2	Gorilla gorilla	100%	100%	100%	100%
3	Pan troglotytes	100%	100%	100%	100%
4	Nomascus Leucogenya	100%	96%	96%	96%
5	Macaca mulatta	100%	96%	96%	96%
6	Hylobates moloch	100%	96%	96%	95%
7	Pongo abelii	98%	98%	98%	97%
8	Rhinopithecus roxellana	98%	98%	98%	97%
9	Rhinopithecus bieti	98%	98%	98%	97%
10	Macaca fascicularis	96%	96%	96%	96%

E-Value					
Sr. No.	Output: Scientific name	Original	Mutation1	Mutation2	Mutation3
1	Homo Sapiens	1.00E-167	6.00E-165	3.00E-160	5.00E-163
2	Gorilla gorilla	1.00E-164	8.00E-162	4.00E-157	6.00E-160
3	Pan troglotytes	3.00E-163	2.00E-160	6.00E-156	2.00E-158
4	Nomascus Leucogenya	2.00E-156	2.00E-127	3.00E-123	2.00E-125
5	Macaca mulatta	8.00E-156	2.00E-123	9.00E-118	4.00E-121
6	Hylobates moloch	1.00E-155	4.00E-127	9.00E-123	6.00E-125

7	Pongo abelii	1.00E-130	1.00E-127	2.00E-123	2.00E-125
8	Rhinopithecus roxellana	1.90E-126	1.00E-123	9.00E-120	9.00E-122
9	Rhinopithecus bieti	1.00E-126	2.00E-123	2.00E-119	1.00E-121
10	Macaca fascicularis	9.00E-124	6.00E-119	1.00E-114	9.00E-119

Percentage Identity					
Sr. No.	Output: Scientific name	Original	Mutation1	Mutation2	Mutation3
1	Homo Sapiens	100%	98.80%	98.80%	98.40%
2	Gorilla gorilla	99.60%	97.20%	97.20%	96.80%
3	Pan troglodytes	98%	96.80%	96.80%	96.40%
4	Nomascus Leucogenya	94.80%	90.42%	90.42%	90.00%
5	Macaca mulatta	93.41%	89.52%	89.52%	89.11%
6	Hylobates moloch	94.80%	90.42%	90.42%	90.34%
7	Pongo abelii	95.92%	94.69%	94.69%	94.65%
8	Rhinopithecus roxellana	94.29%	93.06%	93.06%	93.00%
9	Rhinopithecus bieti	94.29%	93.06%	93.06%	93.00%
10	Macaca fascicularis	89.92%	83.08%	88.71%	88.31%

### Conclusion

Researchers can now assess the genomic makeup of a broad variety of different animals thanks to the concurrent development of huge volumes of sequencing initiatives and bioinformatics tools like BLAST. This has aided bioinformatics in bridging the gap between computer science and biology, which is crucial for the further development of the field. As a



result, the analysis of genome sequence data using bioinformatics techniques has become an essential part of modern scientific research. BLAST is one of the most well-known techniques in the field of bioinformatics, and because it is available to all researchers online, it is commonly used to categorise sequences according to their functional and taxonomic characteristics. Applications include everything from testing to general expansion into sequence-based data, ranging from the analysis of raw sequence data to the assessment of genomic similarities.

### Future Scope

In addition to identifying genes, classifying protein structures, making gene predictions, identifying genes, and diagnosing various illnesses that influence how genes are expressed, among other things, it is possible to analyse the evidence for evolution and to extract useful patterns in gene expression. Data mining, which is crucial for pattern identification, categorization of data, for predicting events, or for induction within genetic networks, gives the ability to analyse bioinformatics information.

### References

- [1] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403-410.
- [2] **NCBI BLAST** [<http://www.ncbi.nlm.nih.gov/BLAST/>]
- [3] **WU-BLAST** [<http://blast.wustl.edu/>]
- [4] Baxevanis AD, Ouellette BFF (eds): *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*. John Wiley; 1998.
- [5] Durbin R, Eddy S, Krogh A, Mitchison G: *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge: Cambridge University Press; 1998.
- [6] Higgins D, Taylor W (eds): *Bioinformatics: Sequence, Structure and Databanks*. New York: Oxford University Press; 2000.
- [7] Kanehisa M: *Post-Genome Informatics*. New York: Oxford University Press; 2000.
- [8] Gibas L, Jambeck P: *Developing Bioinformatics Computer Skills*. Sebastopol, California: O'Reilly and Associates; 2001.
- [9] Wake DB: **Comparative terminology.** *Science* 1994, **265**:268-269.
- [10] Wake DB: **Homoplasy, homology and the problem of 'sameness' in biology.** *Novartis Found Symp* 1999, **222**:24-33.
- [11] Reeck GR, de Haen C, Teller DC, Doolittle RF, Fitch WM, Dickerson RE, Chambon P, McLachlan AD, Margoliash E, Jukes TH, *et al.*: **"Homology" in proteins and nucleic acids: a terminology muddle and a way out of it.** *Cell* 1987, **50**:667
- [12] Pearson WR, Lipman DJ: **Improved tools for biological sequence comparison.** *Proc Natl Acad Sci USA* 1988, **85**:2444-2448.
- [13] Altschul SF, Boguski MS, Gish W, Wootton JC: **Issues in searching molecular sequence databases.** *Nat Genet* 1994, **6**:119-129.
- [14] Pearson WR: **Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms.** *Genomics* 1991,

- [15] Koski LB, Golding GB: **The closest BLAST hit is often not the nearest neighbor.** *J Mol Evol* 2001, **52**:540-542.
- [16] Henikoff S, Henikoff JG: **Amino acid substitution matrices from protein blocks.** *Proc Natl Acad Sci USA* 1992, **89**:10915-10919.
- [17] Dayhoff MO, Schwartz RM, Orcutt BC: **A model of evolutionary change in proteins.** In: *Atlas of Protein Sequence and Structure*, vol.
- [18] 5. Edited by Dayhoff MO. Washington DC: National Biomedical Research Foundation; 1978:345-352.
- [19] States DJ, Gish W, Altschul SF: **Improved sensitivity of nucleic acid database searches using application-specific scoring matrices.** *Methods: A Companion to Methods in Enzymology* 1991, **3**:66-70.
- [20] Henikoff S, Henikoff JG: **Protein family classification based on searching a database of blocks.** *Genomics* 1994, **19**:97-107.
- [21] Henikoff S, Henikoff JG: **Automated assembly of protein blocks for database searching.** *Nucleic Acids Res* 1991, **19**:6565-6572.
- [22] **NCBI FTP directory - BLAST matrices**
- [23] [<ftp://ncbi.nlm.nih.gov/blast/matrices>]
- [24] Needleman SB, Wunsch CD: **A general method applicable to the search for similarities in the amino acid sequence of two proteins.** *J Mol Biol* 1970, **48**:443-453.
- [25] Smith TF, Waterman MS: **Identification of common molecular subsequences.** *J Mol Biol* 1981, **147**:195-197.
- [26] Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA, Wheeler DL: **GenBank.** *Nucleic Acids Res* 2000, **28**:15-18.
- [27] **GenBank** [<http://www.ncbi.nlm.nih.gov/Genbank/>]
- [28] Bairoch A, Apweiler R: **The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000.** *Nucleic Acids Res* 2000, **28**:45-48.
- [29] **SWISS-PROT** [<http://www.expasy.ch/sprot/>]
- [30] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
- [31] Karlin S, Altschul SF: **Applications and statistics for multiple high-scoring segments in molecular sequences.** *Proc Natl Acad Sci USA* 1993, **90**:5873-5877.
- [32] Lamperti ED, Kittelberger JM, Smith TF, Villa-Komaroff L: **Corruption of genomic databases with anomalous sequence.** *Nucleic Acids Res* 1992, **20**:2741-2747.
- [33] Kristensen T, Lopez R, Prydz H: **An estimate of the sequencing error frequency in the DNA sequence databases.** *DNA Seq* 1992, **2**:343-346.
- [34] **Full list of the BLAST Advanced options**
- [35] [[http://www.ncbi.nlm.nih.gov/BLAST/full\\_options.html](http://www.ncbi.nlm.nih.gov/BLAST/full_options.html)]
- [36] States DJ, Botstein D: **Molecular sequence accuracy and the analysis of protein coding regions.** *Proc Natl Acad Sci USA* 1991, **88**:5518-5522.
- [37] Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult

- CJ, Tomb JF, Dougherty BA, Merrick JM, *et al.*: **Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.** *Science* 1995, **269**:496-512.
- [38] Ichikawa T, Suzuki Y, Czaja I, Schommer C, Lessnick A, Schell J, Walden R: **Identification and role of adenylyl cyclase in auxin signalling in higher plants.** *Nature* 1997, **390**:698-701.
- [39] Ichikawa T, Suzuki Y, Czaja I, Schommer C, Lessnick A, Schell J, Walden R: **Identification and role of adenylyl cyclase in auxin signalling in higher plants.** *Nature* 1998, **396**:390.
- [40] Karlin S, Altschul SF: **Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes.** *Proc Natl Acad Sci USA* 1990, **87**:2264-2268.
- [41] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3007598/>