GENETIC INSIGHTS INTO EPIMORPHIC REGENERATION IN VERTEBRATES: EVOLUTIONARY PATTERNS AND IMPLICATIONS FOR REGENERATIVE MEDICINE

Komal Mahmood

Department of Pharmacology, Dow University of Health Sciences, Karachi, Pakistan
mahmoodkomal43@gmail.com

ABSTRACT

This research paper delves into the genetic underpinnings of epimorphic regeneration within the monophyletic class of vertebrates, investigating the potential implications for regenerative capabilities in this taxonomic group. Starting with a comprehensive literature review, the paper establishes the evolutionary context, from the emergence of vertebrates in the late Ordovician period to the varying regenerative abilities among species, including amphibians like the axolotl and certain lizard species.

The study employs a structured methodology, encompassing species selection, data preprocessing, a percent identity matrix analysis, Principal Component Analysis (PCA), and the construction of phylogenetic trees. Key candidate genes, BMP2 (bone morphogenetic protein 2), BMP4 (bone morphogenetic protein 4), SMAD2 (Mothers against decapentaplegic homolog 2), and Neurogenin1, are assessed for their genetic similarities across selected species, which encompass humans, mice, reptiles, amphibians, birds, and fish. The results reveal intriguing genetic relationships. This investigation advances our understanding of the genetic basis for regenerative abilities in vertebrates, shedding light on the evolutionary and developmental aspects that underpin these phenomena. The findings have significant implications for regenerative medicine and tissue engineering, offering insights into the potential for enhancing regenerative capacities in humans and other vertebrates.

Keywords: Epimorphic Regeneration, Genetics, Phylogeny, Vertebrates

ABBREVIATIONS

BMP2 - Bone Morphogenetic Protein 2
BMP4 - Bone Morphogenetic Protein 4
DNMT3A - DNA (Cytosine-5)-Methyltransferase 3A
RCc6a - Genome Reference Consortium Chicken Build 6a
GRCz11 - Genome Reference Consortium Zebrafish Build 11
GRCm38 - Genome Reference Consortium Mouse Build 38
GRCh38 - Genome Reference Consortium Human Build 38
JGI_4.2 - Joint Genome Institute 4.2 (Xenopus tropicalis genome version)
MUSCLE - Multiple Sequence Comparison by Log-Expectation
NCBI - National Center for Biotechnology Information
PCA - Principal Component Analysis
SMAD2 - Mothers against Decapentaplegic Homolog 2
m10 - Mouse Genome Version 10 (GRCm38/mm10)

INTRODUCTION

Homo sapiens belong to the monophyletic class of vertebrates which itself belongs to the monophyletic class of chordates (Ahlberg, 2001). Phylum Chordata is one among the group of deuterostomes (organisms with bilateral symmetry) besides the echinoderms and the hemichordates. Although echinoderms such as starfish possess the remarkable capacity to fully regenerate amputated tissues, this ability gradually diminishes as we go downward in geological time (Dupont and Thorndyke, 2007).

Evidence suggests that vertebrates emerged in the late Ordovician period (approx. 450 million years ago) and the jawed vertebrates boomed in the Silurian (approx. 420 million years ago) (Zhu *et al*., 2022). Of the jawed vertebrates, certain amphibian species specifically the Ambystoma mexicanum (axolotl) have retained the capacity to fully regrow their amputated limbs and apparently with original fidelity. Many lizard species among the reptiles have retained in their tails the autotomy planes in the caudal vertebrae with sufficient progenitor cell populations to regenerate whenever they lose their tails (Alibardi, 2009). Some studies have reported that the regrown tail may

have an anatomical structure relatively compromised as compared to the original. The vertebral segmentation may be replaced by a long cartilaginous rod which only serves as a functional substitute. The question arises whether tissue regrowth recapitulates embryonic development or phylogeny since mineralized bone is substituted with cartilage like the agnatha among chordate fish (Ritzman *et al*., 2012). It has long been held that regeneration might recapitulate development, yet the axolotl regrown limb lacks the characteristic apical ectodermal ridge cells (AER) of the embryo. Instead, they are believed to recycle the AER-like cells into an apical epidermal cap; a feature absent in mammals (Zhong *et al*., 2023).

The model biological organism, *Xenopus laevus* (African clawed frog) can restore lost limb tissue during early development however this process is compromised once digit differentiation initiates (Goss and Holt, 1992) indicating the gradual loss of this ability as cell fate is determined. But blastema-mediated tissue restoration is found in mammals, albeit diminished. *Mus musculus* (mice) can restore their amputated digit tips via the blastema (Fernando *et al*., 2011), however, they can only do so till the extent of the distal digit tip (P3). The proximal the amputation, the more compromised the epimorphic regeneration favoring fibrotic scar tissue healing without regrowth (Lee *et al*., 2013). This process is identical in Homo sapiens (humans). It is also why BMP2 was shortlisted as one of the candidate genes for the study, because of its critical role in the dorsoventral patterning of early organogenesis stage embryo (Rossant and Tam, 2002) and signaling in digit tip blastema (Fernando *et al*., 2011). BMP4 is expressed by the AER and the developing limb bud (Zuniga and Zeller, 2020).

However, the blastema is not a homogenous tissue. It is comprised of a heterogeneous population of cells whose fates and lineage are already determined (Carlson, 2011). Therefore the dedifferentiation process only goes back enough to maintain the information needed to generate a specified tissue type. SMAD2 gene is potent in regeneration and early embryonic development. SMAD2 regulation is critical for cell fate specification (Dick *et al.*, 2000).

Janvier (Janvier, 1981) despite the obvious absence of the mineralized skeleton, classified hagfish alongside vertebrates to form a singular, umbrella class of 'craniates'. We, however, have included only the organisms with a biomineralized skeletal system to increase the robustness of the study and limit bias in species selection. This is why Neurogenin1 was also selected as a potential gene candidate for multiple sequence alignment. Neurogenin1 is a key signaling molecule for neuroplastic changes. A neurogenic transcription factor, its overexpression increases mesenchymal stem cell-mediated therapeutic effect on the ischemic rat brain (Kim *et al*., 2020).

MATERIALS AND METHODS

Literature Review: A comprehensive literature search was conducted to identify genes involved in epimorphic regeneration within the phylum Chordata, focusing on vertebrates. The search strategy included the use of relevant keywords and phrases. The databases, including the National Center for Biotechnology Information (NCBI), were queried to extract pertinent information regarding orthologous sequences of four candidate genes: BMP2, BMP4, SMAD2, and Neurogenin1.

Species Selection: Species were meticulously selected based on the availability of highly annotated genome sequences. The chosen species were as follows: two mammals (*Mus musculus* and *Homo sapiens*), one reptile (*Anolis carolinensis*), one amphibian (*Xenopus tropicalis*), one bird (*Gallus gallus*), and one bony fish (*Danio rerio*). The genome versions of these species, utilized in this study, are as follows: GRCh38 (Genome Reference Consortium human build 38), GRCm38 (Genome Reference Consortium mouse build 38)/mm10 (mouse genome version 10), GRCc6a (Genome Reference Consortium chicken build 6a), GRCz11 (Genome Reference Consortium zebrafish build 11), JGI_4.2 (Joint Genome Institute 4.2) (*Xenopus tropicalis*).

Data Preprocessing: The orthologous sequences of the selected genes were downloaded in the FASTA format and subsequently imported into the Jalview software. Data preprocessing involved a rigorous curation process, which included the removal of gaps and empty columns, ensuring that sequences with less than 70% alignment coverage threshold were excluded.

Sequence Identity Matrix: The preprocessed sequences were subjected to multiple sequence alignment using the Clustal Omega tool and the MUSCLE program was run. A sequence identity matrix was generated to establish sequence similarity between species.

Statistical Analysis: The sequence matrix was used to perform Principal Component Analysis (PCA). PCA is a robust dimensionality reduction technique used to minimize data variance while retaining critical information (Konishi *et al*., 2019). Online resources or software tools are available and were utilized to compute eigenvalues, and only the first two principal components (PC1 and PC2) were retained, capturing more than 80% of the variance. A scatter plot was generated to visually represent the species as vectors projected from an origin, with circular gene

markers scattered across the grid. The interpretation of this plot provided insights into the relationships among the species and the candidate genes.

Phylogenetic Tree: A phylogenetic tree constructed with neighbor-joining without gap correction and a heatmap for percent identities offered a visual representation of sequence relationships. A tree was calculated with average distance for sequence identity and hierarchical clustering produced a dendrogram. Branch lengths represent evolutionary distances. The resulting phylogenetic tree was utilized to create a heatmap based on the percent identity matrix, visually representing the genetic relationships among the selected species.

By following this structured methodology, the research aims to provide a robust and transparent analysis of the genetic relationships and potential implications for epimorphic regeneration in the monophyletic chordate taxon, the vertebrates.

RESULTS

Vertebrates are a complex monophyletic class of organisms that share certain characteristic features: a lower single mandibular bone, mineralized skeleton, and metameric segmentation being a few of the most common ones. **Fig 1** depicts the phylogenetic trees with parameters set to establish neighbor joining without gap corrections for the six species Homo sapiens, *Anolis carolinensis, Xenopus tropicalis, Gallus gallus, Mus musculus* and *Danio rerio*. The heat map based on the percent identity similarities between species for each of the 4 candidate proteins BMP2, BMP4, SMAD2, and Neurogenin1 establishes the genetic likeness. **Fig 1 (b)-(d)** show *Homo sapien* and *Mus musculus* have remarkable genetic similarity for BMP4, SMAD2, and Neurogenin1, but *Homo sapien* share a genetic likeness with Danio rerio for BMP2 as depicted in **Fig 1(a)**. **Fig 1(b)** for BMP4 is revealing, for it depicts that it segregated later in evolutionary development among *Xenopus tropicalis, Gallus gallus, Homo sapien*, and *Mus musculus*, indicative of a closer percent similarity in nucleotide sequence. **Fig 1(c)** shows the genetic similarity for SMAD2 between zebrafish, *Mus musculus*, and *Homo sapien*, a vindication of prior research that gave evidence for the presence of SMAD2 signaling in zebrafish caudal fin regeneration (Liu, et al., 2018) and the mouse and human digit tip restoration via a blastema mediation. **Fig 1(d)**, however, shows zebrafish as an outlier, apparently retaining most of the nucleotide makeup of the ancestral progenitor prior to the evolutionary split between species.

Fig 2(a) represents a hierarchical clustering dendrogram depicting the nucleotide similarity relationship of all 4 candidate genes (BMP2, BMP4, SMAD2, Neurogenin1) in each of the 6 selected species (mouse, human, zebrafish, green anole, West African clawed frog, and chicken) separated on the basis of average distance since the evolutionary split. The SMAD2 genes of the *Mus musculus* and *Anolis carolinensis* appear to be closer in similarity with the *Anolis carolinensis* SMAD2 emerging only after a second lineage split. Zebrafish SMAD2 still appears as the outlier having no other immediate sister group. **Fig 2(b)** depicts a principal component plot obtained after calculating using the nucleotide similarity scores for the 4 candidate genes among each of the 6 species. The scatter of the circular gene markers and their distance against the vectors indicate an association. The zebrafish vector is isolated in the lower left quadrant, and its gene markers are clustered far from the vector. This suggests that the genes for zebrafish have a distinct pattern of sequence similarity compared to the other species. They are significantly different.

Mouse and humans share genetic similarities for BMP2, BMP4, and Neurogenin1 with a difference in SMAD2, this is reflecting of their sequence similarity governed by their shared mammalian heritage. Xenopus and Anole vectors are both in the top left quadrant, indicating some degree of similarity between these two species but the dispersion and scatter of the genetic markers also indicate a difference of genetic pattern from other species.

Also depicted in **Fig 2(a). Fig 3** is a bar graph with computed eigenvalues. Generally, the number of principal components is equal to the number of data points, but **Fig 3** displays that the first two principal components retain more than 80% of the percent cumulative variance and therefore we can select only PC1 and PC2 as key for our orthogonal principal component analysis biplot as shown in **Fig 2(b)**. Eigenvalues are used to determine the important of each principal component (PC). Eigenvalues are always non-negative, and they are typically arranged in descending order, with the first eigenvalue explaining the most variance, and the second eigenvalue explaining the second most variance. The sum of all eigenvalues is equal to the total variance in the data.

Fig 4 are conserved domains for (top to bottom) BMP2, BMP4, Neurogenin1, and SMAD2 in green anole, zebrafish, chicken, human, mouse, and xenopus. Viewed in Jalview, the alignment obtained after MUSCLE is curated by maintaining a percent identity threshold at 70% because highly divergent vertebrate species have been selected. **Fig 4(a)** and **4(c)** depict greater zebrafish divergence than other species, and also displays highly conserved cytosine residues for both genes BMP2 and Neurogenin1.

Figure 1

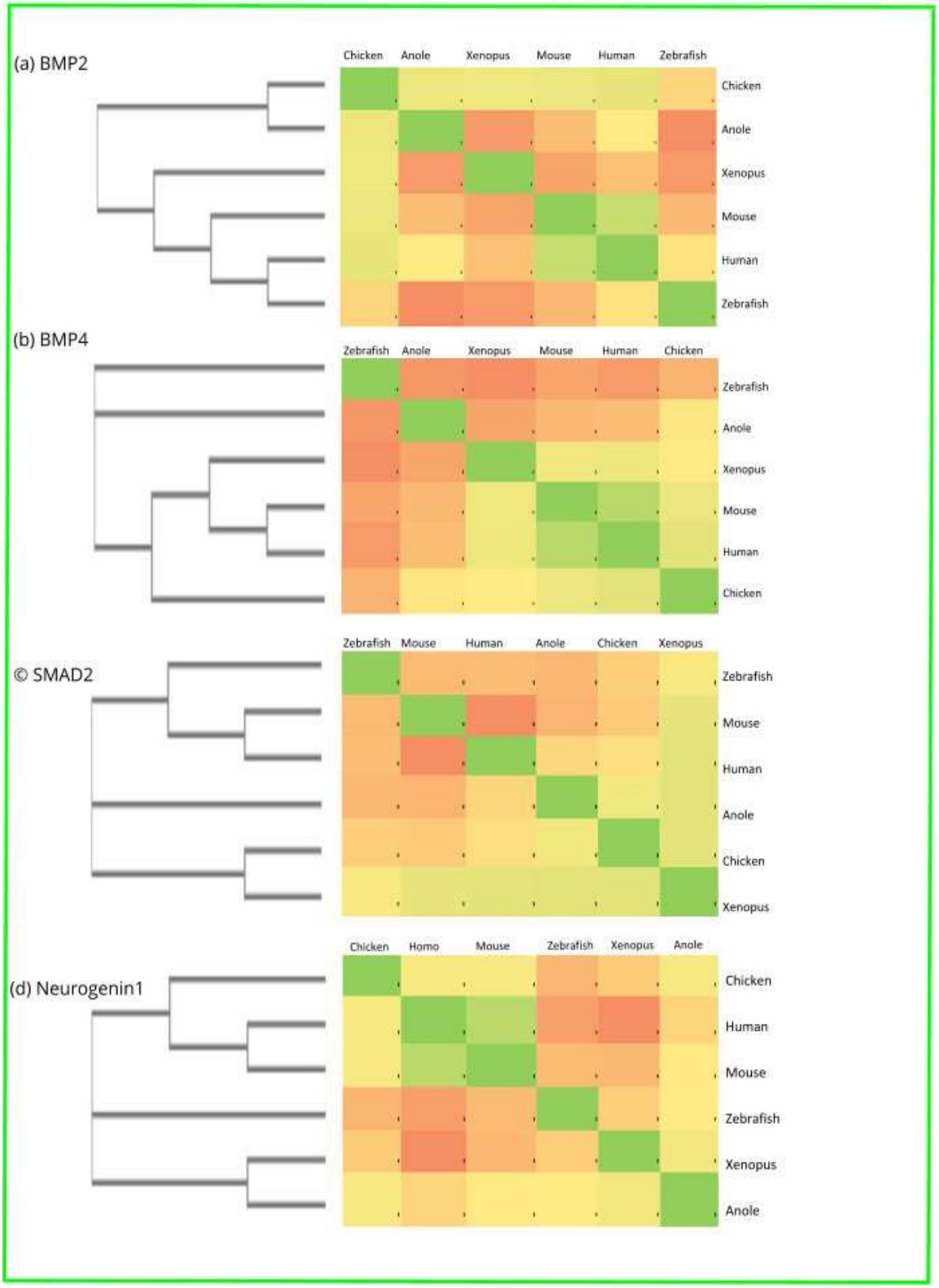

Fig. 1. A phylogenetic tree with neighbor-joining without gap corrections and heat-map depicting the percent identity matrix for sequence relationship after a multiple sequence alignment.

Figure 2

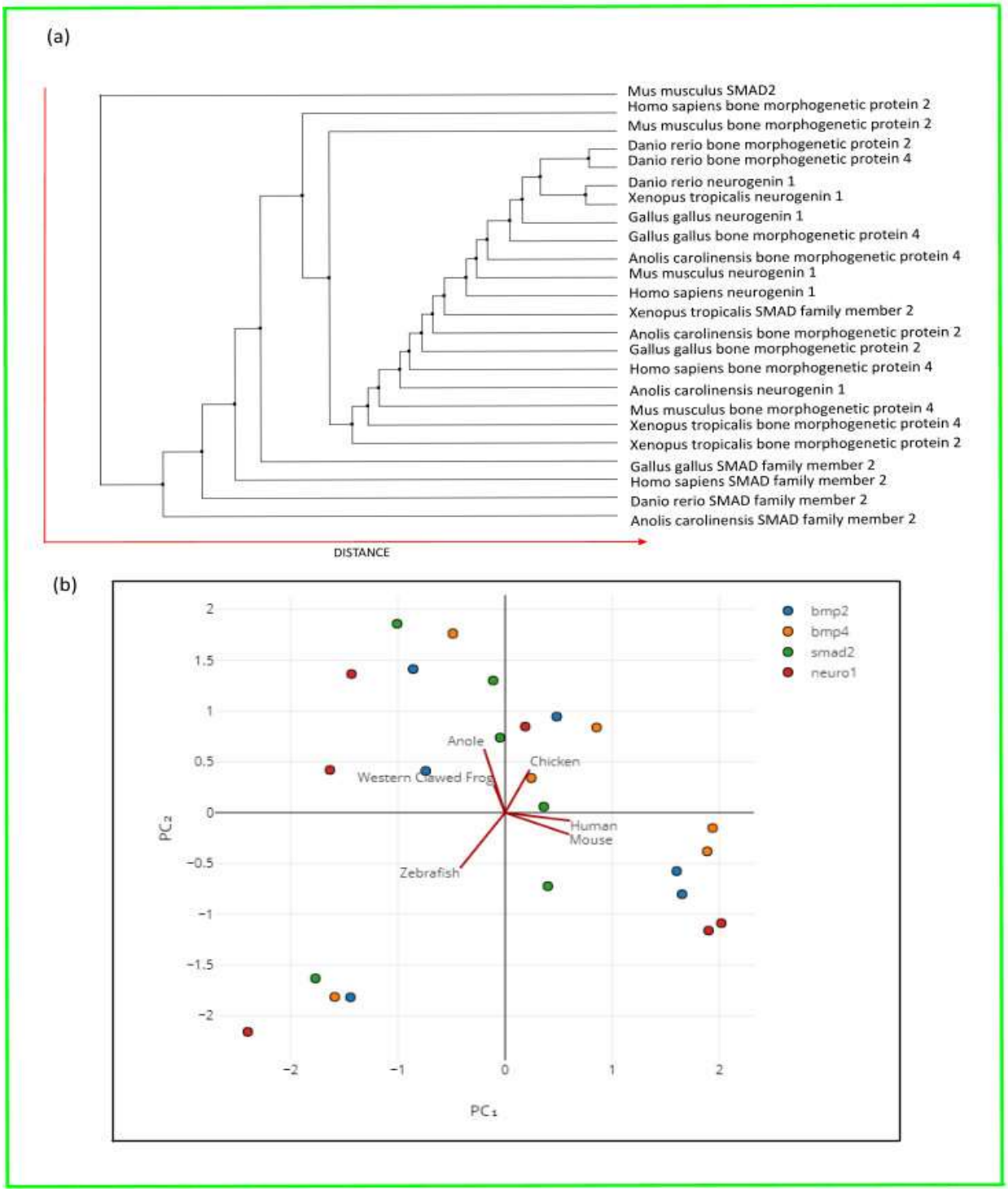

Fig. 2(a). Hierarchical clustering dendrogram for the chosen candidate genes among the selected species together displaying the average distance genetic dissimilarity/similarity between species since the evolutionary split from the ancestral progenitor in the mid to late Ordovician period. Vectors extending from the origin represent each of the six selected vertebrate species: Green anole, Xenopus, Chicken, Human, Mouse, Zebrafish with scattered (circular, colored) markers representing each of the 4 candidate genes BMP2, BMP4, SMAD2, Neurogenin1. The locations of these gene markers on the plot reflect how similar or dissimilar these genes are to the corresponding genes in other species. **Fig. 2(b)** depicts the scattering and genetic variance between species represented as vectors. The greater the distance between vector and gene markers, the greater the species variation from other species.

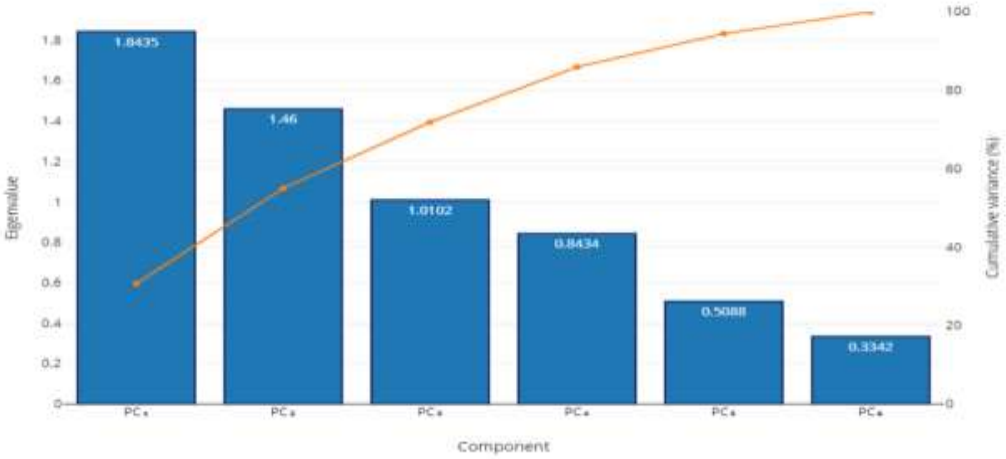

Fig. 3. The scree plot with bars depicting the principal components. The percent cumulative variance and eigenvalues as given by the length of the bars indicate the first two principal components, PC1 and PC2 display more than 85% of the total variance in the data. Normally the number of principal components is equal to the number of data points.

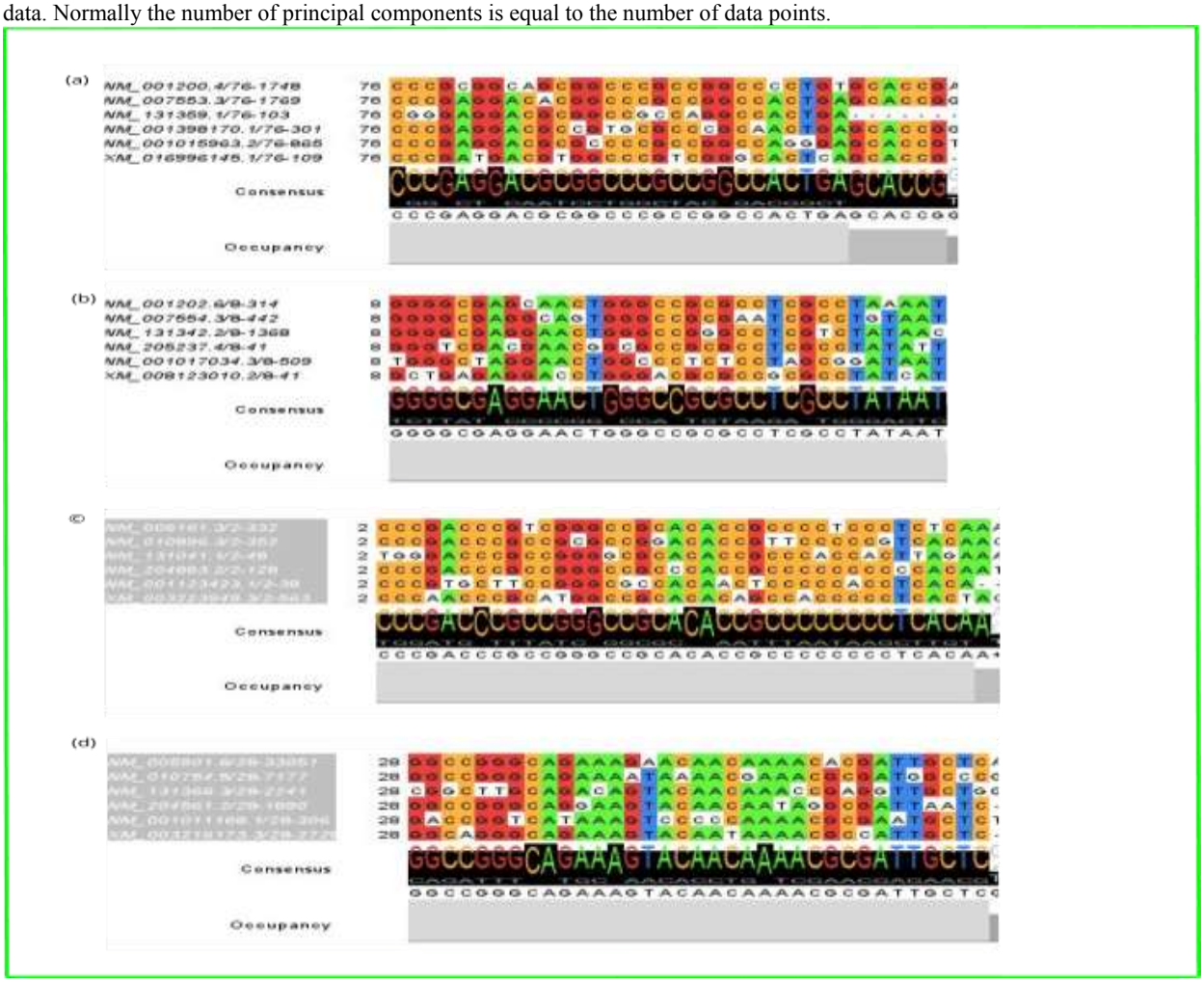

Fig. 4. Conserved sequences obtained after MUSCLE alignment, arranged at 70% percent identity threshold, and color coded for nucleotides. **Fig 4(a)** depicts BMP2 with highly conserved cytosine domains at 83%. Zebrafish differs the most among all species. **Fig 4(b)** depicts BMP4 shows highly conserved guanine domains**. Fig 4(c)** depicts Neurogenin1 highly conserved cytosine residues in orange. Most cytosine residues are conserved at 83%. **Fig 4(d)** depicts SMAD2 with conserved adenine residues. Zebrafish differs in nucleotide conserved sequences more than the other species.

DISCUSSION

The results suggest a shared relationship of certain genetic candidates critical for tissue regeneration as a whole, but epimorphic regeneration in particular. It would appear that certain genomic blocks on epimorphic regeneration have been placed on the vertebrate taxon as evolution progressed after species transitioned onto land. The first vertebrates with morphological similarity to humans with segmentation into a head, torso, and limbs are the amphibia but even they have a differential genetic expression and ability for tissue restoration among individual species. The Xenopus traits differ from those of Ambystema and ecology may have something to do with it. Axolotls retain neotenous characteristics and live their lives underwater as a juvenile species. Some researchers believe that the differential capacity between the axolotl and the lizard for tail regeneration has to do with their distinct environments.

The axolotl requires the tail for greater mobility in water whereas the reptilian transition and entire mobility requires greater use of the limbs. The choice of which organ is to be restored is therefore an evolutionary switch determined by ecology. But switches can also be mediated by epigenetic modification of the genome. Despite the common genes among all these species, the ability of tissue cells post-amputation to dedifferentiate and form a blastema varies greatly. Salamander blastema cells retain positional information and know the extent of injury and restoration required. This positional information is missing in mammals, although experiments are underway to study DNA methylation and post-transcriptional modification. Some studies lean heavily on the notion that morphogenetic processes in full swing during a developmental program during embryogenesis are not lost after birth but rather 'silenced' primarily by epigenetic switches and DNMT3a (DNA methyl transferase 3a) is a candidate for such research. Our results reveal glaring similarities between different species for similar genes yet the phenotypic expression is very distinct. It is of note that the axolotl genome is about ten times as large as homo sapiens with 32 billion base pairs, but many of those genes are epigenetically silenced. Our study reveals an amity with genetic composition among all the selected organisms. The PCA plot displays a remarkable similarity between Mus musculus and Homo sapiens and all genetic markers lie on the same side of the plot for principal components. Zebrafish is astoundingly distinct from all organisms involved yet is frequently employed as a key tool to study embryogenesis and brain function. All segmented organisms likely segregated from a fish-like ancestor early on in evolution and therefore Zebrafish suggests this relationship.

REFERENCES

Ahlberg, P. E. (2001). *Major events in early vertebrate evolution.* CRC Press.

Alibardi, L. (2009). Morphological and cellular aspects of tail and limb regeneration in lizards: a model system with implications for tissue regeneration in mammals. *Adv Anat Embryol Cell Biol*., 207: 1-109.

Carlson, B. M. (2011). *Principles of regenerative biology.* Elsevier.

Dick, A., T. Mayr, H. Bauer, A. Meier and M. Hammerschmidt (2000). Cloning and characterization of zebrafish smad2, smad3 and smad4. *Gene,* 246: 69–80.

Dupont, S. and M. Thorndyke (2007). Bridging the regeneration gap: insights from echinoderm models. *Nature Reviews Genetics,* 8: 320–320.

Fernando, W. A., Leininger, E., Simkin, J., Li, N., Malcom, C. A., Sathyamoorthi, S., . . . Muneoka, K. (2011). Wound healing and blastema formation in regenerating digit tips of adult mice. *Developmental biology,* 350: 301–310.

Goss, R. J. and R. Holt (1992). Epimorphic vs. tissue regeneration in Xenopus forelimbs. *Journal of Experimental Zoology,* 261: 451–457.

Janvier, P. (1981). The phylogeny of the Craniata, with particular reference to the significance of fossil" agnathans". *Journal of Vertebrate Paleontology*, 121–159.

Kim, G.-H., M. Subash, J. S. Yoon, D. Jo, J. Han, J. M. Hong, . . . and H. Suh-Kim (2020). Neurogenin-1 overexpression increases the therapeutic effects of mesenchymal stem cells through enhanced engraftment in an ischemic rat brain. *International Journal of Stem Cells,* 13: 127–141.

Konishi, T., S. Matsukuma, H. Fuji, D. Nakamura, N. Satou and K. Okano (2019). Principal component analysis applied directly to sequence matrix. *Scientific Reports,* 9: 19297.

Lee, J., L. Marrero, L. Yu, L. A. Dawson, K. Muneoka and M. Han (2013). SDF-1α/CXCR4 signaling mediates digit tip regeneration promoted by BMP-2. *Developmental biology,* 382: 98–109.

Lehoczky, J. A., B. Robert and C. J. Tabin (2011). Mouse digit tip regeneration is mediated by fate-restricted progenitor cells. *Proceedings of the National Academy of Sciences,* 108: 20609–20614.

Liu, F.-Y., T.-C. Hsu, P. Choong, M.-H. Lin, Y.-J. Chuang, B.-S. Chen and C. Lin (2018). Uncovering the regeneration strategies of zebrafish organs: a comprehensive systems biology study on heart, cerebellum, fin, and retina regeneration. *BMC systems biology,* 12: 33–46.

Ritzman, T. B., L. K. Stroik, E. Julik, E. D. Hutchins, E. Lasku, D. F. Denardo . . . and R. E. Fisher (2012). The gross anatomy of the original and regenerated tail in the green anole (Anolis carolinensis). *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology,* 295: 1596–1608.

Rossant, J. and P. T. Tam (2002). *Mouse development: patterning, morphogenesis, and organogenesis.* Gulf Professional Publishing.

Zhong, J., R. Aires, G. Tsissios, E. Skoufa, K. Brandt, T. Sandoval-Guzman and C. Aztekin (2023). Multi-species atlas resolves an axolotl limb development and regeneration paradox. *BioRxiv*, 2023–03.

Zhu, Y.-A., Q. Li, J. Lu, Y. Chen, J. Wang, Z. Gai, . . . and others. (2022). The oldest complete jawed vertebrates from the early Silurian of China. *Nature,* 609: 954–958.

Zuniga, A. and R. Zeller (2020). Dynamic and self-regulatory interactions among gene regulatory networks control vertebrate limb bud morphogenesis. *Current topics in developmental biology,* 139: 61–88.

(*Accepted for publication December 2023*)