

TP53

Martin Sanchez & Ryder Sabale

Background

Understanding TP53



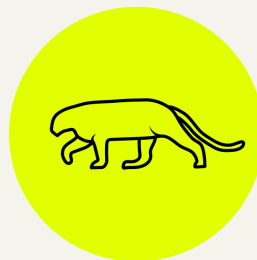
Tumor Suppressor Function

p53 protein role in DNA repair, cell cycle regulation.



Importance in Cancer Studies

Mutations linked to various cancers; potential targets for gene therapy.



Cross-Species Perspective

TP53 found across mammals, amphibians, and more. Potential insights from cancer-resistant species.



Established Research

Studies show controlled p53 regulation vital for regeneration



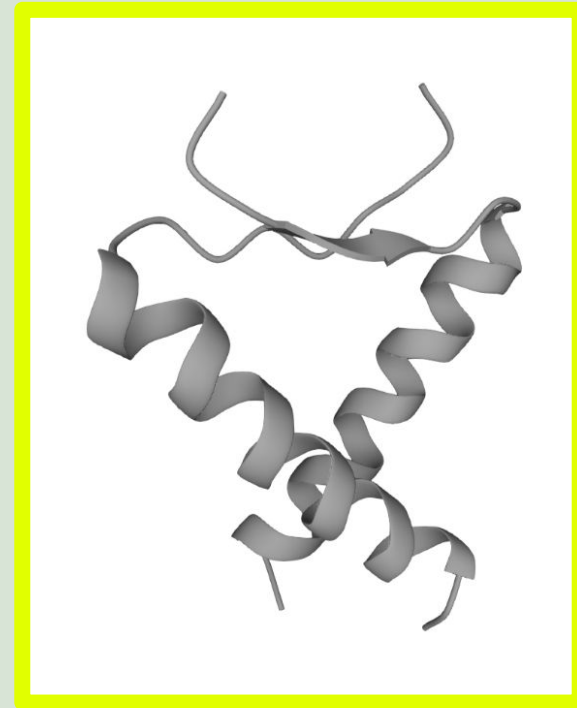
Our Approach

Gather TP53 sequences, align them, construct phylogenetic trees to reveal evolutionary patterns.

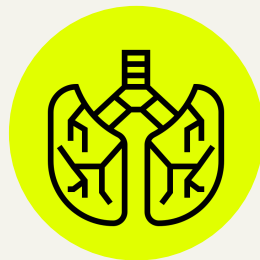
What is TP53?

Understanding it's function

- Often referred to as: “Guardian of the genome”.
- Gene function: detects damaged DNA, activates DNA repair mechanisms, and triggers cell cycle apoptosis if necessary.
- Its role helps suppress the formation of cancerous tumors.



Importance in Cancer Research



TP53 is susceptible to mutations, linked to various cancers (breast, lung, bladder) and syndromes (Li-Fraumeni)



Why understanding TP53 matters:
Improves gene therapy and cancer treatments



Cross-Species Perspective



Ambystoma Mexicanum



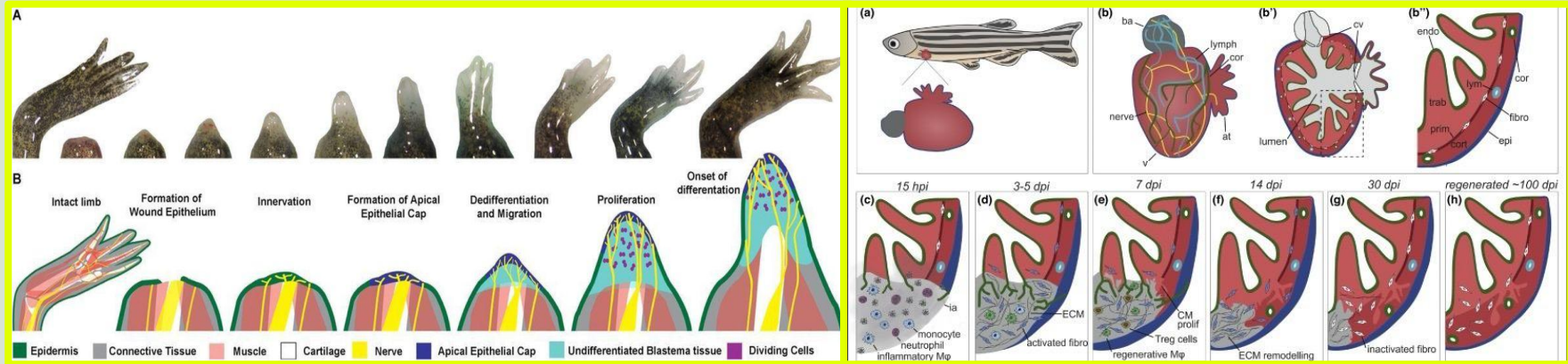
Loxodonta Africana



Podarcis muralis

- **tp53 is not unique to humans: found in axolotls, lizards, elephants, chimpanzees etc.**
- **Some species show enhanced cancer resistance or regenerative capabilities**
- **Comparing sequences may reveal evolutionary clues to cancer resistance**

Prior Research



Ambystoma Mexicanum

Danio Rerio

- A study found that controlled regulation of p53 is essential for the regeneration of limbs in salamander and fish species.
- The activity of tp53 initially decreases and then returns to baseline throughout each phase.
- Axolotl TP53 differs only by 38 amino acids; potential CRISPR applications.

Our Plan

Research questions: What increases cancer resistance?

- Is age/evolutionary stress a factor?
- Are there unique TP53 gene patterns that correlate with cancer resistance or regenerative traits?
- Does this gene vary across animal classes?
- Do similar sequences mean similar resistances?

Expected findings

- We expect to see clustering between animals within the same category.
- Clustering between animals with shared traits.

The Process

- Gather sequences of the TP53 gene of different animals.
- Align these sequences and create a phylogenetic tree, using software.
- Ascertain the evolutionary relationships of tp53 sequences.

Hypothesis

Animals with more developed/complex TP53 genes are able to counteract the susceptibility to diseases, leading to increased cancer resistance.

Data Collection

Gathering high-quality genetic data

Data Source: NCBI GenBank

- **Mammalian Sequences:** Includes various primates, horses, whales, naked mole rats, bats, etc.
- **Non-Mammalian Sequences:** Includes axolotls, salamanders, birds, fish, and reptiles.
- All sequences saved in FASTA format for compatibility with alignment tools.

Special Considerations

- Tumor Protein P53 Ortholog Database
- Primary source for wild-type TP53 sequences across various taxa.
- **Disease-Associated Sequences:** Horse TP53 (SSC) & Woodchuck TP53 (Hepatitis Virus):

Data Retrieval and Organization

- Focused on coding regions to highlight functional differences.
- Disease-associated sequences provide insights into TP53 mutations and their broader impacts.

Alignment Method

using python

Needleman Wunsch Algorithm Implementation

- **Alignment Matrix Construction:** The algorithm computes a two-dimensional scoring matrix. Scores are based on the match (+1), mismatch (-1), and gap penalties (-2) defined in our scoring system.
- **Traceback:** After building the alignment matrix, the code traces back through the matrix to find the optimal alignment path. This ensures that the best possible alignment is generated for the sequences.

Motif Sequence Calculation: A consensus or motif sequence is identified by analyzing aligned regions, highlighting conserved amino acids that are functionally significant.

- **Gapping Function:** Code refines alignments by removing unnecessary gaps, making the results more readable.

```
# STEP 3: FILL IN THE MATRIX.
match = 1
mismatch = -1
gap = -2
for i in range(1, s1Length + 1):
    for j in range(1, s2Length + 1):
        deletion = Matrix2D[i - 1][j] + gap # Deletion (gap in seq2)
        insertion = Matrix2D[i][j - 1] + gap # Insertion (gap in seq1)
        matchScore = Matrix2D[i - 1][j - 1] + (match if seq1[i - 1] == seq2[j - 1] else mismatch)
        Matrix2D[i][j] = max(deletion, matchScore, insertion) # Pick the best score

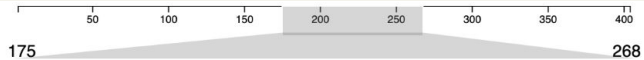
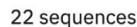
# STEP 4: TRACEBACK
align1 = []
align2 = []
i = s1Length
j = s2Length

while j and i > 0:
    currScore = Matrix2D[i][j]
    # statement handles match and mismatch case
    if Matrix2D[i - 1][j - 1] + (match if seq1[i - 1] == seq2[j - 1] else mismatch) == currScore:
        align1.append(seq1[i - 1])
        align2.append(seq2[j - 1])
        j = j - 1
        i = i - 1
    # gap in seq2
    elif Matrix2D[i - 1][j] + gap == currScore:
        align2.append("-")
        align1.append(seq1[i - 1])
        i = i - 1
    # gap in seq1
    else:
        align2.append(seq2[j - 1])
        align1.append("-")
        j = j - 1
# add all remaining gaps if ran out of characters in one sequence
while i > 0:
    align2.append("-")
    align1.append(seq1[i - 1])
    i = i - 1
while j > 0:
    align1.append("-")
    align2.append(seq2[j - 1])
    j = j - 1
# reverse the sequences to return in the correct order
align1 = ''.join(reversed(align1))
align2 = ''.join(reversed(align2))
return align1, align2
```

Alignment Interpretations

Seq1	PPPGTVRAMATYKQSQHTEVVRCP--HIER	CSDSDGLAPQHLTRVEGNLRVEYLDNDTRFHS
Seq2	PPPGTVRAMATYKQSQHTEVVRCP--HIER	CSDSDGLAPQHLTRVEGNLRVEYLDNDTRFHS
Seq3	PPPGTVRAMATYKQSQHTEVVRCP--HIER	CSDSDGLAPQHLTRVEGNLRVEYLDNDTRFHS
Seq4	PPPGTVRAMATYKQSQHTEVVRCP--HIER	CSDSDGLAPQHLTRVEGNLRVEYLDNDTRFHS
Seq5	PPPGTVRAMATYKQSFEETVVRCP--HIER	CSDSDGLAPQHLTRVEGNLRVEYLDNDTRFHS
Seq6	PPPGTVRAMATYKQSFEETVVRCP--HIER	CSDSDGLAPQHLTRVEGNLRVEYLDNDTRFHS
Seq7	TPPGTVRAMATYKQSQHTEVVRCP--HIER	CSDSDGLAPQHLTRVEGNLRVEYLDNDTRFHS
Seq8	PPPGTVRAMATYKQSQHTEVVRCP--HIER	CSDSDGLAPQHLTRVEGNLRVEYLDNDTRFHS
Seq9	PPPGTVRAMATYKQSFEETVVRCP--HIER	CSDSDGLAPQHLTRVEGNLRVEYLDNDTRFHS
Seq10	PPPNTRVAMATYKQSFEETVVRCP--HIER	CSDSDGLAPQHLTRVEGNLRVEYLDNDTRFHS
Seq11	PPPGTCVRAMATYKQSFEETVVRCP--HIER	CPDS--SDGLAPQHLTRVEGNLRVEYLDNDTRFHS
Seq12	TSTPPGTVRAMATYKQSQHTEVVRCP--HIER	CSDSDGLAPQHLTRVEGNLRVEYLDNDTRFHS
Seq13	PPPGSVRAMATYKQSQHTEVVRCP--HIER	CSDSDGLAPQHLTRVEGNLRVEYLDNDTRFHS
Seq14	PPRGTVRAMATYKQSQHTEVVRCP--HIER	CSDSDGLAPQHLTRVEGNLRVEYLDNDTRFHS
Seq15	PPPGTVRAMATYKQSQHTEVVRCP--HIER	CSDSDGLAPQHLTRVEGNLRVEYLDNDTRFHS
Seq16	PPPGTVRAMATYKQSQHTEVVRCP--HIER	CSDSDGLAPQHLTRVEGNLRVEYLDNDTRFHS
Seq17	PPPGTVRAMATYKQSQHTEVVRCP--HIER	CSDSDGLAPQHLTRVEGNLRVEYLDNDTRFHS
Seq18	PPPGTVRAMATYKQSEYMEVVRCP--HIER	CSDSDGLAPQHLTRVEGNLRVEYLDNDTRFHS
Seq19	PPPGTVRAMATYKQSEYMEVVRCP--HIER	CSDSDGLAPQHLTRVEGNLRVEYLDNDTRFHS
Seq20	PPPNTRVAMATYKQSEYMEVVRCP--HIER	CSDSDGLAPQHLTRVEGNLRVEYLDNDTRFHS
Seq21	PPPGTVRAMATYKQSQHTEVVRCP--HIER	CSDSDGLAPQHLTRVEGNLRVEYLDNDTRFHS
Seq22	PPPGTVRAMATYKQSEYMEVVRCP--HIER	CSDSDGLAPQHLTRVEGNLRVEYLDNDTRFHS

Mammalian Sequences



RAT_WILD_TP53
BOLIVIANSCUIRRELMONKEY_WILD_TP
GREENMONKEY_WILD_TP35
MACAQUEMONKEY_WILD_TP53
ORANGUTAN_WILD_TP53
HUMAN_MUTANT3302_TP53
HUMAN_MUTANT1951_TP53
GORILLA_WILD_TP53
HUMAN_WILD_TP53
CHIMPANZEE_WILD_TP53
HUMAN_MUTANT1414_TP53
NAKEDMOLERAT_WILD_TP53
WOODCHUCK_WILD_TP53
WOODCHUCK_HBX_TP53
CAT_WILD_TP35
DOG_WILD_TP53
REDFOX_WILD_TP53
HORSESHOEBAT_WILD_TP53
BISON_WILD_TP53
BLUEWHALE_WILD_TP53
HORSE_WILD_TP53

[illegible]

Non-mammalian Sequences



WALLLIZARD_WILD_TP53
ATLANTICCOD_WILD_TP53
COMMONSEADRAGON_WILD_TP53
SWANGOOSE_WILD_TP53
RATTLESNAKE_WILD_TP53
KOMODODRAGON_WILD_TP53
AXOLOTL_WILD_TP53
CHINOOKSALMON_WILD_TP53
CHANNELCATFISH_WILD_TP53
ZEBRAFISH_WILD_TP53
GOLDFISH_WILD_TP53
COMMONCARP_WILD_TP53
GREATWHITESHARK_WILD_TYPE
GIANTTORTOISE_WILD_TP53
CHINESE-ALLIGATOR_WILD_TP53
AFRICAN-OSTRICH_WILD_TP53
GOSHAWK_WILD_TP53
COMMONMALLARD_WILD_TP53
CHICKEN_WILD_TP53

```

-----FDLPFSEAG--SAGKSHSEATILPSTASPTSEPEEHFELEFEPSETASVICTYFDNLKLYQVQCTPTLKVITPTPPFAVIT
-----SITY-EDLFF-----DPPAP-LDAPTHAAPTFTTSVAEGEGRIFRQNSETA SVTSVSLNLFQCLATPTQVNLVCAFFFAAIL
-----PECFENFL-----VIAEAPVSDCVPPASTPVITTPDCELEGRINSETASVTSVSNLNLQCLAKTSPVENLKEFFPAAGVIL
-----SCTCECKIFD-----PIEPFPTNEVNNPITPTPTPTPFCSELELRQCKSETASVTSVSEETLNLQCLATSPVENLKEFFPAAGVIL
EGYPTALSDFLEN-----S-OFVPLADPQVAGEOSCFPTEDPEEGRLEFECSGASTVTSVSPDLNLFQCLATPTQVNLVCAFFFAAIL
LSFPEGR-----YSRAEENIPPLETVQFICTSSPSTEDVWNGFELVCESSGASTVICTVSPILNLFQCLATCKLNLVCLFFPAAGVIL
EG-----VSSTLNGLVGLVPIPEARL-ESEAFTSTPSTEDPTFNLTQDSCGASTVTSVSPDLNLFQCLATPTQVNLVCAFFFAAIL
VE-----FDPSLFEVSAE-PAPQSIETLDGSPSTPTPTVPEALGFLRFLQCLASTVICTVSPDLNLFQCLATPTQVNLVCAFFFAAIL
-----CDVLSCDLVCGS-----SSPPTSTPVSTPDLNLLFTHPDSSETVICTVSPDLNLFQCLATPTQVNLVCAFFFAAIL
GS-----FDNKFENVL-EEQDIP-----STLPSTSTPESSPDEKFEURLPQSETASVICTVSPDLNLFQCLATPTQVNLVCAFFFAAIL
SP-----FDNIFDNLV-TEQDIP-----STSPPTASPLATVPEEHKLGFGQSETASVICTVSNLNLQCLATPTQVNLVCAFFFAAIL
GS-----FDNKFENVL-TEQDIP-----STSPSTASPLATVPEEHKLGFGQSETASVICTVSNLNLQCLATPTQVNLVCAFFFAAIL
LEFVPLNDPLNLY-SSQAG-----FATDINVAQCTLVATTEPPEHFERLQCGQSETASVICTVSPSNLNLQCLATPTQVNLVCAFFFAAIL
-----ADPSLLPGAGSGDQVAGVLPAPPPPTTSPTSTEDVAEHFELEFGQSETASVICTVSPSNLNLQCLATPTQVNLVCAFFFAAIL
-----GLPLLEELEGVAVGLGREAPPALCTSIPTSTEDPGRFGEVAFQVQSETASVICTVSPSNLNLQCLACSPQVNLVCAFFFAAIL
-----EA-AFEWVPLQDQVAPVPLPSTEDPGRFGEVFLQGLQSETASVICTVSPSNLNLQCLACSPQVNLVCAFFFAAIL
-----GPFPGG-----VF-VPLPADPFPFPPFSPSTEDPGRHNFRLGLEAGTASVICTVSPSNLNLQCLACSPQVNLVCAFFFAAIL
-----GFPF-----STPFAAPFPPFPPSPSTEDPGRVYFGLQGETEATASVICTVSPSNLNLQCLACSPQVNLVCAFFFAAIL
GGGGGGLLAAA-----PRLVRRTEGRAGSGDQVAGVLPAPPPPTTSPTSTEDPGRFGEVAFQVQSETASVICTVSPSNLNLQCLACSPQVNLVCAFFFAAIL

```

Phylogenetic Analysis

Multiple Sequence Alignments

Mammals

Produced through R tools



Non-mammals

Produced though R tools



Union

Produced through Phylogeny.fr



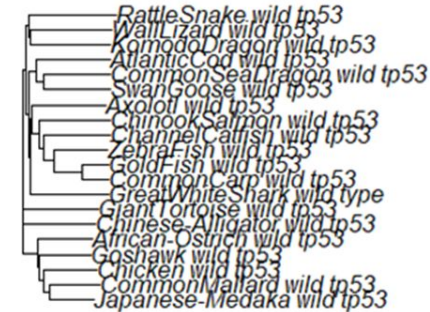
Phylogenetic Trees

Construction with R tools

Mammal Phylogenetic Tree

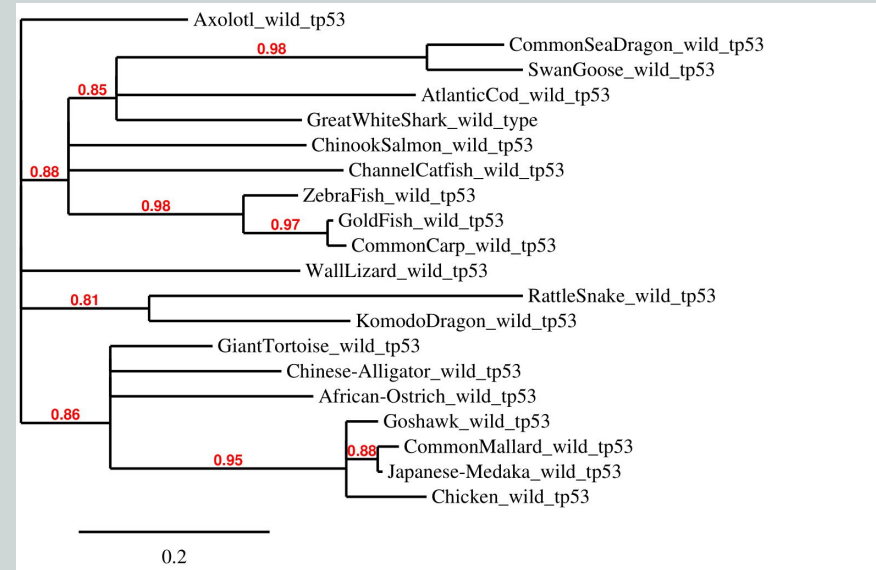
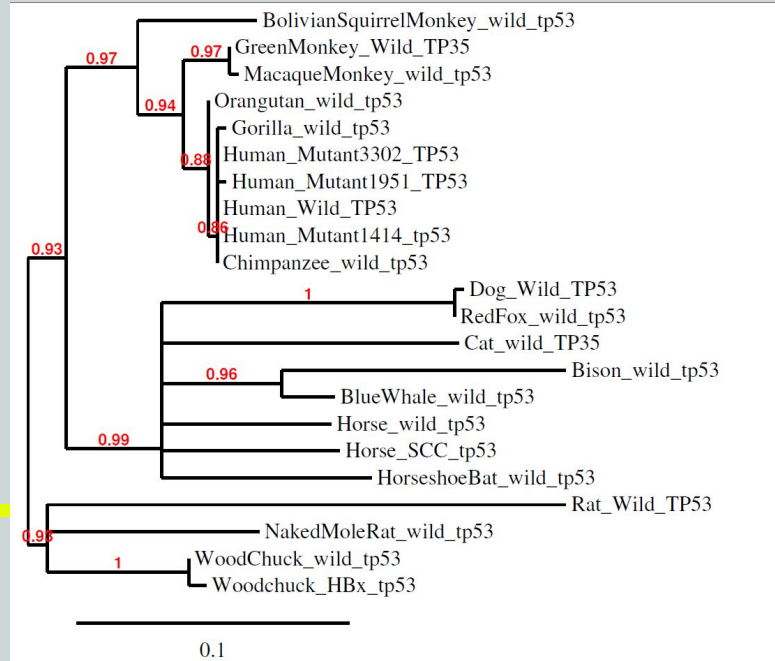


Non-Mammal Phylogenetic Tree



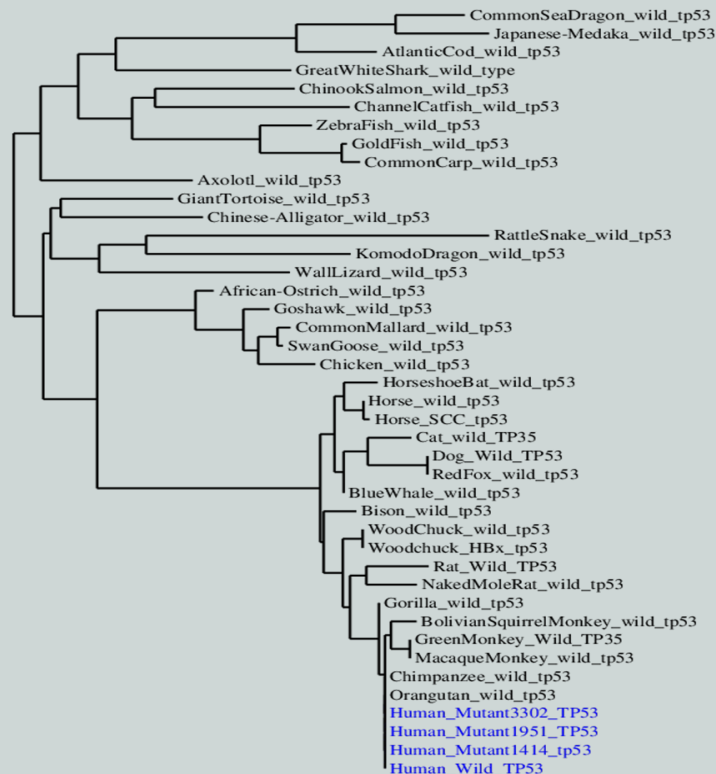
Phylogenetic Trees

Validation with Phylogeny.fr



Phylogenetic Trees

Validation with Phylogeny.fr



Findings

Taxons do not coincide with traits

Many of the taxons of the trees reveal evolutionary relationships between who do not exhibit similar cancer resistance traits (ie. Blue Whale, Bison, and Swan Goose, Mallard)

Furthermore, animals who share cancer resistant traits like the mole rat and axolotl are far removed from one another suggesting multiple types of cancer resistance methods among genes

Axolotl's unique TP53 gene

Axolotls, whose TP53 gene usage is linked to their regenerative and cancer resistance ability, seem to unique and appear as an outgroup in both trees

Limitations and Improvements

MSA Improvements:

- Some sequences in the alignment appear to have excessive or misaligned gaps. Reviewing how the gap penalties are applied in the scoring matrix, could improve accuracy.
- Potential issues in matrix initialization: If sequences are not aligning correctly at the ends or mid-regions, the problem could lie in the initial matrix setup.
- Code was very slow, taking upwards of 40 seconds to output an alignment.

Limitations:

- Linking phylogenetic analysis to traits was cloudy at best given the multiple confounding factors between evolutionary relationships and cancer resistance traits
 - Analysis of the correlation between life expectancy, evolutionary stress and cancer resistance was further stunted by this limitation

Future Work

Key Findings / Limitations

Recommendations

01 Taxons don't coincide with traits

02 Confounding factors between evolutionary relationships and traits

03 Axolotl's Unique Gene



Long Term Study of the development of certain cancers in certain environments for species with observed traits to determine how each gene deals with cancer resistance uniquely

Study comparing Axolotl's TP53 gene and upkeep to other animals with observed resistances

Thank You!

Thank you for listening! Questions?
