Phylobone: a comprehensive database of bone extracellular matrix proteins in human and model organisms

Margalida Fontcuberta-Rigo[1], Miho Nakamura[1,2,3,*] and Pere Puigbò[4,5,6,*]

[1] Medicity Research Laboratory, Faculty of Medicine, University of Turku, Tykistökatu 6, 20520 Turku, Finland.
[2] Institute of Biomaterials and Bioengineering, Tokyo Medical and Dental University, 2-3-10 Kanda-Surugadai, Chiyoda, Tokyo 1010062 Japan
[3] Graduate School of Engineering, Tohoku University, 6-6 Aramaki Aza Aoba, Aoba-ku, Sendai, Miyagi 9808579 Japan
[4] Department of Biology, University of Turku, 20500 Turku, Finland
[5] Eurecat, Technology Center of Catalonia. Nutrition and Health Unit, Reus, 43204, Catalonia, Spain.
[6] Department of Biochemistry and Biotechnology, University Rovira i Virgili, 43007 Tarragona, Catalonia, Spain

* To whom correspondence should be addressed. Email: pepuav@utu.fi (P.P.), miho.nakamura@utu.fi (M.N.)

ABSTRACT

The bone extracellular matrix (ECM) contains minerals deposited on highly cross-linked collagen fibrils, and hundreds of non-collagenous proteins. Some of these proteins are determinant to regulate bone formation and regeneration *via* signaling pathways, and play important regulatory and structural roles. However, the complete list of bone extracellular matrix proteins, their roles, and the extent of individual and cross-species variation are not yet well understood in both humans and model organisms. Here, we introduce the most comprehensive resource of bone bone extracellular matrix (ECM) proteins that can be used in future studies in research fields such as bone regeneration, osteoporosis and mechanobiology. The Phylobone database (available at https://phylobone.com) includes 255 proteins potentially expressed in the bone extracellular matrix (ECM) of humans and 30 species of vertebrates. A bioinformatics pipeline has been utilized to identify evolutionary relationships of bone ECM proteins. This analysis facilitates the identification of potential model organisms to study the molecular mechanisms of bone regeneration. A network analysis shows the high connectivity of bone ECM proteins. A total of 214 functional protein domains have been identified, including collagen and domains involved in bone formation and resorption. Information from public drug repositories is utilized to identify potential repurposing of existing drugs. The Phylobone database provides a platform to study bone regeneration and osteoporosis in the light of (biological) evolution, and will substantially contribute to the identification of molecular mechanisms and drug targets.

INTRODUCTION

The bone extracellular matrix (ECM) consists of both organic compounds, which make up approximately 40% of the matrix, and inorganic compounds, which account for the remaining 60% [1]. The organic fraction of the ECM is mostly composed of collagen (90%) and by hundreds of non-collagenous proteins [2] that play structural and regulatory roles [3,4]. The liquid Chromatography Tandem Mass Spectrometry (LC-MS-MS) proteomics allow to elucidate regulatory mechanisms, therapeutic strategies, and biomarkers for bone regeneration and osteoporosis research [5]. Several studies of bone proteomics have reported using secreted proteins and from cells, including mesenchymal stem cells, osteoblasts and osteoclasts [6]. However, since the bone ECM contains the minerals deposited on highly crosslinked collagen fibrils, it is highly challenging to solubilize for proteomics analyses [7]. Usually, a combination of decalcification and chemical treatments is necessary for the extraction of proteins and LC-MS-MS-based proteomics [8]. Thus, the availability of bioinformatics resources is necessary to facilitate comparative studies of bone extracellular matrix proteomes both within human populations and across different model organisms. Although several non-collagenous proteins of the bone ECM proteome have been identified, the role and the list of most proteins are not well understood [1], as well as individual and cross-species variations. Therefore, the identification of non-collagenous proteins in humans and model organisms will significantly advance the field of bone regeneration and osteoporosis.

In this article, we introduce the so far most comprehensive database of putative bone ECM proteins in 39 species, including *Homo sapiens* and the most common animal models (e.g. *Danio rerio*, *Mus musculus, and Xenopus laevis*), useful for the study of osteoporosis [9]. Due to deer antlers being a speculated model of bone regeneration and osteoporosis [10–14], the database includes proteins from six species of the Cerevidae family. The Phylobone database, which includes 255 (28 collagenous and 227 non-collagenous) proteins, includes information on protein sequences, functional characterizations and potential drugs that interact with bone ECM proteins. Thus, the database provides a robust tool for the study of bone weakness and regeneration. Moreover, emerging therapies targeting novel disease mechanisms provide a powerful strategy for osteoporosis management in the future [15]. Thus, the phylobone database will be a suitable resource for the identification of novel target proteins and therapeutic peptides for the treatment and prevention of osteoporosis.

MATERIAL AND METHODS

Preliminary list of bone ECM proteins

A list of 255 seed proteins, previously identified in the literature as proteins present in the ECM, was gathered from the UniProt database [16]. These seed proteins were previously identified in the proteome of the bone ECM of *D. rerio* (n = 243), *H. sapiens* (n = 255), and *Cervus nippon* (n = 57) [17,18] (supplementary table S1). UniProt codes of *D. rerio* (Zebrafish) proteins that were obsolete (or fractions of protein sequences) were updated with current UniProt codes. Additional putative bone ECM proteins from *H. sapiens* were identified from the literature [11,18–21] and gathered from the UniProt database [16]. Each protein group was initially identified, based on the human bone ECM proteome, with a unique code ranging from PB0001 to PB0255.

Selection of species of interest

We have selected 31 representative species of vertebrates to cover a wide range of phylogenetic groups (supplementary table S2). The list includes a wide coverage of vertebrates, including most common model organisms and other vertebrates that are potentially important as model organisms of bone regeneration and osteoporosis (e.g. six members of the family Cervidae, including *Cervus hanglu yarkandensis*, *Odocoileus virginianus texanus*, *Cervus canadensis*, *Cervus elaphus*, *Muntiacus muntjak*, and *Muntiacus reevesi*). Moreover, we included in the search some species of

invertebrates (n = 8) and other taxonomic groups that are phylogenetically distant to vertebrate (e.g., *Arabidopsis*, Archaea, Bacteria, Choanoflagellata, Cnidaria, Ctenophora, Fungi, and Porifera).

Gathering ECM proteins from public databases

Orthologous proteins were gathered from public repositories of the National Center for Biotechnology Information (NCBI). This resource gets its results from NCBI's Eukaryotic Genome Annotation pipeline [22] and NCBI Gene dataset [23] in order to find orthologous sequences of some vertebrates (e.g. *M. musculus*, and *Rattus norvegicus*, *Gallus gallus*, and *Bos taurus*). Additional protein sequences of vertebrates were gathered from protein BLAST searches (blastp) [24] with a threshold of 10^{-6} on the NCBI's non-redundant protein sequences (nr). Synthetic construct sequences (taxid:32630) and partial proteins were not included in the Phylobone database.

Phylogenetic analysis and reconstruction of phyletic patterns

Phylogenetic analyses were performed with the NGPhylogeny web server [25]. We utilized the *One Click Workflow* option to obtain multiple sequence alignments (MAFFT algorithm [26]), clean alignments (BMGE algorithm [27]) and phylogenetic trees (PhyML algorithm [28]) of the 255 protein. Phyletic patterns of each protein group were built with custom made perl scripts and information from NCBI's Taxonomy [29]. We selected various species of different taxonomic groups including 5 primates, 1 rabbit, 2 rodents, 2 carnivores, 10 even-toed ungulates (including 5 members of the family Cervidae), 2 reptiles, 3 birds, 2 frogs, 4 bony fishes and 8 invertebrates (Supplementary table 3).

Functional analysis

COG functional categories and KEGG annotations of all the selected proteins in the database are retrieved from eggnog [30,31]. The dataset of 255 proteines groups was mapped onto Eggnog mapper [30,31] to identify COG functional categories involved in bone ECM proteins. COGs are classified into 26 functional groups [32]. GO enrichment data for human and zebrafish proteins were collected from Gene Ontology [33] with the Panther resource. Benjamini-Hochberg False Discovery Rate (FDR) correction was used for the searches. Pfam [34] was used to retrieve data about the domains of each human protein in the dataset. Pfam has structural and functional information for each domain as well as links to other databases such as InterPro [35]. In addition, information of functional domains of all protein sequences in the dataset was annotates with CD-Search (with default parameters; E-Value cut-off of 0.01 Composition-corrected scoring: Applied Low-complexity regions: Not filtered Maximum aligns: 500) [36].

Protein-protein interaction network

Protein-protein interactions of each human protein in the database were retrieved from EBI's Intact webserver [37] and visualized with Cytoscape [38]. Betweenness centrality was used to evaluate the centrality of each protein in the network. This parameter gives a value to a node calculating according to the following equation 1, where C_n is the betweenness centrality value of the node n, σ_{st} is the number of shortest paths from *s* to *t* and $\sigma_{st/n}$ is the number of shortest paths from *s* to *t* that *n* lies in between.

$$C_n = \sum_{s \neq t \neq n} \frac{\sigma_{st/n}}{\sigma_{st}} \qquad \textit{Equation 1}$$

Targets by existing drugs

In order to facilitate the research on repurposing drugs for osteoporosis treatment, we used information from the Drug Bank [39] and KEGG [40] databases to identify bone ECM proteins that are

targets of currently available drugs in the market.

RESULTS AND DISCUSSION

Phyletic patterns

The phyletic distribution of proteins in the selected group of species show a high conservation of the number of bone ECM proteins within vertebrate species (Figure 1 and Supplementary table 3). Seed proteins to build the Phylobone database were obtained from humans and zebrafish, which are present in 255 (100%) and 251 (98%) protein groups. Most vertebrates are present in ~90% of the protein groups. This finding underscores the significance of the presence of these proteins in the bone ECM, as they exhibit a high degree of conservation across multiple species. Such conservation may have important implications for the function and regulation of bone tissue across vertebrates. Moreover, certain proteins, such as the amelogenin (PB0250), are only present in mammals, amphibians and reptiles, because of potential evolutionary adaptations in tetrapods [41]. Notice that there are exceptional gaps in certain members of the Cervidae family (e.g. *C. elaphus*, *M. reeversi* and *M. muntjak* have less known proteins than the rest of mammals) due to incomplete annotation of the genomes. As expected, only around one-third of the proteins found in invertebrates have homologs in vertebrate species, and in many cases, the levels of similarity were found to be quite low. None of the protein sequences have homologs in lower taxonomic groups, such as bacteria.

Functional Analysis

Protein functional domains

Not unexpectedly, the most abundant protein functional domain in the Phylobone dataset is Collagen, which it's present in 12% of the human proteins, which covers 90% of the organic fraction of the bone ECM [2]. Collagen is essential for homeostasis maintenance and it serves as scaffold to many other macromolecules and hydroxyapatite, enabling cell attachment and bone resistance to mechanical forces [1,42,43]. Furthermore, 214 non-collagenous functional domains are present in human sequences of the dataset (Figure 2 and Supplementary table S4). Most common in bone formation, resorption, cell attachment or they are intermediaries in a variety of metabolic pathways. Leucine-rich repeats (LRRs) are the second most common functional domains due to large abundance of leucine-rich proteoglycans (SLRPs) in the bone ECM [1]. LRRs have an important role in bone formation and maintenance of bone homeostasis. Moreover, Laminin, von Willebrand factor (VWF), epidermal growth factor (EGF) and trypsin domains are also frequent. Laminins ,VWF and EGF have significant roles in bone regeneration and stability: 1) Laminins are involved in cell adhesion, proliferation and differentiation [44]; 2) VWF enhances the inhibition of osteoclastogenesis [45], and 3) EGF can stimulate bone resorption [46]. Trypsin is an important domain in the ECM because of its ability to cleave proteins. In addition, several matrix-degrading enzymes, such as cathepsin K and matrix metalloproteinases, also contribute to an efficient degradation of bone [47].

Analysis of gene ontology

We have compared GO enrichment analyses of biological process, molecular function and cellular component categories between human and zebrafish (Figure 3 and supplementary figure 1-6)). GO annotations are more abundant in human, but there is certain functional overlap with zebrafish that can be utilized to evaluate the use of zebrafish as model organism of osteoporosis [48]. A common category in biological processes is bone mineralization, which is essential for bone regeneration [49]. In molecular functions, several binding activities are both abundant in humans and zebrafish. As expected, collagen is frequent in GO cellular component categories of both species, as it is the main component in bone ECM [1].

Clusters of orthologous groups

The Phylobone database is a compilation of ECM proteins, thus several proteins are categorized in the COG database under the function of as *extracellular structures* (W) and *signal transduction* (T) (Figure 4a). Signal transduction is an indispensable function of the ECM for the control homeostasis and transmission of molecular signals into the cells [50]. Furthermore, proteins involved in bone signaling regulate bone formation and resorption [49] (Figure 5). The third and fourth most common COG functional categories are *function unknown* (S) (our database may shed some light to the identification of these proteins' function by studying them in suitable animal models) and as well as *post-translational modification, protein turnover and chaperones* (O). Proteins of the ECM are expected to be highly integrated to enable signaling functions and to control the homeostasis.

Protein interaction network of bone ECM proteins in human

We have built a protein interaction network of 5781 human proteins, including putative bone ECM proteins and those that potentially interact (independently of body tissue) with them (Figure 4b). The parameter of betweenness centrality shows higher values in bone ECM proteins compared to those that interact with them (Figure 4c). This result is in agreement with the central role of bone ECM proteins to maintain tissue homeostasis and regulatory processes. Interactions with cellular components also allow signaling (e.g. cell-matrix interactions permits osteoblast differentiation [4]). The Amyloid beta precursor protein, a transmembrane protein that can be cleaved by a secretase and released in the ECM [51], has the highest value of betweenness centrality (Figure 4d). This protein regulates osteoclast function (affecting bone remodeling) and is present in both bone and brain and can link osteoporosis and neurodegenerative diseases [52,53]. Fibronectin has also one of the largest values of betweenness centrality. This glycoprotein of the ECM extracellular matrix has an important role in osteogenesis because it promotes pre-osteoblast mineralization and differentiation [54]. Furthermore, fibronectin plays as a scaffold for the cleavage of procollagen (interacting with both bone morphogenic protein 1 and procollagen) [55]. This ability to bind other ECM proteins gives fibronectin the high centrality value and an important role in bone formation and regeneration with a single mechanism.

How to use the Phylobone database to study bone regeneration and osteoporosis

Osteoporosis is one of the most common bone problems in the middle age and elderly population worldwide [56,57]. Approximately, nine million fractures per year (one every three seconds) are caused by osteoporosis that contribute significantly to the morbidity and mortality in developed countries [15]. Given that life expectancy is increasing globally, osteoporosis has become an emerging topic, as it significantly affects the quality of life of individuals in most countries [15]. Osteoporosis is a metabolic disease caused by an imbalance between bone anabolism and catabolism [48]. In general, the recommendations to reduce osteoporosis risk are to get adequate calcium, sunlight, vitamin D and weight bearing exercises. It is known that the reduction in mechanical loading due to prolonged bed rest or long-term exposure to microgravity can lead to reduction in bone mass [58,59]. The mechanisms of weight bearing exercises to prevent osteoporosis are not fully understood, yet collagen fibrils, which represents 90% of the organic components of the bone ECM, generate electricity during weight bearing exercise that is stored in the inorganic components of the ECM [60]. Since bone ECM has both structural and regulatory roles, non-collagenous organic components have a key role in bone regulation by mechanical loading [3,4].

Several non-collagenous proteins of the bone ECM proteome have been identified previously, including osteocalcin, osteonectin and R-spondins, however the role and list of bone ECM proteins in humans (and model organisms) is not well understood [1]. Some of the ECM proteins have been reported to play an important role in promoting bone formation. Type I collagen, as the most abundant protein in the bone ECM, provides mechanical support such as bone strength and fragility as a structural role [61] as well as integrin receptors for osteoblasts as a regulatory role [62]. Moreover, the

bone ECM proteome comprises hundreds of non-collagenous proteins with structural, regulatory, and homeostatic roles [63] (Figure 5). Osteopontin is a non-collagenous protein in bone ECM related to mechanical stress with multifunctional functions [64]. This protein is produced by osteoblasts as well as osteoclasts [65] and is considered to inhibit osteoblast activity and promote osteoclast activity [66]. In addition, there are proteins responsible for maintaining bone homeostasis, such as bone morphogenetic proteins (BMPs) [67].

Phylobone database to study mechanobiology

Outcomes from the Phylobone project will also have an impact on the field of mechanobiology. Several publications have determined the role of some cell types, cellular membrane receptors, and bone ECM proteins in mechanical loading [64,68,69]. For instance, osteocytes are known as mechanosensing cells in bone tissue by transduction of mechanical signals to biochemical responses [68]. PIEZO1 is a mechanosensitive ion channel component that became a topic of discussion because its discoverer won the Nobel Prize in 2021 [70]. PIEZO1 as a mechanosensing protein regulates homeostasis *via* crosstalk of osteoblasts (bone-making cells) and osteoclasts (bone-eating cells) [69]. In addition, the mechanical properties, especially fluid shear stress in bone ECM has significant effects on cells and their interactions [71]. However, only few non-collagenous proteins of the ECM (e.g. osteopontin) have been related to mechanical stress [64].

Potential repurposing of existing drugs

The Phylobone database is a key resource to advance the research on the prevention of osteoporosis and the development of new treatments for bone regeneration. The most common treatments for osteoporosis are bisphosphonates, monoclonal receptor activator of nuclear factor-kB ligand (RANKL) antibodies, monoclonal sclerostin antibodies, and a parathyroid hormone peptide [72]. The parathyroid hormone peptide increases osteoblast activity and inhibits osteoclast recruitment, whereas the targets of the other treatments inhibit osteoclast resorption [73]. However, these treatments have limitations and problems with side effects and serious risks. For instance, bisphosphonates have been reported to have higher fracture rates after long-term use, characterized as more than six years and require “bisphosphonate holiday” [74]. RANKL antibodies have been reported to have some side effects, such as skin eczema, flatulence, cellulitis, and osteonecrosis of the jaw [75]. Thus, there is still a need for osteoporosis treatment. However finding new drugs is expensive thus repurposing existing drugs is a main goal for the identification of novel treatments and preventive methods. Because several proteins of the bone ECM proteome have been identified as playing a regulatory role in osteogenesis and bone degradation [1], we hypothesize that key proteins of the ECM involved in mechanical loading could be potential drug targets for the treatment and prevention of osteoporosis. The Phylobone database includes 36 proteins out of 255 proteins from the bone ECM that have existing drugs and are putative candidates for repurposing of existing drugs (Table 1 and Supplementary table S5), including 10 drugs that interact with more than one protein (Table 2).

AVAILABILITY

The Phylobone database is freely accessible at https://phylobone.com. The current version of the dataset includes 8,615 putative bone ECM proteins from 39 species of vertebrates and invertebrates, and categorized in 255 protein groups. Each protein of the database is annotated with basic information, that includes its name, organism, a general protein description, a list of gene ontologies (GO) associated, protein-protein interactions (PPI), functional domains, metabolic pathways and drugs. We have precomputed a phyletic profile of proteins and species, based on the identification of orthologous sequences in vertebrates, including common model organisms. Protein sequences, multiple sequence alignments and PhyML phylogenetic trees are available to visualize on the web browser and download. Seed sequences of human and zebrafish, that have been used to build the

database, are also available. Moreover, the database includes several links to external resources, including Uniprot [16], Protein Atlas [76], Intact [37], InterPro [35] and the BLAST alignment tool [24]. All proteins have been mapped onto currently existing drugs and the database includes links to DrugBank [39] and KEGG [40] databases.

CONCLUSIONS AND FUTURE DEVELOPMENTS

The aim of the Phylobone research project is to provide a platform to study bone regeneration and osteoporosis in the light of (biological) evolution. The current database includes the most comprehensive repository of bone ECM proteins in human and animal models. The phylobone database provides a bioinformatics resource that includes sequences, phylogenetic and functional. Thus, this resource will be helpful to address a major technical challenge in bone biology, the identification of non-collagenous proteins involved in bone formation, regeneration and progression of osteoporosis disease. In the future, we anticipate an update of the Phylobone database with more experimental data on the regulatory role of bone ECM proteins in bone regeneration and osteoporosis.

ACKNOWLEDGEMENT

We thank members of the Phylobone team and collaborators for helpful discussions.

FUNDING

This project was supported by continuation funds from the Turku Collegium for Science, Medicine and Technology, the Japan Society for the Promotion of Science (#) and the Sigrid Jusélius Foundation (#230131). M.F-R. internship at the university of Turku was funded by the Erasmus+ program.

CONFLICT OF INTEREST

None.

FIGURE LEGENDS

Figure 1. General parameters of the selected species in the Phylobone database.
Tree visualization was made with iTOL v5 [77] and shows the following datasets: G+C percentage (yellow), genome size (blue), number of proteins (green), number of putative bone extracellular matrix proteins (purple) and taxonomic information.

Figure 2. Distribution of protein functional domains in the Phylobone database.
Relation between the human proteins of the dataset (purple) and their domains (blue). Dots filled in orange are the most common domains in the database (notice that only domains present in at least 5 proteins are shown). The number of proteins that each domain has is annotated in green (n).

Figure 3. Venn diagram of the number of GO categories in the bone extracellular matrix proteins of *Homo sapiens* and *Danio rerio*.
a. Biological process categories.
b. Molecular function categories.
c. Cellular component categories.

Figure 4. Protein-protein interaction network of human proteins in the bone extracellular matrix
a. Percentage of proteins that have each COG functional group. The sequences of all selected species are represented. Signal transduction mechanisms (T), Extracellular structures (W), Function unknown (S), Post-translational modification, protein turnover, chaperones (O), Defense mechanisms (V), Cytoskeleton (Z), Intracellular trafficking, secretion, and vesicular transport (U), Inorganic ion transport and metabolism (P), Carbohydrate transport and metabolism (G), Cell wall/membrane/envelope biogenesis (M), Secondary metabolites biosynthesis, transport and catabolism (Q), Amino acid transport and metabolism (E), Transcription (K), Lipid transport and metabolism (I), and Energy proSignal transduction mechanisms (T).
b. Tree of interactions between proteins. Orange dots are human proteins in the dataset and blue ones are for other proteins that they interact with.
c. Gaussian representation of proteins' betweenness centrality in a logarithmic scale (ln). Green lines represent all 5871 proteins that interact with each other, and orange lines are for the 255 human proteins in the dataset.
d. Human proteins that have a higher betweenness centrality value in our database.

Figure 5. Schematic representation of the state-of-the-art in bone formation and remodeling mediated by proteins of the extracellular matrix (ECM).
a. Simplified representation of processes involving osteocytes, osteoblasts and osteoclasts formation in the bone. ECM proteins can modulate them via activation or inhibition through signaling pathways. Those mechanisms allow to control bone homeostasis.
b. Percentage of proteins with putative structural or regulatory role in the bone ECM based on annotations in the Uniprot database [16].

TABLES

Table 1. List of the 36 bone extracellular matrix (ECM) proteins that are target of at least one available drug in the market

PB Code	Uniprot Code	Uniprot Name	N. of Drugs	Protein Description
PB0124	P00742	FA10_HUMAN	17	Coagulation factor X
PB0003	P00734	THRB_HUMAN	13	Prothrombin
PB0080	P01031	CO5_HUMAN	12	Complement C5
PB0242	P02766	TTHY_HUMAN	10	Transthyretin
PB0107	P14780	MMP9_HUMAN	8	Matrix metalloproteinase-9
PB0178	O14793	GDF8_HUMAN	7	Growth/differentiation factor 8
PB0204	P08253	MMP2_HUMAN	6	72 kDa type IV collagenase
PB0235	P06756	ITAV_HUMAN	5	Integrin alpha-V
PB0238	P00747	PLMN_HUMAN	4	Plasminogen
PB0067	P00740	FA9_HUMAN	3	Coagulation factor IX
PB0008	P02743	SAMP_HUMAN	3	Serum amyloid P-component
PB0132	P05121	PAI1_HUMAN	3	Plasminogen activator inhibitor 1
PB0236	P05556	ITB1_HUMAN	3	Integrin beta-1
PB0112	P45452	MMP13_HUMAN	3	Collagenase 3
PB0160	P61812	TGFB2_HUMAN	3	Transforming growth factor beta-2 proprotein
PB0159	P00751	CFAB_HUMAN	2	Complement factor B
PB0161	P21333	FLNA_HUMAN	2	Filamin-A
PB0100	P43235	CATK_HUMAN	2	Cathepsin K
PB0017	P69905	HBA_HUMAN	2	Hemoglobin subunit alpha
PB0175	P01009	A1AT_HUMAN	1	Alpha-1-antitrypsin
PB0030	P01024	CO3_HUMAN	1	Complement C3
PB0187	P01042	KNG1_HUMAN	1	Kininogen-1
PB0212	P02753	RET4_HUMAN	1	Retinol-binding protein 4
PB0077	P04114	APOB_HUMAN	1	Apolipoprotein B-100
PB0092	P05186	PPBT_HUMAN	1	Alkaline phosphatase, tissue-nonspecific isozyme
PB0174	P07477	TRY1_HUMAN	1	Serine protease 1
PB0234	P08648	ITA5_HUMAN	1	Integrin alpha-5
PB0201	P10600	TGFB3_HUMAN	1	Transforming growth factor beta-3 proprotein
PB0200	P10909	CLUS_HUMAN	1	Clusterin
PB0196	P29279	CCN2_HUMAN	1	CCN family member 2
PB0246	P63000	RAC1_HUMAN	1	Ras-related C3 botulinum toxin substrate 1
PB0036	Q92743	HTRA1_HUMAN	1	Serine protease HTRA1
PB0144	Q9BX93	PG12B_HUMAN	1	Group XIIB secretory phospholipase A2-like protein
PB0254	Q9BXY4	RSPO3_HUMAN	1	R-spondin-3
PB0181	Q9GZV9	FGF23_HUMAN	1	Fibroblast growth factor 23
PB0203	Q9UNI1	CELA1_HUMAN	1	Chymotrypsin-like elastase family member 1

Table 2. Drugs that interact with more than one protein of the bone extracellular matrix (ECM). There are 114 drugs that interact with at least one bone ECM protein (104 drugs interact with one protein, 9 drugs interact with two proteins and 1 drug interacts with three proteins). The complete matrix of drugs and protein interactions is available in supplementary table 5.

Drug	Related proteins	Drug efficacy
Batimastat	P08253, P14780	Antineoplastic, Matrix metalloproteinase inhibitor
Cipemastat, Trocade	P14780, P45452	Antirheumatic, Cartilage protectant, Matrix metalloproteinase inhibitor
Ilomastat	P08253, P14780	Antineoplastic, Matrix metalloproteinase inhibitor
Marimastat	P08253, P14780	Antineoplastic, Matrix metalloproteinase inhibitor
Rebimastat	P08253, P14780	Antineoplastic, Matrix metalloproteinase inhibitor
Tanomastat	P08253, P14780	Antineoplastic, Matrix metalloproteinase inhibitor
Volociximab	P05556, P08648	Aging-related macular degeneration therapeutic agent, Antineoplastic, Angiogenesis inhibitor, Anti-alpha5beta1 integrin antibody
Fresolimumab	P10600, P61812	Immunomodulator, Anti-transforming growth factor beta antibody
Emicizumab, Hemlibra	P00740, P00742	Bleeding suppressant, Factor VIII replacement
Prinomastat	P08253, P14780, P45452	Antineoplastic, Angiogenesis inhibitor, Matrix metalloproteinase inhibitor

** **P08253**: 72 kDa type IV collagenase; **P14780**: Matrix metalloproteinase-9; **P45452**: Collagenase 3; **P05556**: Integrin beta-1; **P08648** Integrin alpha-5; **P10600**: Transforming growth factor beta-3 proprotein; **P61812**: Transforming growth factor beta-2 proprotein; **P00740**: Coagulation factor IX; **P00742**: Coagulation factor X*

REFERENCES

1. Lin, X., Patil, S., Gao, Y.-G. & Qian, A. The bone extracellular matrix in bone formation and regeneration. *Front. Pharmacol.* **11**, 757 (2020).
2. Mansour, A., Mezour, M. A., Badran, Z. & Tamimi, F. * extracellular matrices for bone regeneration: A literature review. *Tissue Eng. Part A* **23**, 1436–1451 (2017).
3. Zhao, D. *et al.* Osteocytes regulate bone anabolic response to mechanical loading in male mice via activation of integrin α5. *Bone Res.* **10**, 49 (2022).
4. Alford, A. I., Kozloff, K. M. & Hankenson, K. D. Extracellular matrix networks in bone remodeling. *Int. J. Biochem. Cell Biol.* **65**, 20–31 (2015).
5. Calciolari, E. & Donos, N. Proteomic and transcriptomic approaches for studying bone regeneration in health and systemically compromised conditions. *Proteomics Clin. Appl.* **14**, e1900084 (2020).
6. Lee, J.-H. & Cho, J.-Y. Proteomics approaches for the studies of bone metabolism. *BMB Rep.* **47**, 141–148 (2014).
7. Mueller, C. *et al.* Unlocking bone for proteomic analysis and FISH. *Lab. Invest.* **99**, 708–721 (2019).
8. Mickleburgh, H. L. *et al.* Human Bone Proteomes before and after Decomposition: Investigating the Effects of Biological Variation and Taphonomic Alteration on Bone Protein Profiles and the Implications for Forensic Proteomics. *J. Proteome Res.* **20**, 2533–2546 (2021).
9. Reinwald, S. & Burr, D. Review of nonprimate, large animal models for osteoporosis research. *J. Bone Miner. Res.* **23**, 1353–1368 (2008).
10. Borsy, A. *et al.* Identifying novel genes involved in both deer physiological and human pathological osteoporosis. *Mol. Genet. Genomics* **281**, 301–313 (2009).
11. Yao, B. *et al.* Comparative transcriptome analysis of the main beam and brow tine of sika deer antler provides insights into the molecular control of rapid antler growth. *Cell. Mol. Biol. Lett.* **25**, 42 (2020).
12. Banks, W. J., Epling, G. P., Kainer, R. A. & Davis, R. W. Antler growth and osteoporosis. I. Morphological and morphometric changes in the costal compacta during the antler growth cycle. *Anat. Rec.* **162**, 387–398 (1968).
13. Wang, D. & Landete-Castillejos, T. Stem cells drive antler regeneration. *Science* **379**, 757–758 (2023).
14. Qin, T. *et al.* A population of stem cells with strong regenerative potential discovered in deer antlers. *Science* **379**, 840–847 (2023).
15. Noh, J.-Y., Yang, Y. & Jung, H. Molecular mechanisms and emerging therapeutics for osteoporosis. *Int. J. Mol. Sci.* **21**, (2020).
16. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
17. Kessels, M. Y. *et al.* Proteomics analysis of the zebrafish skeletal extracellular matrix. *PLoS ONE* **9**, e90568 (2014).
18. Zhang, R., Li, Y. & Xing, X. Comparative antler proteome of sika deer from different developmental stages. *Sci. Rep.* **11**, 10484 (2021).
19. Shi, G.-X. *et al.* Evidence of the Role of R-Spondin 1 and Its Receptor Lgr4 in the Transmission of Mechanical Stimuli to Biological Signals for Bone Formation. *Int. J. Mol. Sci.* **18**, (2017).
20. Frasheri, I. *et al.* Full-length amelogenin influences the differentiation of human dental pulp stem cells. *Stem Cell Res. Ther.* **7**, 10 (2016).
21. Yu, M. *et al.* BMP4 mutations in tooth agenesis and low bone mass. *Arch. Oral Biol.* **103**, 40–46 (2019).
22. Thibaud-Nissen, F., Souvorov, A., Murphy, T., DiCuccio, M. & Kitts, P. Eukaryotic Genome Annotation Pipeline - The NCBI Handbook - NCBI Bookshelf. (2013).
23. Home - Gene - NCBI. https://www.ncbi.nlm.nih.gov/gene.
24. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
25. Lemoine, F. *et al.* NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Res.* **47**, W260–W265 (2019).
26. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
27. Criscuolo, A. & Gribaldo, S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **10**, 210 (2010).
28. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
29. Schoch, C. L. *et al.* NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database (Oxford)* **2020**, (2020).
30. Huerta-Cepas, J. *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).
31. Tatusov, R. L., Koonin, E. V. & Lipman, D. J. A genomic perspective on protein families. *Science* **278**, 631–637 (1997).
32. Galperin, M. Y., Kristensen, D. M., Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Microbial genome analysis: the COG approach. *Brief. Bioinformatics* **20**, 1063–1070 (2019).
33. Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47**, D419–D426 (2019).

34. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021).
35. Blum, M. *et al.* The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* **49**, D344–D354 (2021).
36. Marchler-Bauer, A. & Bryant, S. H. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.* **32**, W327-31 (2004).
37. Orchard, S. *et al.* The MIntAct project - IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* **42**, D358-63 (2014).
38. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
39. Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).
40. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).
41. Diekwisch, T. G. H. *et al.* Amelogenin evolution and tetrapod enamel structure. *Front. Oral Biol.* **13**, 74–79 (2009).
42. Daneault, A., Prawitt, J., Fabien Soulé, V., Coxam, V. & Wittrant, Y. Biological effect of hydrolyzed collagen on bone metabolism. *Crit. Rev. Food Sci. Nutr.* **57**, 1922–1937 (2017).
43. Stock, S. R. The Mineral-Collagen Interface in Bone. *Calcif. Tissue Int.* **97**, 262–280 (2015).
44. Gu, Y.-C. *et al.* Laminin isoform-specific promotion of adhesion and migration of human bone marrow progenitor cells. *Blood* **101**, 877–885 (2003).
45. Mojzisch, A. & Brehm, M. A. The manifold cellular functions of von willebrand factor. *Cells* **10**, (2021).
46. Siddiqui, J. A. & Partridge, N. C. Physiological bone remodeling: systemic regulation and growth factor involvement. *Physiology (Bethesda)* **31**, 233–245 (2016).
47. Edwards, J. R. & Mundy, G. R. Advances in osteoclast biology: old findings and new insights from mouse models. *Nat. Rev. Rheumatol.* **7**, 235–243 (2011).
48. Bergen, D. J. M., Kague, E. & Hammond, C. L. Zebrafish as an Emerging Model for Osteoporosis: A Primary Testing Platform for Screening New Osteo-Active Compounds. *Front Endocrinol (Lausanne)* **10**, 6 (2019).
49. Kenkre, J. S. & Bassett, J. The bone remodelling cycle. *Ann. Clin. Biochem.* **55**, 308–327 (2018).
50. Hastings, J. F., Skhinas, J. N., Fey, D., Croucher, D. R. & Cox, T. R. The extracellular matrix as a key regulator of intracellular signalling networks. *Br. J. Pharmacol.* **176**, 82–92 (2019).
51. Müller, U. C., Pietrzik, C. U. & Deller, T. The physiological functions of the β-amyloid precursor protein APP. *Exp. Brain Res.* **217**, 325–329 (2012).
52. Roos, P. M. Osteoporosis in neurodegeneration. *J. Trace Elem. Med. Biol.* **28**, 418–421 (2014).
53. Cui, S. *et al.* APPswe/Aβ regulation of osteoclast activation and RAGE expression in an age-dependent manner. *J. Bone Miner. Res.* **26**, 1084–1098 (2011).
54. Yang, C. *et al.* Fibronectin 1 activates WNT/β-catenin signaling to induce osteogenic differentiation via integrin β1 interaction. *Lab. Invest.* **100**, 1494–1502 (2020).
55. Saunders, J. T. & Schwarzbauer, J. E. Fibronectin matrix as a scaffold for procollagen proteinase binding and collagen processing. *Mol. Biol. Cell* **30**, 2218–2226 (2019).
56. Compston, J. E., McClung, M. R. & Leslie, W. D. Osteoporosis. *Lancet* **393**, 364–376 (2019).
57. Srivastava, M. & Deal, C. Osteoporosis in elderly: prevention and treatment. *Clin. Geriatr. Med.* **18**, 529–555 (2002).
58. Sun, Z. *et al.* Simulated microgravity inhibits L-type calcium channel currents partially by the up-regulation of miR-103 in MC3T3-E1 osteoblasts. *Sci. Rep.* **5**, 8077 (2015).
59. Leblanc, A. D., Schneider, V. S., Evans, H. J., Engelbretson, D. A. & Krebs, J. M. Bone mineral loss and recovery after 17 weeks of bed rest. *J. Bone Miner. Res.* **5**, 843–850 (1990).
60. Nakamura, M., Hiratai, R. & Yamashita, K. Bone mineral as an electrical energy reservoir. *J. Biomed. Mater. Res. A* **100**, 1368–1374 (2012).
61. Buckwalter, J. A., Glimcher, M. J., Cooper, R. R. & Recker, R. Bone biology. I: Structure, blood supply, cells, matrix, and mineralization. *Instr. Course Lect.* **45**, 371–386 (1996).
62. Jikko, A., Harris, S. E., Chen, D., Mendrick, D. L. & Damsky, C. H. Collagen integrin receptors regulate early osteoblast differentiation induced by BMP-2. *J. Bone Miner. Res.* **14**, 1075–1083 (1999).
63. Alcorta-Sevillano, N., Macías, I., Infante, A. & Rodríguez, C. I. Deciphering the relevance of bone ECM signaling. *Cells* **9**, (2020).
64. Fujihara, S. *et al.* Function and regulation of osteopontin in response to mechanical stress. *J. Bone Miner. Res.* **21**, 956–964 (2006).
65. Merry, K., Dodds, R., Littlewood, A. & Gowen, M. Expression of osteopontin mRNA by osteoclasts and osteoblasts in modelling adult human bone. *J. Cell Sci.* **104 (Pt 4)**, 1013–1020 (1993).
66. Singh, A., Gill, G., Kaur, H., Amhmed, M. & Jakhu, H. Role of osteopontin in bone remodeling and orthodontic tooth movement: a review. *Prog. Orthod.* **19**, 18 (2018).
67. Sánchez-Duffhues, G., Hiepen, C., Knaus, P. & Ten Dijke, P. Bone morphogenetic protein signaling in bone homeostasis. *Bone* **80**, 43–59 (2015).
68. Yan, Y., Wang, L., Ge, L. & Pathak, J. L. Osteocyte-Mediated Translation of Mechanical Stimuli to Cellular Signaling and Its Role in Bone and Non-bone-Related Clinical Complications. *Curr. Osteoporos. Rep.* **18**, 67–80 (2020).
69. Wang, L. *et al.* Mechanical sensing protein PIEZO1 regulates bone homeostasis via osteoblast-osteoclast crosstalk. *Nat. Commun.* **11**, 282 (2020).

70. The Nobel Prize in Physiology or Medicine 2021 - Advanced information - NobelPrize.org. https://www.nobelprize.org/prizes/medicine/2021/advanced-information/.
71. Qin, L., Liu, W., Cao, H. & Xiao, G. Molecular mechanosensors in osteocytes. *Bone Res.* **8**, 23 (2020).
72. LeBoff, M. S. *et al.* The clinician's guide to prevention and treatment of osteoporosis. *Osteoporos. Int.* **33**, 2049–2102 (2022).
73. Gao, Y., Patil, S. & Jia, J. The development of molecular biology of osteoporosis. *Int. J. Mol. Sci.* **22**, (2021).
74. Wang, M., Wu, Y.-F. & Girgis, C. M. Bisphosphonate Drug Holidays: Evidence From Clinical Trials and Real-World Studies. *JBMR PLUS* **6**, e10629 (2022).
75. Zhang, N. *et al.* Pros and cons of denosumab treatment for osteoporosis and implication for RANKL aptamer therapy. *Front. Cell Dev. Biol.* **8**, 325 (2020).
76. Uhlén, M. *et al.* The human secretome. *Sci. Signal.* **12**, (2019).
77. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).

Figure 1

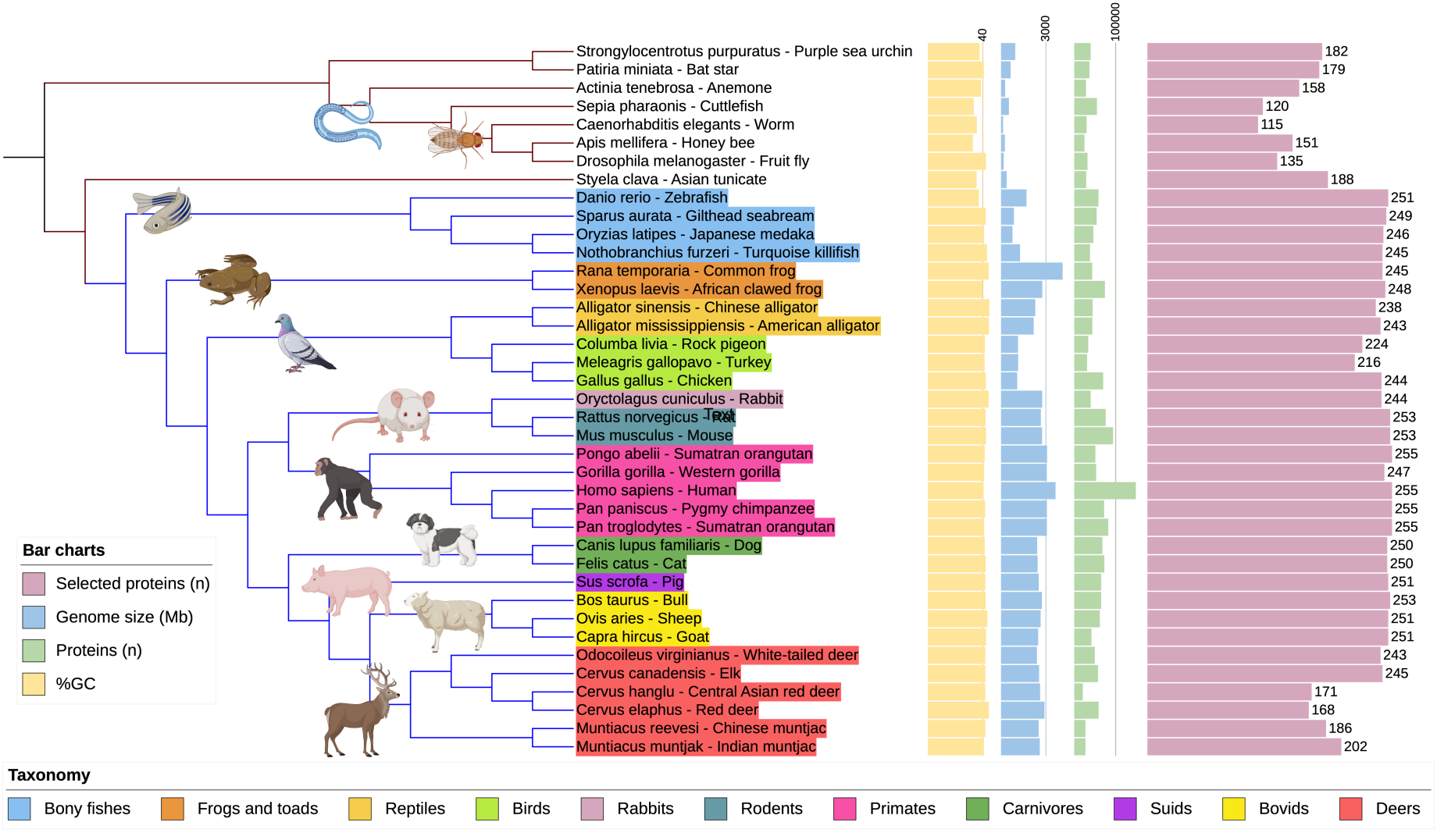

40
3000
100000
Strongylocentrotus purpuratus - Purple sea urchin 182
Patiria miniata - Bat star 179
Actinia tenebrosa - Anemone 158
Sepia pharaonis - Cuttlefish 120
Caenorhabditis elegants - Worm 115
Apis mellifera - Honey bee 151
Drosophila melanogaster - Fruit fly 135
Styela clava - Asian tunicate 188
Danio rerio - Zebrafish 251
Sparus aurata - Gilthead seabream 249
Oryzias latipes - Japanese medaka 246
Nothobranchius furzeri - Turquoise killifish 245
Rana temporaria - Common frog 245
Xenopus laevis - African clawed frog 248
Alligator sinensis - Chinese alligator 238
Alligator mississippiensis - American alligator 243
Columba livia - Rock pigeon 224
Meleagris gallopavo - Turkey 216
Gallus gallus - Chicken 244
Oryctolagus cuniculus - Rabbit 244
Rattus norvegicus - Rat 253
Mus musculus - Mouse 253
Pongo abelii - Sumatran orangutan 255
Gorilla gorilla - Western gorilla 247
Homo sapiens - Human 255
Pan paniscus - Pygmy chimpanzee 255
Pan troglodytes - Sumatran orangutan 255
Canis lupus familiaris - Dog 250
Felis catus - Cat 250
Sus scrofa - Pig 251
Bos taurus - Bull 253
Ovis aries - Sheep 251
Capra hircus - Goat 251
Odocoileus virginianus - White-tailed deer 243
Cervus canadensis - Elk 245
Cervus hanglu - Central Asian red deer 171
Cervus elaphus - Red deer 168
Muntiacus reevesi - Chinese muntjac 186
Muntiacus muntjak - Indian muntjac 202
Bar charts
Selected proteins (n)
Genome size (Mb)
Proteins (n)
%GC
Taxonomy
Bony fishes
Frogs and toads
Reptiles
Birds
Rabbits
Rodents
Primates
Carnivores
Suids
Bovids
Deers

Figure 2

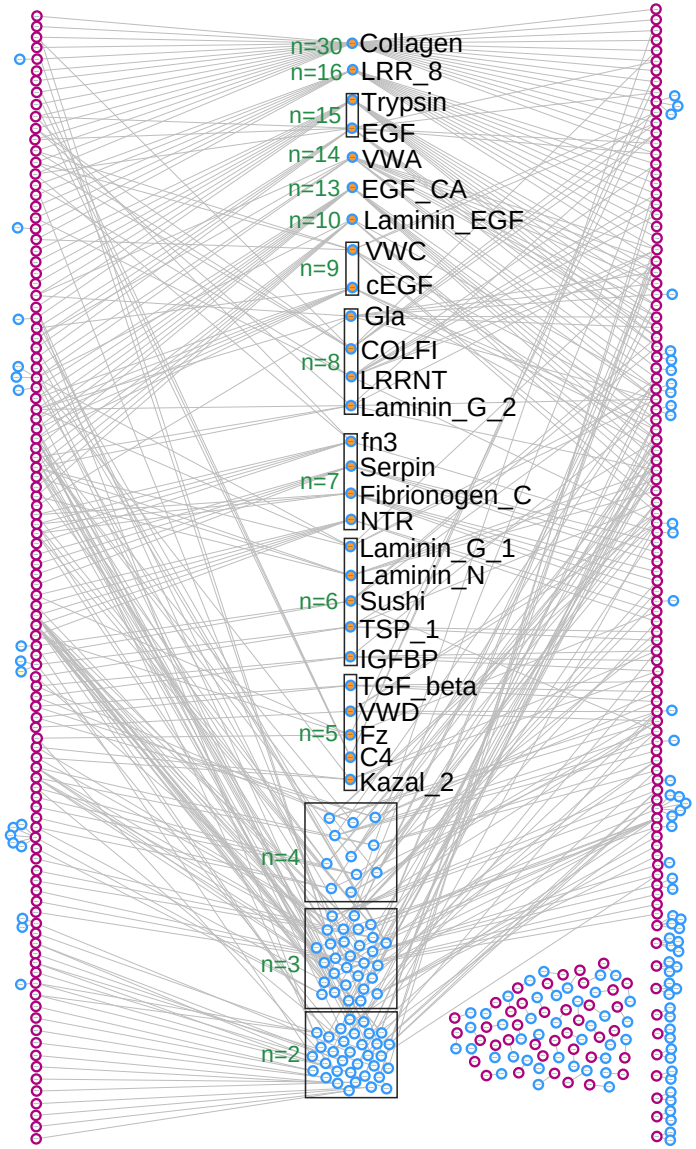

n=30
Collagen
n=16
LRR_8
n=15
Trypsin
EGF
n=14
VWA
n=13
EGF_CA
n=10
Laminin_EGF
n=9
VWC
cEGF
n=8
Gla
COLFI
LRRNT
Laminin_G_2
n=7
fn3
Serpin
Fibrionogen_C
NTR
n=6
Laminin_G_1
Laminin_N
Sushi
TSP_1
IGFBP
TGF_beta
n=5
VWD
Fz
C4
Kazal_2
n=4
n=3
n=2

Figure 3

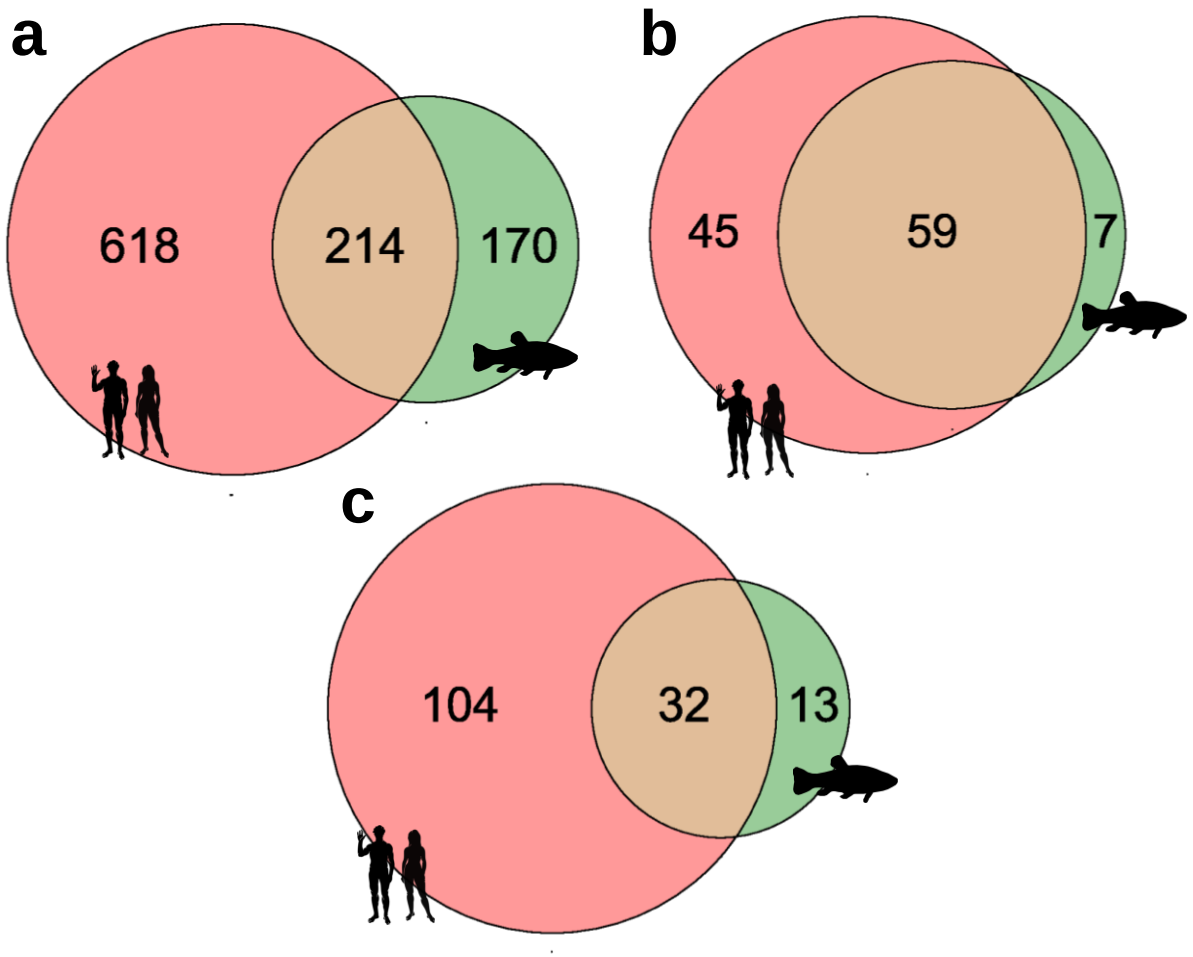


Figure 4

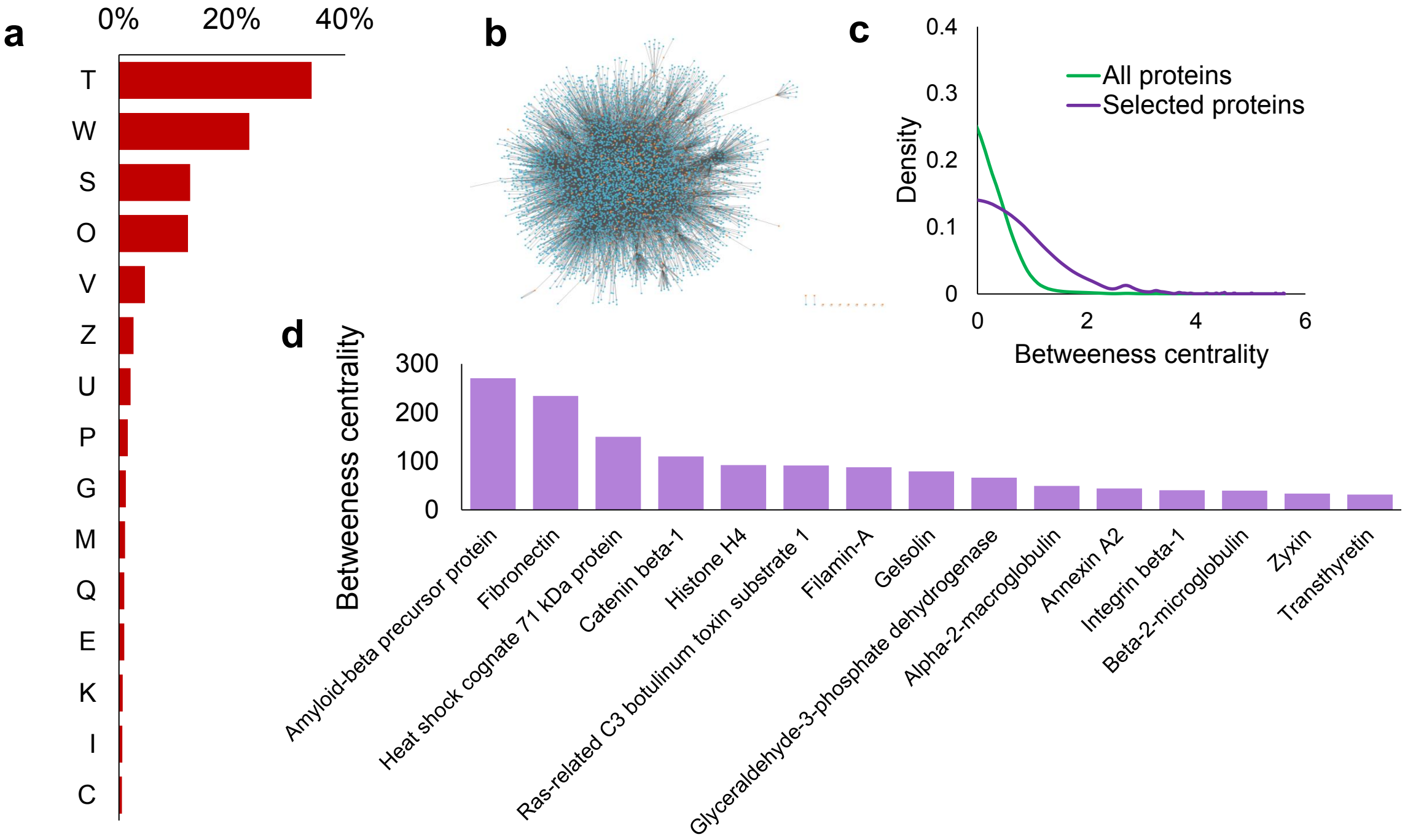
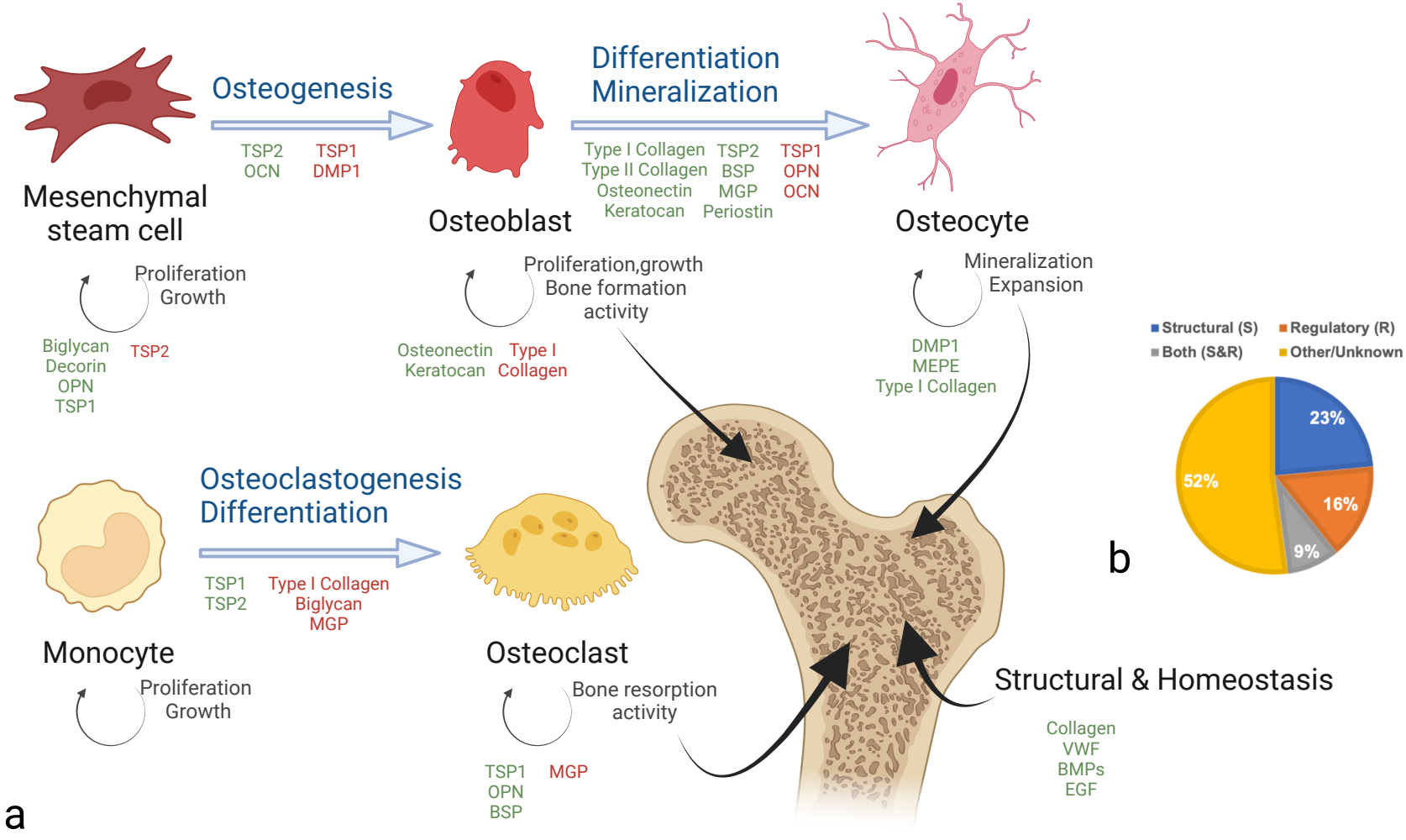

a
0%
20%
40%
T
W
S
O
V
Z
U
P
G
M
Q
E
K
I
C
b
c
0.4
0.3
0.2
0.1
0
Density
All proteins
Selected proteins
0
2
4
6
Betweeness centrality
d
Betweeness centrality
300
200
100
0
Amyloid-beta precursor protein
Fibronectin
Heat shock cognate 71 kDa protein
Catenin beta-1
Histone H4
Ras-related C3 botulinum toxin substrate 1
Filamin-A
Gelsolin
Glyceraldehyde-3-phosphate dehydrogenase
Alpha-2-macroglobulin
Annexin A2
Integrin beta-1
Beta-2-microglobulin
Zyxin
Transthyretin

Figure 5


Osteogenesis
TSP2
OCN
TSP1
DMP1
Mesenchymal steam cell
Proliferation
Growth
Biglycan
Decorin
OPN
TSP1
TSP2
Osteoblast
Proliferation,growth
Bone formation activity
Osteonectin
Keratocan
Type I Collagen
Differentiation
Mineralization
Type I Collagen
Type II Collagen
Osteonectin
Keratocan
TSP2
BSP
MGP
Periostin
TSP1
OPN
OCN
Osteocyte
Mineralization
Expansion
DMP1
MEPE
Type I Collagen
Osteoclastogenesis
Differentiation
TSP1
TSP2
Type I Collagen
Biglycan
MGP
Monocyte
Proliferation
Growth
Osteoclast
Bone resorption activity
TSP1
OPN
BSP
MGP
Structural & Homeostasis
Collagen
VWF
BMPs
EGF
a
Structural (S)
Regulatory (R)
Both (S&R)
Other/Unknown
23%
16%
9%
52%
b