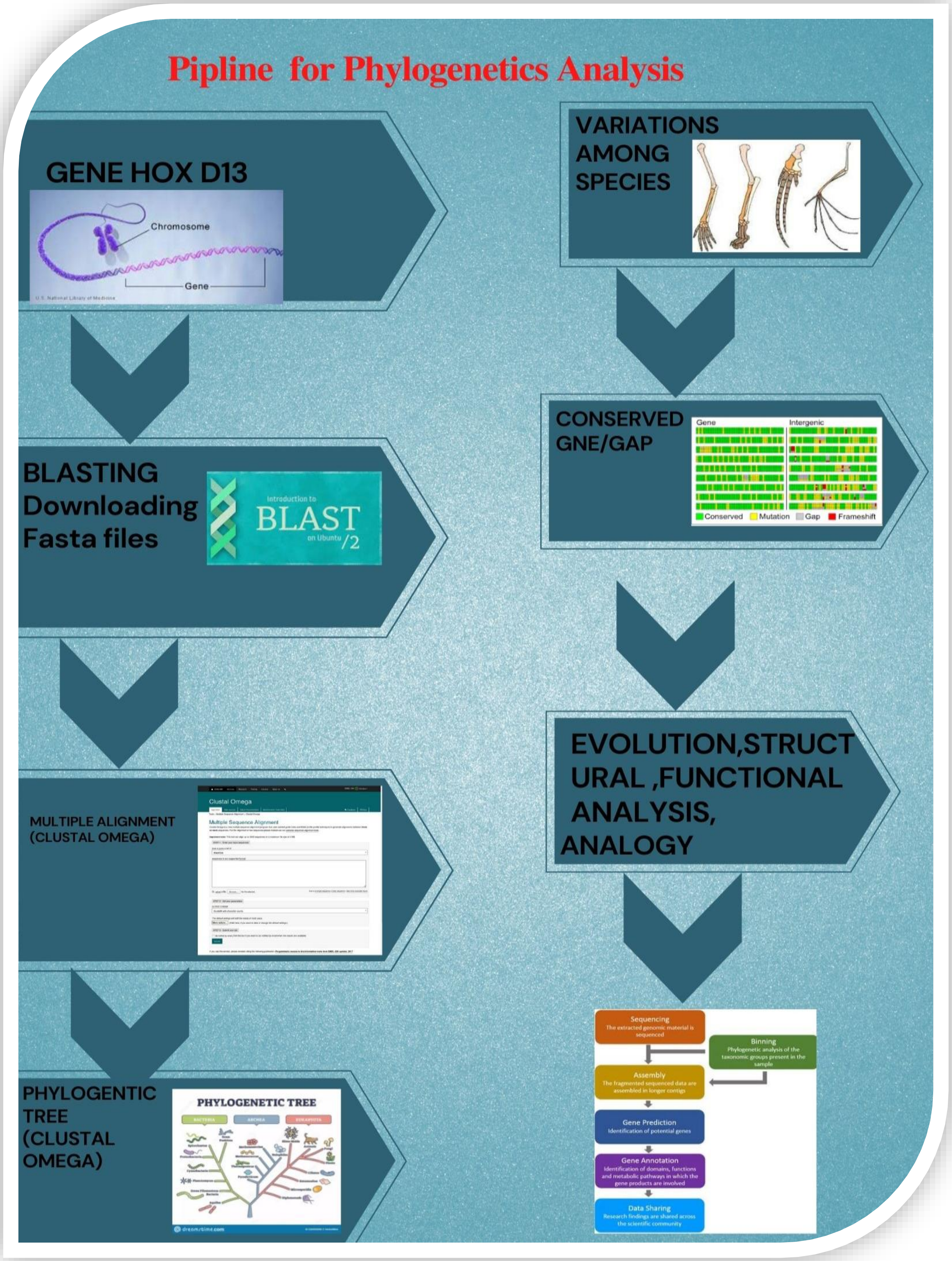


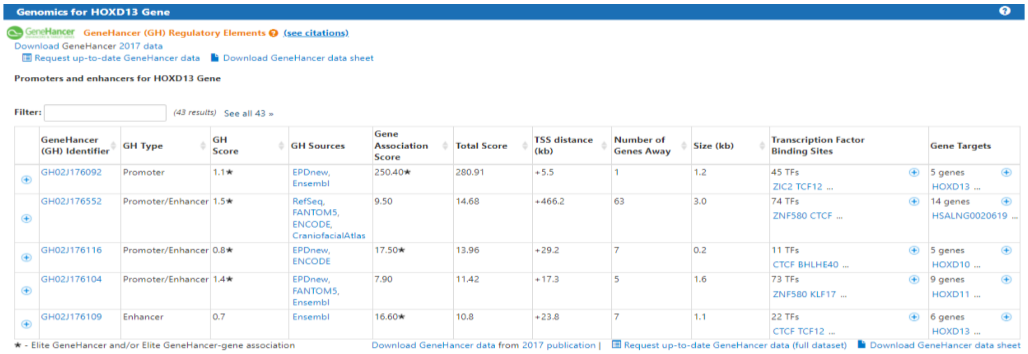
Figure 1 Graphical Abstract: Methodological Pipeline

INTRODUCTION:

The gene HOXD13 is a member of the homeobox gene family. The homeobox genes encode a highly conserved family of transcription factors involved in morphogenesis in all multicellular organisms. This gene is one of several homeobox HOXD genes found on chromosome 2. Its cytogenetic location is 2q31.1 and the genomic coordinates (GRCh38) are 2:176,087,486-176,095,943 (Gene, 2022) Homeobox (HOX) genes control patterning, differentiation, and morphogenesis during development. HOXD13 is a master regulator of autopod skeletal morphogenesis. (OMIM, 2019)

GENOMIC STRUCTURE:

The genomic structure of HOXD13 established in the study from Akarsu (1996) consists of two exons that encode a polypeptide of 335 amino acids. The downstream exon at the 3′ end of this gene contains the homeodomain sequences that are highly conserved. 63 bp upstream of this exon lies a stretch of intronic CA-repeats that are polymorphic in two different populations. The upstream exon encodes 75% of the entire protein and contains a stretch of 15 normal alanines at its 5′ end. A total of 24 alanine residues that resulted from a duplication of nine polyalanines are present at this position in the homozygous affected individuals identified. (Akarsu, 1996)



Genomics for HOXD13 Gene

GeneHancer (GH) Regulatory Elements (see citations)

Download GeneHancer 2017 data

Request up-to-date GeneHancer data Download GeneHancer data sheet

Promoters and enhancers for HOXD13 Gene

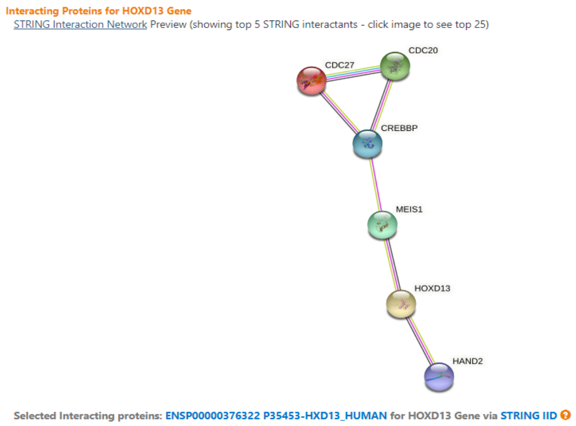
Filter: (43 results) See all 43 »

	GeneHancer (GH) Identifier	GH Type	GH Score	GH Sources	Gene Association Score	Total Score	TSS distance (kb)	Number of Genes Away	Size (kb)	Transcription Factor Binding Sites	Gene Targets
⊕	GH02J176092	Promoter	1.1★	EPDnew, Ensembl	250.40★	280.91	+5.5	1	1.2	45 TFs ⊕ ZIC2 TCF12 ...	5 genes ⊕ HOXD13 ...
⊕	GH02J176552	Promoter/Enhancer	1.5★	RefSeq, FANTOM5, ENCODE, CraniofacialAtlas	9.50	14.68	+466.2	63	3.0	74 TFs ⊕ ZNF580 CTCF ...	14 genes ⊕ HSALNG0020619 ...
⊕	GH02J176116	Promoter/Enhancer	0.8★	EPDnew, ENCODE	17.50★	13.96	+29.2	7	0.2	11 TFs ⊕ CTCF BHLHE40 ...	5 genes ⊕ HOXD10 ...
⊕	GH02J176104	Promoter/Enhancer	1.4★	EPDnew, FANTOM5, Ensembl	7.90	11.42	+17.3	5	1.6	73 TFs ⊕ ZNF580 KLF17 ...	9 genes ⊕ HOXD11 ...
⊕	GH02J176109	Enhancer	0.7	Ensembl	16.60★	10.8	+23.8	7	1.1	22 TFs ⊕ CTCF TCF12 ...	6 genes ⊕ HOXD13 ...

★ - Elite GeneHancer and/or Elite GeneHancer-gene association

Download GeneHancer data from 2017 publication | Request up-to-date GeneHancer data (full dataset) Download GeneHancer data sheet

Figure 2Promoters and enhancers for HOXD13 Gene



Interacting Proteins for HOXD13 Gene

STRING Interaction Network Preview (showing top 5 STRING interactants - click image to see top 25)

Selected Interacting proteins: ENSP00000376322 P35453-HXD13_HUMAN for HOXD13 Gene via STRING IID

Figure 3Top 5 Interacting Proteins for HOXD13 Gene

MUTATIONS IN THE *HOXD*13 GENE:

HOXD13 is the first HOX gene known to be linked to human developmental disorders. Mutations in HOXD13 are associated with limb deformities in both humans and mice, suggesting a critical role in limb development. Polyalanine repeat expansions in the coding region of the HOXD13 gene commonly causes Synpolydactyly 1 (SPD1). SPD1 is a dominantly inherited limb malformation that is characterised by an increased number of digits. Additionally, missense and out-of-frame deletion mutations in this gene are also known to cause SPD caused by haploinsufficiency. (Kurban et al., 2011) Amino acid substitutions in the homeodomain are associated with brachydactyly types D and E. In the study from Zhao et al., 2007, the link between *HOXD13* and two additional limb phenotypes; syndactyly type V and brachydactyly type A4 was established.

PHYLOGENETIC ANALYSIS:

Phylogenetic analysis is the study of the evolutionary development of a species or a group of organisms or a particular characteristic of an organism. In phylogenetic analysis, phylogenetic trees are made to represent the evolutionary history or relationship between different species, organisms, or characteristics of an organism that are developed from a common ancestor. Phylogenetic analysis is important for gathering information on biological diversity, genetic classifications, as well as learning developmental events that occur during evolution. (Dutta, 2021)

MEGA:

Molecular Evolutionary Genetics Analysis (MEGA) is computer software designed for comparative analysis of homologous gene sequences either from multigene families or from different species with a special emphasis on inferring evolutionary relationships and patterns of DNA and protein evolution. (Kumar et al., 2008) It is an integrated tool for conducting automatic and manual sequence alignment, inferring phylogenetic trees, mining web-based databases, estimating rates of molecular evolution, inferring ancestral sequences, and testing evolutionary hypotheses. (MEGA. Get the Software Safely and Easily., 2022) Data acquisition in MEGA is done by searching homologous sequences in the BLAST by using either a gene name or a query sequence. MEGA also supports sequence alignment using both the ClustalW and MUSCLE programs. (Kumar, 2004b). We have employed MEGA-X version for the purpose of our experiment.

METHODOLOGY:

Following is the methodology employed for carrying out the phylogenetic analysis.

RETRIEVAL OF GENE SEQUENCE:

The methodology for the project is started by comparing our reference sequence which is of *homo sapiens* and the name of our gene is HOXD13 with an excession id-3239 with all the other 14 sequences of different species. This query gene is present on chromosome number 2 and the locus is NC_000002 and its length is equal to 8458bp. These sequences are obtained in FASTA format from NCBI using BLAST. The FASTA sequences with their excession numbers as follows:

1. *Zootoca vivipara* (common lizard) 118092831
2. *Rana temporary* (common frog)] 120943982
3. *Gallus gallus* (chicken)-NM_205434.1
4. *Tursiops truncates* (common bottlenose dolphin)-XM_033860404
5. *Phoca vitulina* (harbor seal)- XM_032420586
6. *Falco peregrinus* (peregrine falcon)-XM_005237146.1
7. *Ursus maritimus* (polar bear)-XM_0406375
8. *Eptesicus fuscus*(big brown bat)- XM_008138753.2
9. *Thamnophis elegans* (western terrestrial garter snake)- XM_032234116.1

10. *Chlorocebus sabaeus* (green monkey)
11. *Charadrius vociferous* (killdeer)-XM_009894634.1
12. *Panthera pardus* (leopard)-NW_017619846.1
13. *Camelus dromedarius* (Arabian camel)- NC_044515.1
14. *Equus caballus* (horse)- NC_009161.3

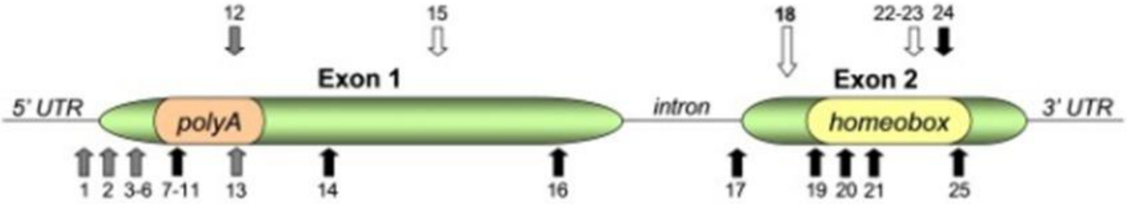
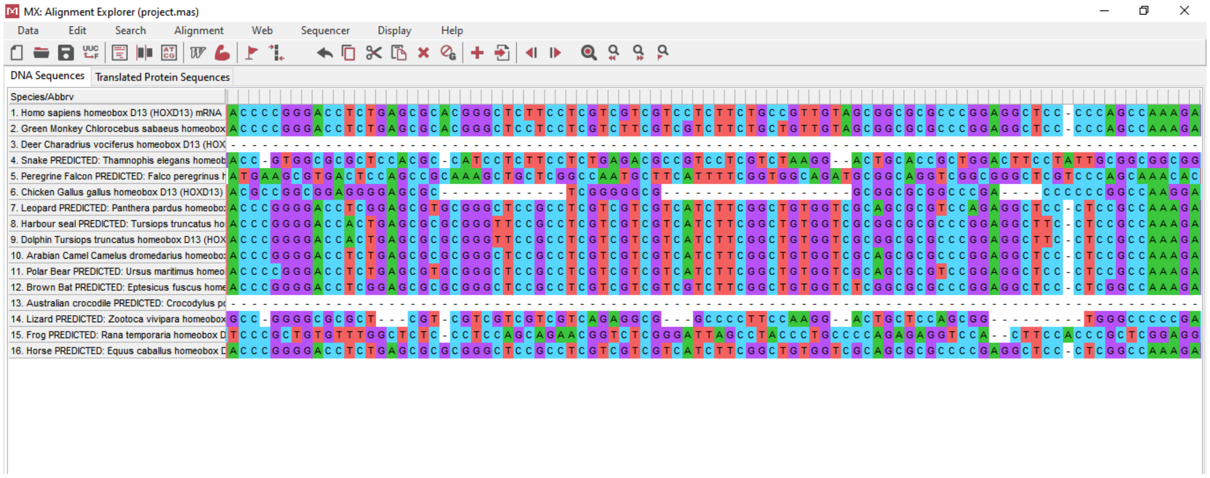


Figure 4 Query sequence gene of homosapiens

BASE-SEQUENCE ALIGNMENT:

MEGA Algorithm (https://www.megasoftware.net) is used for the alignment of retrieved sequence through multiple sequence alignment. The query sequence and templates are aligned using clustalW.



Species/Abbrv	
1. Homo sapiens homeobox D13 (HOXD13) mRNA	ACCCCGGGACCTCTGAGCGCACGGGCTCTTCCTCGTCGTCGTCCTCTTCTGCCGTTGTAGCGGCGCGCCCGGAGGCTCC-CCCAGCCAAAGA
2. Green Monkey Chlorocebus sabaeus homeobox	ACCCCGGGACCTCTGAGCGCACGGGCTCCTCCTCGTCTTCGTCGTCTTCTGCTGTTGTAGCGGCGCGCCCGGAGGCTCC-CCCAGCCAAAGA
3. Deer Charadrius vociferus homeobox D13 (HOX	--
4. Snake PREDICTED: Thamnophis elegans homeob	ACC-GTGGCGCGCTCCACGC-CATCCTCTTCCTCTGAGACGCCGTCCTCGTCTAAGG--ACTGCACCGCTGGACTTCCTATTGCGGCGGCGG
5. Peregrine Falcon PREDICTED: Falco peregrinus h	ATGAAGCGTGACTCCAGCCGCAAAGCTGCTCGGCCAATGCTTCATTTTCGGTGGCAGATGCGGCAGGTCGGCGGGCTCGTCCCAGCAAACAC
6. Chicken Gallus gallus homeobox D13 (HOXD13)	ACGCCGGCGGAGGGGAGCGC------------TCGGGGGCG------------------GCGGCGCGGCCCGA----CCCCCCGGCCAAGGA
7. Leopard PREDICTED: Panthera pardus homeobo	ACCCGGGGACCTCGGAGCGTGCGGGCTCCGCCTCGTCGTCGTCATCTTCGGCTGTGGTCGCAGCGCGTCCAGAGGCTCC-CTCCGCCAAAGA
8. Harbour seal PREDICTED: Tursiops truncatus ho	ACCCGGGGACCACTGAGCGCGCGGGTTCCGCCTCGTCGTCGTCATCTTCGGCTGTGGTCGCGGCGCGCCCGGAGGCTTC-CTCCGCCAAAGA
9. Dolphin Tursiops truncatus homeobox D13 (HOX	ACCCGGGGACCACTGAGCGCGCGGGTTCCGCCTCGTCGTCGTCATCTTCGGCTGTGGTCGCGGCGCGCCCGGAGGCTTC-CTCCGCCAAAGA
10. Arabian Camel Camelus dromedarius homeobo	ACCCGGGGACCTCTGAGCGCGCGGGCTCCGCCTCGTCGTCGTCATCTTCGGCTGTGGTCGCAGCGCGCCCGGAGGCTCC-CTCCGCCAAAGA
11. Polar Bear PREDICTED: Ursus maritimus homeo	ACCCCGGGACCTCTGAGCGTGCGGGCTCCGCCTCGTCGTCGTCATCTTCGGCTGTGGTCGCAGCGCGTCCGGAGGCTCC-CTCCGCCAAAGA
12. Brown Bat PREDICTED: Eptesicus fuscus home	ACCCGGGGACCTCGGAGCGCGCGGGCTCCGCCTCGTCGTCGTCGTCTTCGGCTGTGGTCTCGGCGCGCCCGGAGGCTCC-CTCGGCCAAAGA
13. Australian crocodile PREDICTED: Crocodylus p	--
14. Lizard PREDICTED: Zootoca vivipara homeobox	GCC-GGGGCGCGCT---CGT-CGTCGTCGTCGTCAGAGGCG---GCCCCTTCCAAGG--ACTGCTCCAGCGG---------TGGGCCCCCGA
15. Frog PREDICTED: Rana temporaria homeobox D	TCCCGCTGTGTTTGGCTCTC-CCTCCAGCAGAACGGTCTCGGGATTAGCCTACCCTGCCCCAGAGAGGTCCA--CTTCCACCCGCTCGGAGG
16. Horse PREDICTED: Equus caballus homeobox D	ACCCGGGGACCTCTGAGCGCGCGGGCTCCGCCTCGTCGTCGTCATCTTCGGCTGTGGTCGCAGCGCGCCCCGAGGCTCC-CTCGGCCAAAGA

Figure 5 Multiple Sequence Alignment using MEGA X

Then the sequence is saved in MEGA format followed by observation of conserved and variable regions using Amino acid replacing nucleotide sequence algorithm. This gives the following data:

Table 1 showing no. of sites.

Site regions	Conserved region	Variable region	parsimony-informative	singleton	Total websites
No. of sites	1112	1612	717	841	2927

Figure a

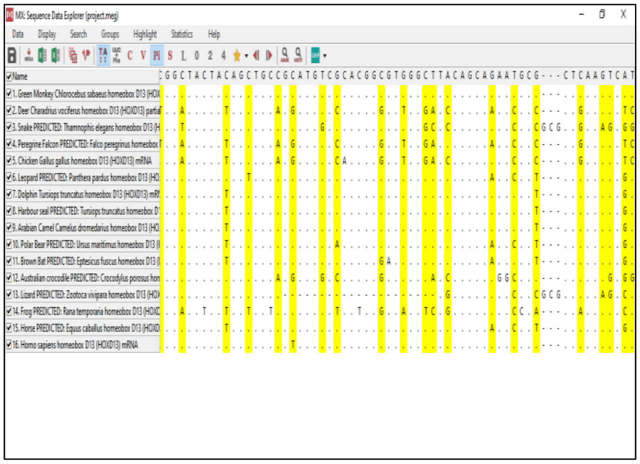


Figure b

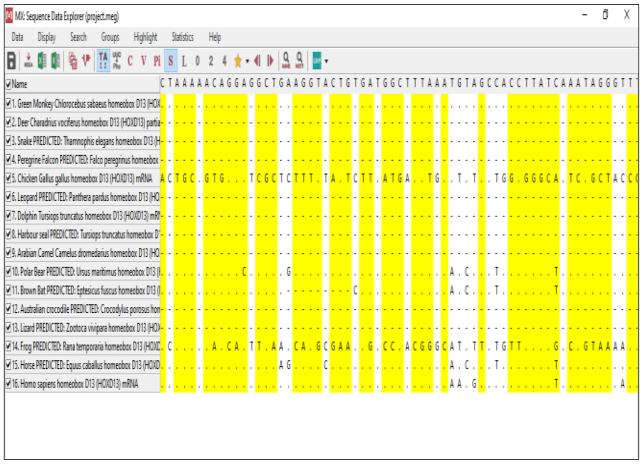


Figure c

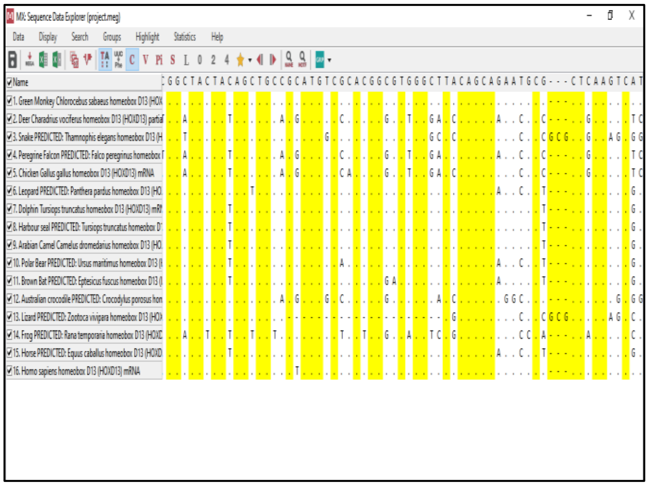


Figure d

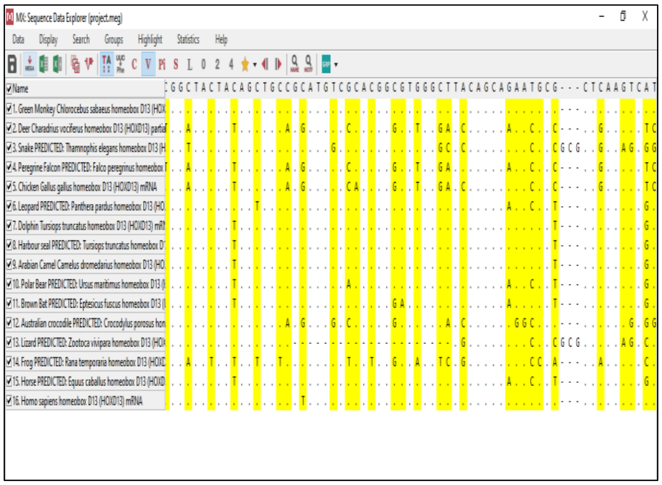
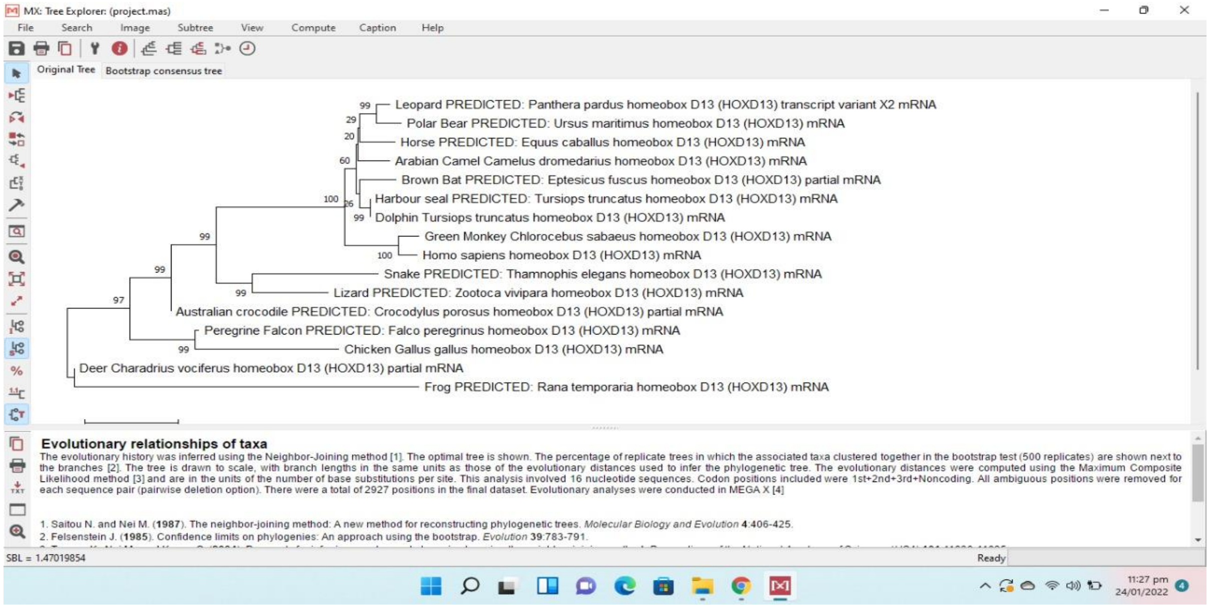


Figure 6 Patterns in gene sequence Fig (a) shows conserved regions. Fig (b) shows variable regions, Fig (c) shows Pi, Fig (d) shows Singleton.

PHYLOGENETIC RELATIONSHIPS AND TIMING OF DIVERSIFICATION:

The saved file in MEGA format which contains all the aligned sequences is followed by the construction of a phylogenetic tree. This construction is done by using MEGA software, neighbor joining method with bootstrap values are selected. The principle of bootstrap method is to find pairs of operational taxonomic units (OTUs [= neighbors]) that minimize the total branch length at each stage of clustering of OTUs starting with a star-like tree. The branch lengths as well as the topology of a parsimonious tree can quickly be obtained by using this method (Saitou & Nei, 1987). The constructed phylogenetic tree is shown in fig. 4.



Evolutionary relationships of taxa

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method [3] and are in the units of the number of base substitutions per site. This analysis involved 16 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 2927 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [4]

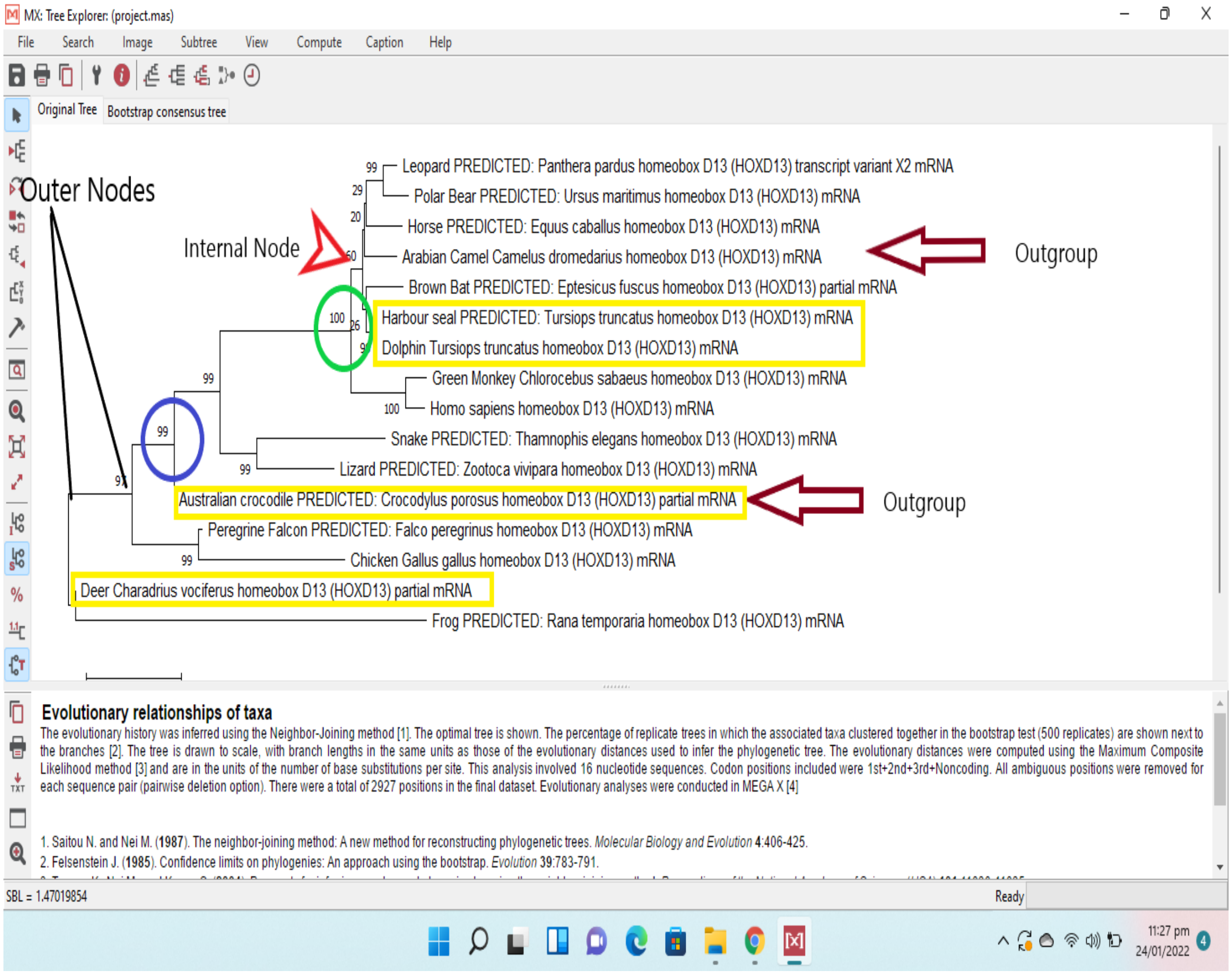
1. Saitou N. and Nei M. (**1987**). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**:406-425.
2. Felsenstein J. (**1985**). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**:783-791.

Figure 7 Phylogenetic tree of HOXD13 gene in various species

RESULTS AND DISCUSSION:

PHYLOGENETIC ANALYSIS:

After the multiple alignment of the selected species, the following phylogenetic tree was constructed using MEGAX:

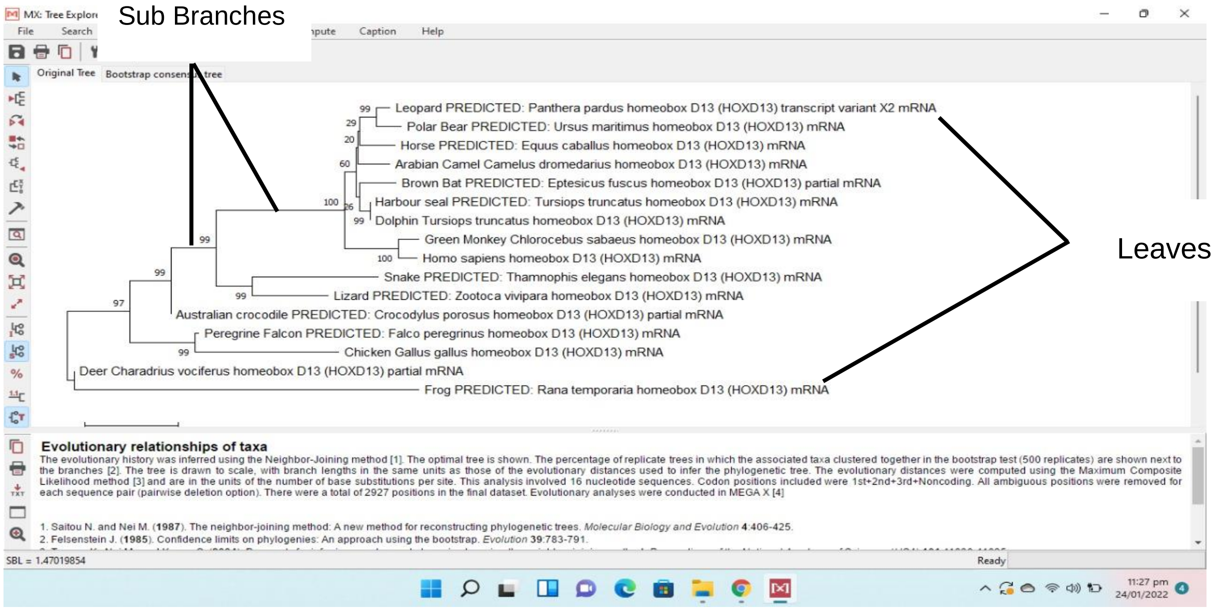


Evolutionary relationships of taxa

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method [3] and are in the units of the number of base substitutions per site. This analysis involved 16 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 2927 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [4]

1. Saitou N. and Nei M. (**1987**). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**:406-425.
2. Felsenstein J. (**1985**). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**:783-791.

Figure 8 Phylogenetic tree of 16 species showing evolution of HOXD13 gene



Evolutionary relationships of taxa

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method [3] and are in the units of the number of base substitutions per site. This analysis involved 16 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 2927 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [4]

1. Saitou N. and Nei M. (**1987**). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**:406-425.
2. Felsenstein J. (**1985**). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**:783-791.

Figure 9 Labelling of Sub Branches and Leaves

Following are the main characteristics of this tree:

OUTER NODES:

As labelled above, two nodes are bifurcating to produce the following branching pattern:

1. The upper node is branching extensively to produce several internal nodes, which keep producing sub-branches until 14 species are arranged on the leaves of the upper node. This indicates that 14 species are connected to one common ancestor.
2. Only two species are located on the lower node: *Charidrius vociferous* (Deer) and *Rana temporaria* (Frog) are located on the second node.

MAIN INTERNAL NODES:

i. The upper internal node (labelled in green) with a bootstrap value of 100 shows very extensive branching, indicating that the gene underwent multiple mutagenic events from this point onwards. It can be seen that 9 species are produced from the common ancestor at this node. Arabian camel *Camelus dromedarius* is an **outgroup** here.

ii. The lower internal node (labelled in blue) with a bootstrap value of 99 produces an **outgroup**, *Crocodilus porosus* (Australian crocodile).

IMMEDIATE SPECIATION:

As indicated in the tree, there are 4 species which have **zero** horizontal distance, indicating that they did not take any time for the gene to evolve further before the formation of the respective species. These are:

i. Harbor Seal (PREDICTED) *Tursiops truncatus*
ii. Dolphin *Tursiops truncatus*
iii. Australian Crocodile *Crocodilus porosus*
iv. Deer *Charidrius vociferous*

CLOSELY LINKED INDIVIDUALS:

The internal node (labelled with a red arrow) with boot strap value 60, shows the 7 species that are most closely related in the tree, forming sub-branches at small intervals:

iii. Leopard *Panthera pardus*
iv. Polar Bear *Ursus maritimus*
v. Horse *Equus caballus*
vi. Arabian Camel *Camelus dromedarius*
vii. Brown Bat *Eptesicus fuscus*
viii. Harbour Seal (PREDICTED) *Tursiops truncatus*
ix. Dolphin *Tursiops truncates*

Other than these, we have the following four pairs of closely linked species, excluding the outgroup:

i. Green monkey *Chlorocebus sabaeus* and Humans *Homo sapiens*
ii. Snake *Thamnophis elegans* and Lizard *Zootoca vivipara*
iii. Peregrine Falcon *Falco peregrinus* and Chicken *Gallus gallus*
iv. Deer *Charidrius vociferus* and Frog *Rana temporaria*

DISTANTLY LINKED SPECIES:

Following are some pairs of distantly linked species displayed in the tree:

i. Leopard *Panthera pardus* and Frog *Rana temporaria*
ii. Arabian camel *Camelus dromedarius* and Chicken *Gallus gallus*
iii. Brown Bat *Eptesicus fuscus* and Australian crocodile *Crocodilus porosus*

iv. Polar bear *Ursus maritimus* and Lizard *Zootoca vivipara*
v. Deer *Charidrius vociferus* and Human *Homo sapiens*

INTERPRETATION:

Based on the results of the phylogenetic tree, we can interpret the following:

i. Amongst all the species selected, the prominent outgroup is the Arabian camel, which is two-toed i.e. it has two phalanges, which is why it is separated at the internal node level before it divided further. This is a clear example of how the HOXD-13 gene underwent mutagenesis to produce bidactyl feet in camels, despite it being a Primate, which are usually pentadactyl.

ii. The Primates have been grouped in one internal node, which indicates that the gene consensus sequence was widely common in these. This is confirmed by the fact that they all have pentadactyl limbs. However, deer, also being a Primate, is distantly separated from the rest of these due to the same reason: mutagenesis. It has four digits, which separates it from its pentadactyl primate fellows.

iii. Due to epigenetic factors, the crocodile is also separated from its reptilian partners i.e. lizard and snake, despite also having five digits.

iv. Frogs, which are Amphibians, have four fingers and five toes, which is why it is grouped in the lower node with the Primate deer, which also has 4 digits.

CONCLUSION:

Due to the mutagenesis occurring at various speciation events, the 16 different species have been divided into various groups and subgroups. Based on the amount of conserved sequence that they have inherited, the species are classified as closely linked and distantly linked. For the same reasons, outgroups are also produced. Thus, it can be seen that the mutations occurring in a singular gene can produce such wide-ranging results in the structure of toes and fingers in various species.

REFERENCES:

1. Gene[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2004 – cited 2022 January 26. Available from: https://www.ncbi.nlm.nih.gov/gene/3239
2. Online Mendelian Inheritance in Man, OMIM®. Johns Hopkins University, Baltimore, MD. MIM Number: 142989: 11/08/2019: . World Wide Web URL: https://omim.org/
3. Akarsu, A. (1996). Genomic structure of HOXD13 gene: a nine polyalanine duplication causes synpolydactyly in two unrelated families. Human Molecular Genetics, 5(7), 945–952. https://doi.org/10.1093/hmg/5.7.945
4. Kurban, M., Wajid, M., Petukhova, L., Shimomura, Y., & Christiano, A. M. (2011). A nonsense mutation in the HOXD13 gene underlies synpolydactyly with incomplete penetrance. Journal of Human Genetics, 56(10), 701–706. https://doi.org/10.1038/jhg.2011.84
5. Zhao, X., Sun, M., Zhao, J., Leyva, J. A., Zhu, H., Yang, W., Zeng, X., Ao, Y., Liu, Q., Liu, G., Lo, W. H., Jabs, E. W., Amzel, L. M., Shan, X., & Zhang, X. (2007). Mutations in HOXD13 Underlie Syndactyly Type V and a Novel Brachydactyly-Syndactyly Syndrome. The American Journal of Human Genetics, 80(2), 361–371. https://doi.org/10.1086/511387
6. Dutta, S. S., PhD. (2021, March 9). What is Phylogenetic Analysis? News-Medical.Net. https://www.news-medical.net/health/What-is-Phylogenetic-Analysis.aspx
7. Kumar, S., Nei, M., Dudley, J., & Tamura, K. (2008). MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. Briefings in Bioinformatics, 9(4), 299–306. https://doi.org/10.1093/bib/bbn017
8. MEGA. Get the software safely and easily. (2022, January 20). Software Informer. https://mega.software.informer.com/
9. Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution, 4(4), 406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454