



Published in final edited form as:

Microb Physiol. 2023 ; 33(1): 49–62. doi:10.1159/000531468.

Sequence Similarity among Structural Repeats in the Piezo Family of Mechanosensitive Ion Channels

Kevin J. Hendargo^a, Ashay O. Patel^a, Onyeka S. Chukwudozie^a, Gabriel Moreno-Hagelsieb^b, J. Andrés Christen^c, Arturo Medrano-Soto^a, Milton H. Saier Jr.^a

^aDepartment of Molecular Biology, School of Biological Sciences, University of California, San Diego, CA, USA

^bDepartment of Biology, Wilfrid Laurier University, Waterloo, ON, Canada

^cDepartamento de Probabilidad y Estadística, Centro de Investigación en Matemáticas, CIMAT, Guanajuato, Mexico

Abstract

Members of the Piezo family of mechanically activated cation channels are involved in multiple physiological processes in higher eukaryotes, including vascular development, cell differentiation, touch perception, hearing, and more, but they are also common in single-celled eukaryotic microorganisms. Mutations in these proteins in humans are associated with a variety of diseases, such as colorectal adenomatous polyposis, dehydrated hereditary stomatocytosis, and hereditary xerocytosis. Available 3D structures for Piezo proteins show nine regions of four transmembrane segments each that have the same fold. Despite the remarkable similarity among the nine characteristic structural repeats in the family, no significant sequence similarity among them has been reported. Using bioinformatics approaches and the Transporter Classification Database (TCDB) as reference, we reliably identified sequence similarity among repeats based on four lines of evidence: (1) hidden Markov model-profile similarities across repeats at the family level, (2) pairwise sequence similarities between different repeats across Piezo homologs, (3) Piezo-specific conserved sequence signatures that consistently identify the same regions across repeats, and (4) conserved residues that maintain the same orientation and location in 3D space.

This article is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC) (<http://www.karger.com/Services/OpenAccessLicense>). Usage and distribution for commercial purposes requires written permission.

Correspondence to: Milton H. Saier Jr., msaier@ucsd.edu, Arturo Medrano-Soto, 11medranosoto@ucsd.edu.

Author Contributions

K.J.H. performed most of the work for this project, developed programs for the project, and participated in methodology design and writing of the manuscript. A.O.P. performed initial feasibility analyses that revealed similarities among Piezo repeats. O.C. performed the feasibility analysis to identify structurally conserved/collocated residues in Piezo repeats. G.M.H. contributed to strategy design, developed programs, and performed analyses to identify HMM-profile similarities among repeats. J.A.C. designed and supervised the implementation of the strategy to estimate the upper limit for the MPSAT p value. A.M.S. supervised, designed strategies, performed analyses, developed programs for the project, and participated in writing the manuscript. M.H.S. defined the project, designed strategies, supervised the project, obtained funding, and participated in writing the manuscript.

Statement of Ethics

Ethical approval and consent were not required as this study was based on publicly available data.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Keywords

Ion channels; Piezo family; Repeat unit; Homology; Bioinformatics

Introduction

Multiple physiological processes respond to mechanical stimuli, including touch perception, pain sensation, hearing, and blood pressure regulation in animals. Mechanosensitive ion channels of the Piezo family are critical for normal physiology and have been identified in vertebrates with homologs in invertebrates, plants, and protozoa [Coste et al., 2010; Kim et al., 2012; Moroni et al., 2018; Saotome et al., 2018]. Members of the Piezo family do not show sequence and structural similarities to any other known classes of ion channels, such as voltage- or ligand-gated channels [Kawate et al., 2009; Payandeh et al., 2011; Zhang et al., 2012], transient receptor potential channels [Liao et al., 2013; Paulsen et al., 2015], prokaryotic mechanosensitive channels [Chang et al., 1998; Bass et al., 2002; Liu et al., 2009; Kung et al., 2010], or eukaryotic mechanosensitive two-pore domain potassium channels [Brohawn et al., 2012].

Mechanical cues are crucial for vascular development and proper cell differentiation. Proteins Piezo1 and Piezo2 have been established as mechanosensitive cation channels in mammals [Coste et al., 2010, 2012; Ge et al., 2015; Zhao et al., 2016], and they are expressed in various cell types, especially in vascular smooth muscle and endothelial cells [Karlsson et al., 2021; Shah et al., 2022]. For example, Piezo1 is expressed in blood vessels, is involved in sensing blood-flow-associated shear stress for blood vessel development [Ranade et al., 2014a; Li et al., 2014], calcium homeostasis, cytoskeletal dynamics, and pressure-dependent outflow [Yarishkin et al., 2021], and is necessary for maintaining perfusion in muscles [Bartoli et al., 2022]. On the other hand, Piezo2 mediates touch [Ranade et al., 2014b; Ikeda et al., 2014; Maksimovic et al., 2014; Woo et al., 2014], proprioception [Woo et al., 2015], airway stretching, and lung inflation [Nonomura et al., 2017].

Despite the high sequence and structure similarity between Piezo1 and Piezo2 (>40% identity), they display very different patterns of expression in humans and other metazoans. While Piezo1 is expressed at broadly similar levels across a wide range of cell types, Piezo2 is very strongly expressed in oligodendrocytes and undetectable in various cells for which Piezo1 expression is characteristic, for example, blood and immune cells, lung cells, cells lining the gastrointestinal tract, and eye cells [Karlsson et al., 2021].

Roughly half of all known Piezo-related pathologies are localized to the channel domain and its immediate neighbors [Wu et al., 2017], but a significant number can also be traced to mutations throughout the nine repeats present in minimally spliced Piezo isoforms. Congenital malformations related to the transmembrane repeats of Piezo1 include dehydrated hereditary stomatocytosis [Andolfo et al., 2013; Volkers et al., 2015; Wu et al., 2017], which has been linked to mutations in repeats 4, 5, 6, 8, and 9, colorectal adenomatous polyposis, which has been related to mutations in repeats 4, 6, and 8 [Wu et al., 2017], and generalized lymphatic dysplasia, which has been traced to various missense

mutations in repeats 4, 5, 6, and 9 [Wu et al., 2017]. In a number of cases, these mutations are more salient at the tissue level and have fewer detectable effects at the single-cell level, an example of which is a point substitution in repeat 4 (G718S) [Andolfo et al., 2013], known to be sufficient to effect a dehydrated hereditary stomatocytosis phenotype [Wu et al., 2017].

Functional studies show a link between mutations in Piezo1 and hereditary xerocytosis [de Meira Oliveira et al., 2020], and this channel is believed to play a role in glaucoma involving mechanical stress [Chen et al., 2022]. In part, as noted above, Piezo2 has a more narrow tissue distribution, plays a role in rapidly adapting mechanically activated currents in somatosensory neurons [Coste et al., 2010], is found in tissues that respond to physical touch such as Merkel cells [Wu et al., 2017], and is thought to regulate the response to light touch [Faucherre et al., 2013]. Mutations in Piezo2 repeats result in a smaller and more closely related set of phenotypes than those in Piezo1 repeats. The best characterized such phenotype is distal arthrogryposis type 5 with associated muscle atrophy, for which both dominant and recessive variants are known for residues in repeats: 1 [Li et al., 2018], 2 [Haliloglu et al., 2017; Li et al., 2018; Matias-Perez et al., 2018], 3 [Delle Vedove et al., 2016; Li et al., 2018], 4 [Volkers et al., 2015; Delle Vedove et al., 2016; Wu et al., 2017; Li et al., 2018], 5 [Volkers et al., 2015; Wu et al., 2017; Li et al., 2018], and 9 [Volkers et al., 2015; Wu et al., 2017; Li et al., 2018]. Many of the dominant variants appear to be gain-of-function mutations in which the resulting channel recovers more quickly from inactivation [Wu et al., 2017] or has a higher threshold for inactivation [Coste et al., 2013]. Mutations in Piezo2 have been linked to Gordon syndrome, Marden-Walker syndrome, and distal arthrogryposis [Ma et al., 2019].

In addition to mechanical stimuli, Piezo channels are also powerfully modulated by voltage and can even switch to a purely voltage-gated mode. Mutations in Piezo channels that cause diseases in humans shift voltage sensitivity in Piezo channels toward the resting membrane potential and promote voltage gating [Moroni et al., 2018]. The curved transmembrane domains of Piezo1 channels create a convex membrane deformation, which is predicted to flatten in response to increased membrane tension [Jiang et al., 2021]. Piezo channels may sense membrane tension through changes in the local curvature of the membrane [Liang and Howard, 2018]. However, they are biochemically and functionally tethered to the actin cytoskeleton via the cadherin-beta-catenin mechanotransduction complex, whose perturbation impairs Piezo-mediated responses. Mechanistically, the adhesive extracellular domain of E-cadherin interacts with the cap domain of Piezo1, which controls the transmembrane gate, while its cytosolic tail might interact with the cytosolic domains of Piezo1, which are in close proximity to its intracellular gates, allowing a direct focus of adhesion-cytoskeleton-transmitted force for gating [Wang et al., 2022].

3D structures are now available at various resolutions for both Piezo1 [Ge et al., 2015; Saotome et al., 2018; Zhao et al., 2018] and Piezo2 [Wang et al., 2019]. The extracellular domains of the trimeric propeller-like structure of Piezo channels resemble three distal blades with a central cap. The blades form three peripheral wings and a central pore module that encloses a potential ion-conducting pore. The flexible extracellular blade domains are connected to the central intracellular domain by three long beam-like structures. These

peripheral regions may act as force sensors to gate the central ion-conducting pore. In the intracellular region, three long beam-like domains support the transmembrane regions and connect the mobile peripheral regions to the central pore module. This suggests that the mechanism of Piezo trimeric complexes functions by similar principles as how propellers sense and transduce force to control ion conductivity [Ge et al., 2015]. Piezo1 and Piezo2 assemble as transmembrane triskelions to combine force sensing with regulated calcium influx. They are important for multiple physiological processes, and, as we have described above, their malfunction is associated with a variety of diseases. These channels are versatile force sensors [Beech and Kalli, 2019].

The Transporter Classification Database (TCDB) is specialized with respect to curated information about the functions and evolution of transporters from all domains of life [Saier et al., 2021]. Considering current members of the Piezo family (TC: 1.A.75) in TCDB and their homologs, Piezo1 and Piezo2 proteins are ~2,500 amino acids (aas) long, contain over 30 TMSs, consist of up to 9 (or sometimes 10) structural repeats, each with 4 TMSs [Wang et al., 2019], and require no other proteins for activity. The domain containing the beam helix (Pfam: PF15917) is confined to Animalia. The C-terminal domain (Pfam: PF12166) is also present in Viridiplantae and is made up of three hydrophobic α -helices immediately following the ninth repeat, two TMSs that make up the cation translocation channel bracketing a large extracellular domain involved in inactivation as in the calcium-gated potassium channel MthK [Fan et al., 2020] and other voltage-gated channels.

The lack of sequence similarity among repeats [Zhao et al., 2018, 2019] contrasts with the remarkable similarity of their 3D folds [Saotome et al., 2018]. Despite their sequence divergence, we here present several lines of sequence-based evidence supporting the notion that all repeats are homologous. These include (a) hidden Markov model (HMM) profile and hydropathy similarities across repeats at the family level; (b) pairwise sequence similarity between different repeats across Piezo homologs; (c) Piezo-specific conserved sequence signatures that identify the same regions across repeats; and (d) conserved potentially functional residues that maintain the same orientation and location in 3D space.

Results

Topology of the Piezo Family

To investigate the sequence diversity and architectures of Piezo homologs, we constructed multiple sequence alignments (MSAs) for each structural repeat across 279 members of the Piezo family (see Methods). Considering positions with less than 30% gaps in the MSAs, repeats are on average 130 aas long, and their average hydropathy profiles generated with the program AveHAS [Zhai and Saier, 2001] indicate that repeats have increasing levels of conservation as they get closer to the C-terminal channel domain (Fig. 1 and Methods). This can also be observed by the increasing number of highly conserved residues in the sequence logos for each individual repeat (online suppl. Fig. S1; for all online suppl. material, see <https://doi.org/10.1159/000531468>). Furthermore, except for repeat 9, all repeats have similar hydropathy profiles and TMS organizations (online suppl. Fig. S2). That is, TMSs 1–2 form a single hydrophobicity peak, while TMSs 3 and 4 are each observed within single

hydrophobicity peaks. Multiple alignments for full-length proteins and individual repeats are available in file S1 (FigShare: <https://doi.org/10.6084/m9.figshare.21638138>).

Compositional analyses with the program TMSoc [Wong et al., 2012] indicate that all TMSs in all repeats are primarily complex (online suppl. Fig. S3; see Methods), which suggests that their functional role is more important than just acting as anchors to the membrane. Online suppl. Figure S3 shows violin plots of TMSoc z-score profiles for all nine repeats. Z-scores above the green line indicate that TMSs are simple, whereas z-scores below the red line correspond to complex TMSs. Overall, TMSs 1 and 4 have higher levels of complexity across repeats than TMSs 2 and 3. This may have functional implications that remain to be explored.

Sequence-Based Evidence of Homology

To search for evidence of homology among repeats, we generated HMM profiles for the MSAs characterizing each structural repeat unit and performed pairwise comparisons with HHalign [Steinegger et al., 2019] as described in the Methods section. Table 1 shows the alignment probabilities and coverages reported by HHalign for all repeat pairs with E-value $<10^{-3}$ and coverage-weighted score (WScore) > 10 (see Methods). Repeat 9 had no matches better than our thresholds with any of the other repeats and was not further considered. The lack of sequence similarity of repeat 9 with the other repeats could be explained in part by sequence differences that split the hydrophobic peak spanning the first two TMSs in repeats 1–8 into two clear hydrophobic peaks in repeat 9. The length of the loop between TMSs 1 and 2 in the ninth repeat (Fig. 1; online suppl. Fig. S2) is on average 3 times longer than the loops between TMSs 1 and 2 in the first eight repeats, such that its hydrophathy, averaged over 19 residues, actually falls below zero, contrasting sharply with the first 8 repeats. This produces alignments with poor coverage and pushes scores below significance thresholds. In addition, repeat 9 is the closest to the channel domain (Fig. 1), and although different from the other repeats, it has the highest level of conservation in the family. This suggests that stronger selective constraints may exist which restrict the sequence and functional diversification of repeat 9.

We used the WScore to generate a network plot illustrating the strength of the connections among the repeats (Fig. 2). The figure reveals that repeat 5 has the strongest connections with repeats 4, 6, 7, and 8. Repeat 3 has the weakest connections. To have a better idea of the relationships among repeats, we performed hierarchical clustering based on the WScores and generated a dendrogram (Fig. 3; see Methods). The clustering structure is good (agglomerative coefficient: 0.75), supporting the closer relationships between TMSs 4–8 and the weaker relationships between TMSs 1–3. This correlates well with the level of conservation observed in the repeats (Fig. 1 and online suppl. Fig. S1), as well as the strength of their connections in the network (Fig. 2).

Sequence similarity among repeats can also be observed when we compare the sequences of individual repeats across different Piezo homologs. Given that transmembrane regions have compositional biases that can artificially increase alignment scores [Wong et al., 2010, 2011], we consider 2 repeats to be homologous only if their alignment has high coverage and has better scores than the alignments obtained within a negative control set

of randomized nonhomologous sequences that have similar TMS topologies, amino acid compositions, and lengths (see Methods). This is illustrated in online suppl. Figure S4 using hydropathy plots to show that repeat 2 in Piezo homolog XP_013404042 (panel A) has the same number of hydrophobic peaks and TMS locations as 5 of its shuffled versions (panels B–F). Our program Membrane Protein Sequence Alignment Tool (MPSAT) compares the coverage-weighted alignment score (WScore) between a query and subject repeats to the distribution of scores obtained between the query repeat and shuffled versions of the subject repeat, finally returning the corresponding *p* value of the reference alignment (see Methods). Figure 4 shows six examples of alignments involving repeats 1–8 from different Piezo homologs that support the results obtained with HHalign (Table 1; Fig. 2, 3). For each panel in Figure 4, out of 10,000 coverage-weighted scores produced with the shuffled sequences, no single score was better than the reference score. Because alignments with shuffled sequences tend to have poor coverage (see Methods), the net result of shuffling sequences while preserving their TMSs' positions and adjusting the alignment scores with their coverage involves a tendency toward increasing the difference of alignment scores between homologous sequences and the negative control (see online suppl. Fig. S5, S6). This is desirable to better discriminate homologous repeats.

We searched for conserved motifs across repeats using the MEME suite [Bailey et al., 2015] (see Methods). We identified three regions of 25 aas (Fig. 5) that map consistently to the same locations within repeats but have few highly conserved residues across all repeats. Although each repeat has highly conserved residues (online suppl. Fig. S1), most are poorly conserved across repeats. However, due to their high specificity, we regarded these three “motifs” as conserved signature sequences (CSs). Figure 5 shows the sequence logos of the three identified CSs, and online suppl. Figure S7 illustrates where the CSs match in the context of the hydropathy profiles for two representative Piezo members with solved structures: Q8CD54 (TC: 1.A.75.1.2; PDB: 6KG7) and E2JF22 (TC: 1.A.75.1.14; PDB: 5Z10). It can be observed in online suppl. Figure S7 that repeats 1–8 share the same pattern: CS 1 (blue bars) matches the third TMS, CS 2 (red bars) covers a region central to TMSs 1–2, and CS 3 (green bars) matches the fourth TMS. Online suppl. Figure S7 also shows that although repeat 9 has high scoring hits with the three CSs (MAST *p* values $<10^{-10}$), it has a different matching pattern compared to repeats 1–8. That is, CS 2 matches TMS 1, and CS 3 matches TMS 4, as expected, but CS 1 matches TMS 2, and TMS 3 is not matched by any CS. The atypical CS matching pattern of repeat 9 is conserved throughout the Piezo family and correlates with the aforementioned observation that this repeat did not have high coverage similarities above our significance thresholds with the other eight repeats.

As a negative control, we scanned for the three CSs against all proteins in TCDB, excluding the Piezo family. No MAST hits with E-value $<10^{-3}$ were identified in the negative control, indicating that these CSs are specific to the Piezo family. The consistent mapping of the CSs to the same regions within repeats and the lack of hits within the negative control further support the hypothesis that the repeats are homologous. File S2 (FigShare: <https://doi.org/10.6084/m9.figshare.21638204>) contains the results of running MEME on our training set and MAST on our test set and negative control.

Structural Evidence of Homology

Consistent with previous reports [Zhao et al., 2018], TMalign superpositions indicate that Piezo repeats all share the same fold with TM-score >0.55 and RMSDs <3.83 Å (online suppl. Table S1). Figure 6 shows the structural relationships among repeats with a dendrogram of the superposition RMSDs (agglomerative coefficient: 0.34). Despite the RMSD dendrogram having a much weaker clustering structure than the dendrogram obtained based on HMM profile-profile alignment scores (agglomerative coefficient: 0.75; Fig. 3), both dendrograms show topological similarities. In both dendrograms, repeats 5, 7, 8, and 4 share a major branch, and repeats 1–3 are more distant. With the exception of repeat 6, the major differences are observed among repeats that show the weakest relationships (repeats 1–3), which may contribute more to the weaker clustering structure of the RMSD dendrogram in Figure 6. This indicates that RMSDs also have problems identifying the relationships of the most distant repeats.

Because the overall fold is the same for any two Piezo repeats disregarding the RMSD values, they were all aligned to repeat 6, which is close to the “center” of the folds of the repeats in conformational space (see Methods and online suppl. Table S1). After discarding residue classes unlikely to perform functional roles (see Methods), we identified a list of residues that appeared in both Piezo1 and Piezo2 in the same TMS (online suppl. Table S2). These residues were further inspected for relative position and orientation in 3D space with respect to the membrane plane (see Methods). We found that Tyr/Trp account for 4 of the 9 surviving residue classes, which may be a function of their abundance in buried contexts. Cysteine alone made up another 3 classes, all of which survived the repeat presence filter. Although Cys may be prevalent because of its natural enrichment in membrane environments, the high conservation observed for some cysteines may be a result of intrinsically high selective pressure to maintain a specific number/placement of the residue. The only charged residue surviving this part of the analysis, aspartate, was buried such that it participated only in intra-repeat interactions. Residues conserved in at least three repeats with similar orientations are listed in Table 2.

Of the key residues identified in this analysis, listed in Table 2 and displayed in online suppl. Figure S6, those participating in inter-repeat contacts are more likely than those participating in intra-repeat contacts or only intermolecular contacts to be closely involved in transmitting stresses from the periphery of the Piezo complex to the channel domain. However, most of the residues we identified interact not with residues found in neighboring repeats but with lipids or with other residues on the same repeat. Only certain serine residues found in helix 4 of repeats 6, 7, and 8 (online suppl. Fig. S8A, S9F) and tyrosine residues also found in helix 4 but also in repeats 4, 5, and 9 (online suppl. Fig. S8B, S9G) passed our thresholds and were inferred as making inter-repeat contacts. These serine residues do not make particularly close contacts with previous repeats, nor are their possible interacting partners conserved; as a result, the relevance of their collocation is in question. The tyrosine residues appear to be participating in cation- π interactions with basic residues on neighboring repeats, but this is likely to be true of any such amphiphilic residue present around the band associated with lipid head groups, and such a similarity is not much more indicative of homology than the presence of hydrophobic residues deep inside a TMS. The frequent appearance of tyrosine

residues in transmembrane helices is well known and is readily apparent in the emission probabilities given for transmembrane environments for transmembrane helix prediction tools such as HMMTOP and TMHMM [Tusnady and Simon, 1998; Krogh et al., 2001].

For intra-repeat contacts, the differences between repeats are likely just as crucial as the similarities in ensuring the proper folding of the repeats. As energetic differences between conformations are lower in proteins containing identical repeats, reaching a physiologically appropriate fold is significantly easier when incorrect folds are appropriately penalized, for example, by minimizing favorable off-target self-interactions [Li et al., 2018]. In helix 1, only a cysteine residue present in repeats 4, 6, and 8 passed our criteria. It points to helix 4, which contains cysteine residues in appropriate locations in repeats 4 and 6, but it does not form disulfide bonds under the conditions used to solve either 5Z10 or 6KG7. As these cysteines are located on the cytoplasmic side of the protein, this is not surprising; it has long been known that among metazoa, the interior of a cell is a more reducing environment than the exterior.

Aside from this cysteine, a number of aromatic residues occur on the same helix in different helices which all point toward the core. This is not surprising, and neither are the interactions they participate in, which range from pi stacking to cation-pi interactions. More interesting is a core-facing aspartate occurring in repeats 6, 7, and 8 in helix 4 which appears to be buried relative to the band of basic residues interacting with extracellular head groups. Acidic residues are relatively depleted in alpha-helical membrane proteins, although less so in extracellular contexts [Baker et al., 2017], and although the exact implications of the positions of those residues elude us at this level of analysis, it is possible that aspartate plays a key role in stabilizing these repeats, which are close to, but not adjacent to, the channel or ball domain.

The lack of variety in crystallographically solved structures for many families remains an obstacle to the identification of allosteric functional residues. Even in cases where sufficient structural data exist, it is often the case that, taken together, extant structures seldom form a sufficiently representative sample with adequate coverage in all major clusters of interest. As the only reasonably complete Piezo structures solved to date are currently two Piezo paralogs from mice, the extrapolation of results obtained using these two proteins from this model organism to the rest of the Piezo family is not straightforward. However, some degree of extrapolation should be possible because residues from 5Z10 and 6KG7 in trimmed MSAs are, on average, similar with 42% and 38% identity, respectively, to residues in the rest of the Piezo homologs (online suppl. Fig. S10). Potential approaches to circumventing these issues include homology modeling for reasonably close homologs and the use of high-confidence AlphaFold predictions to expand the library of accessible structures. However, the use of AlphaFold2 predictions is not without its risks (e.g., spurious conservation of structure or the development of biologically implausible folds).

Discussion

It has been observed that Piezo repeats bear little sequence similarity with each other [Zhao et al., 2018, 2019]; to our knowledge, no formal analysis has been reported that

identifies homology among repeats at the primary sequence level in this family. Here we present evidence indicating that all repeats in the Piezo family are homologous. Homology is supported by (1) significant alignment scores between HMM profiles at the whole family level (Table 1; Fig. 2); (2) similarity of hydrophathy profiles (online suppl. Fig. S1, S2); (3) pairwise sequence alignments between different repeats across Piezo homologs, where significance was computed using a negative control set of shuffled sequences that preserve the length, TMS topology, and composition of Piezo repeats (see online suppl. Fig. S4, S5, S6 and Methods); (4) identification of sequence signatures (Fig. 5) that consistently identify the same regions across repeats and Piezo homologs (online suppl. Fig. S7); and (5) conservation of residues and their locations/orientations in 3D space across repeats (Table 2; online suppl. Fig. S8–S9). Whether the results obtained from the analysis of structures 5Z10 and 6KG7 from mice translate well to Piezo proteins in other species for which no structures have been solved remains an important question. The observed degrees of residue conservation within repeats across the family (online suppl. Fig. S1, S10) suggest that extrapolation to other organisms is reasonable. We are confident that the evidence herein presented reliably shows that homology among Piezo repeats can be detected at the primary sequence level, confirming the general conclusion based on 3D structural data.

Methods

Extraction of Homologs

Homologs of systems belonging to the Piezo family (TC# 1.A.75) were retrieved from the NCBI nonredundant protein sequence database using the program famXpander [Medrano-Soto et al., 2018]. We selected 1,229 hits that showed a minimal coverage of 50%, E-value $<10^{-5}$, and redundant sequences were removed with CD-HIT [Fu et al., 2012] using a 90% identity threshold.

Construction of Multiple Alignments

Given that Piezo proteins are thousands of residues long, in order to construct multiple alignments, we randomly selected 260 homologs of the Piezo family that showed at least 70% coverage, E-value $<10^{-20}$ and were nonredundant using a 70% identity threshold. In addition, we included all 21 members of the Piezo family in TCDB. The resulting 281 homologs were aligned using the L-INS-i algorithm as implemented in MAFFT [Rozewicki et al., 2019]. The program TrimAl [Capella-Gutierrez et al., 2009] was used to remove positions with gap fractions greater than 30% and to cut the resulting multiple alignment into bundles of four TMSs that correspond to each of the nine known structural four-helix repeats. Two sequences were further removed because they aligned poorly in repeat regions. File S1 multiple alignments for full-length proteins and individual repeats are available in File S1 (FigShare: <https://doi.org/10.6084/m9.figshare.21638138>) contains the MSA for full-length proteins. Cut points were selected to minimize contributions from inter-repeat loops and non-repeat membrane-associated helices (as occurs between repeat 9 and the channel domain) without truncating transmembrane helices. File S1 also contains the MSAs for each individual repeat. For each MSA sequence, logos were constructed using the program WebLogo [Crooks et al., 2004], and regions covered by individual TMSs were highlighted with tan boxes (online suppl. Fig. S1).

MEME/MAST Searches for CSs

From the MSAs constructed for each repeat, random samples of 50 sequences were taken for motif discovery using the MEME suite [Bailey et al., 2015]. At most, 5 motifs per sequence were searched for all widths between 10 and 50 amino acids, showing E-values $<10^{-10}$, and using both modes zoops (0–1 match per sequence) and oops (1 match per sequence). For verification, MAST [Bailey et al., 2015] was used to scan each motif produced by MEME against the remainder of the 1,229 Piezo homologs identified with famXpander (E-value $<10^{-3}$). All of the 1,741 families in TCDB, except Piezo, were used as negative controls. Three motifs of 25 aas were identified (Fig. 5), which recovered 1,200 Piezo members (97.6%) while returning zero hits with E-value $<10^{-3}$ within the negative control (see File S2; FigShare: <https://doi.org/10.6084/m9.figshare.21638204>). A couple of examples are shown in online suppl. Figure S7, which illustrates the locations where the three motifs consistently match in the context of the hydrophathy profiles of two Piezo members with solved structures: Q8CD54 (TC: 1.A.75.1.2; PDB: 6KG7) and E2JF22 (TC: 1.A.75.1.14; PDB: 5Z10).

HMM Profile-Profile Comparisons of Repeat Units

For each repeat unit, trimmed MSAs were compiled into HMM profiles with the program hmake and aligned against each other with HHalign from the HH-suite [Steinegger et al., 2019]. To adjust our confidence levels of alignments between pairs of repeats showing E-value $<10^{-3}$, we calculated a weighted score (WScore) where the probabilities reported by HHalign (as percentages) were multiplied by the coverage of the alignments relative to the shorter MSA to correct for alignment length. This is critical for comparisons of short, well-defined domains, in which low coverage is more often a sign of misalignments than of genuine local similarities. The WScore was used to construct a network graph using Gephi (<https://gephi.org/>) that illustrates the strength of the connections among the repeats (Fig. 2) and to perform hierarchical clustering analysis to reveal their relatedness. Clustering analysis was carried out using the Ward algorithm (agglomerative coefficient: 0.75) as implemented in R (<https://www.r-project.org/>). Figure 3 illustrates the result as a rectangular dendrogram generated with FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>). Repeat 9 was excluded from the plots because all alignments with other repeats reported E-values $>10^{-3}$ and WScores <10 .

Pairwise Sequence Comparisons of Repeats

To confirm the results of HMM profile-profile comparisons and MEME/MAST signature sequence identification, we performed individual pairwise sequence alignments between different repeats across Piezo homologs. Because unrelated membrane proteins may yield alignment scores beyond thresholds of significance due to inherent local compositional biases in TMSs [Wong et al., 2010, 2011], we estimated the significance of pairwise alignments by taking into consideration the TMS topologies of the repeats being compared. We developed the program MPSAT to estimate a *p* value of the alignment score between the reference query and subject proteins (reference score) given a negative control set of scores obtained between the query protein and nonhomologous shuffled versions of the subject protein that preserve the same number of TMSs, locations of TMSs, and amino

acid composition of the subject protein. Residues in the subject protein belonging to a particular class (e.g., TMS and not-TMS) were shuffled exclusively with other residues in the same class, which by default ignores outside/inside and loop/tail differences. As a result, a shuffled sequence has a much more similar hydropathy profile to the reference protein than a completely random shuffle would yield. Online suppl. Figure S4 illustrates this concept by showing the hydropathy of repeat 2 from Piezo homolog XP_013404042 alongside the hydropathies of five shuffled versions of its sequence. The query protein is aligned against shuffled versions of the subject protein using the Smith-Waterman (S-W) algorithm as implemented in ssearch36 [Pearson, 1991]. The raw S-W scores of the aligned regions follow an extreme value distribution (EVD) as expected from theory [Altschul et al., 1997; Pearson, 1998]. Online suppl. Figures S5A and S6A show that an EVD closely fits the histogram of the scores from shuffled alignments and highlights where the reference score is located relative to the EVD (an orange dot on the X-axis). A p value can be easily calculated by integrating the EVD function from the reference score to infinity. This quantifies the probability that the reference alignment score between the original query/subject proteins is no different from alignments performed between the query and shuffled proteins. If an alignment is significant, it must yield a better S-W score than alignments with nonhomologous shuffled sequences that have the same TMS topology and similar compositional biases as the subject sequence.

We noted that a large majority of the scores derived from shuffled alignments have poor coverage, so we adjusted the S-W score by multiplying it by the coverage. This discriminated the reference scores from random alignments better because the distribution of shuffled scores shifts conspicuously toward lower scores while the reference scores, having larger coverages, barely move (online suppl. Fig. S5, S6). However, the distribution of coverage-adjusted scores does not always follow an EVD as illustrated in online suppl. Figures S5B and S6B. In these cases, estimating the p value by fitting and integrating a Gaussian Kernel Density function did not work as expected. Even though the distribution of adjusted scores is well fitted by the Gaussian KDE, the reference score adjusted by coverage is, for multiple cases, far enough from the negative control's distribution that the Gaussian Kernel Density function returns p value = 0.0 as an artifact due to the lack of data points (shuffled scores) near the reference score (online suppl. Fig. S5B, S6B). The larger distances between the reference scores and the distributions of shuffled coverage-adjusted scores indicate that the corresponding p value should be smaller than the one estimated from raw scores using the EVD (online suppl. Fig. S5, S6). Given the uncertainty of the nature of the distribution of coverage-adjusted scores, we estimated an upper boundary for the p value based on the scores obtained with the shuffled sequences.

Based on the alignment scores (S_i) obtained between a query protein and the shuffled versions of the subject protein, we estimate an empirical p value (p) describing the probability that the score observed between the query and subject proteins (s_o) can be found by chance in a negative control set of size n , namely,

$$p = \mathbb{P}[S_i > s_o \mid H_0]$$

where S_i ; $i = 1, \dots, n$ represents the observed alignment scores under the null hypothesis H_0 . This problem may be formulated as Bernoulli sampling, that is, let

$$Y_i = I(S_i > s_o) \quad 2$$

be a function that returns $Y_i = 0$ if $S_i < s_o$ and $Y_i = 1$ otherwise. Indeed, given the null hypothesis, this function follows the Bernoulli distribution

$$Y_i | H_0 \sim \text{Be}(p) \quad 3$$

where p is the p value to be estimated. In Bayesian estimation, a random variable is constructed, for example, P , whose distribution represents our uncertainty on the value of p . We may then rewrite the distribution in (equation 3) as

$$Y_i | P = p \sim \text{Be}(p) \quad 4$$

To avoid making strong assumptions about the prior distribution of p , we use a noninformative conjugate beta prior for $P \sim \text{Beta}(\alpha, \beta)$. The analytical solution for the posterior distribution of a Bernoulli-Beta process is also a beta distribution [Lee, 2012] and has the form

$$P | S_i = s_i \sim \text{Beta}(\alpha + r, \beta + n - r) \quad 5$$

where $\alpha = 1, \beta = 1, i = 1, \dots, n$ and r is calculated based on (equation 2) as $r = \sum_{i=1}^n Y_i; r \geq 0$.

An upper limit for the target p value (p_m) can be inferred from the posterior distribution of P (equation 5) as the quantile q_m with probability.

$$\mathbb{P}[P < p_m | S_i = s_o] = 0.999$$

That is, we can say that when $S_i = s_o$, the p value $p < p_m$ with 0.999 certainty. The quantile q_m was calculated with the Python libraries NumPy and SciPy.

Analysis of TMS Sequence Complexity

Based on the AveHAS algorithm [Zhai and Saier, 2001], average hydropathy profiles were generated for the MSAs corresponding to each of the nine repeats. We used our program TMweaver to obtain the TMS boundaries by projecting the alignment coordinates of hydrophobic peaks (putative TMSs) to each sequence. TMweaver uses an interactive

graphical interface to plot the hydropathy of a sequence or MSA, allowing the user to define regions in the protein(s) (e.g., hydrophobic peaks, hydrophilic domains, etc.). The program then returns the specific boundaries of the defined regions for each input protein in several formats. After this step, only repeat sequences where all TMSs are at least 14 aas long were kept due to the smallest TMS size observed in the two available Piezo structures (PDBs: 5Z10 and 6KG7). The program TMSoC [Wong et al., 2012] was slightly modified to allow the computation of complexity for TMSs with lengths under 19 aas. The z-scores reported by TMSoC were used to generate violin plots (online suppl. Fig. S3) that include the thresholds used by Wong et al. [Wong et al., 2012] to classify TMSs into complex (z-score > -3.29) or simple (z-score < -5.41). TMweaver is available via our public repository (<https://github.com/SaierLaboratory>).

Structural Analysis of Piezo Repeats

Nonoverlapping four-helix bundles were cut out for all solved Piezo structures containing repeats with assistance from the OPM's [Lomize et al., 2012] indices. Figure 1 is a representation of the cutting points for each repeat within the context of the average hydropathy profile of the Piezo family. All repeat bundles in 5Z10 and 6KG7 were aligned against each other using TMalign [Zhang and Skolnick, 2005], and the pairwise score matrices using both TM-scores and RMSDs are shown in online suppl. Table S1. A dendrogram based on the RMSDs in online suppl. Table S1 was built with the Ward algorithm as implemented in R (Fig. 6).

A search for a suitable structural representative with which to anchor further analyses initially suggested repeat 5, whose WScore scores (see Table 1; Fig. 2) and TM-scores (online suppl. Table S1) are higher than those of other repeats. However, inspection of one-vs-all superpositions relative to repeats 5 and 6 suggested that repeat 6 is a better reference to search for conserved/collocated residues. The lowest TM-score observed between repeat 6 and other repeats was 0.62, and the largest RMSD value was 3.34 Å (online suppl. Table S1).

Identification of Conserved/Collocated Residues

We analyzed two available Piezo1 (PDB: 5Z10; resolution: 3.97 Å) and Piezo2 (PDB: 6KG7; resolution: 3.80 Å) structures for their levels of completeness because they include the structural repeats targeted in this study. We searched for possible functional residues with similar orientation shared between repeats of the two structures. Three classes of residues were initially ignored in this analysis: (a) residues occurring outside TMS boundaries marked as loop residues; (b) the most common residues found in transmembrane regions due to their low projected signal-to-noise ratios in membrane environments; and (c) residues that failed to appear in both 5Z10 and 6KG7 in the same repeat and in the same helix. Of the surviving residues, we selected those that appeared in the same helix and same repeat in both structures (online suppl. Table S2). Next, residues that did not co-occur in at least three repeats in the same helix in both structures were discarded as they were deemed unlikely to perform similar functional roles. Owing to their rarity in membrane environments, charged amino acids (particularly the acidic residues Glu and Asp) were deemed to be the most likely to provide useful information on inter-repeat relationships.

Residues passing these criteria were further inspected for their relative positions in 3D space with respect to the membrane plane. Residues beyond two turns away (i.e., 7–8 residues) were not considered to be part of the same class barring cases with exceptional agreement in orientation. Residues with angular differences of more than 90 degrees were similarly not considered to be part of the same residue class. As summarized in Table 2, this process identified Cys in helix 1, Tyr, Cys, and Trp in helix 2, Trp in helix 3, and Ser, Tyr, Asp, and Cys in helix 4. These residues were regarded as potentially functionally significant.

Conservation of Residues in the Piezo Family and Available Structures

Based on the MSA of the Piezo family, the number of sequences with identical residues and positions in structures 5Z10 and 6KG7 was noted and divided by the number of sequences in the multiple alignments excluding the two structures. The resulting one-dimensional vector of fractions was further averaged with a window of 19 residues, spanning full helices, to yield the mean sequence-MSA identity for the purpose of locating regions of higher or lower conservation relative to the structures. The results of this analysis are presented in online suppl. Figure S10.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding Sources

This work was supported by Grant GM077402 to M.H.S. from the National Institutes of Health (<https://www.nih.gov/>) and by UCSD Academic Senate Grant RG104506 for the support of K.J.H. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability Statement

All data used in this study are available as supplementary material in this article, the TCDB website (<https://tcdb.org>), and FigShare (<https://doi.org/10.6084/m9.figshare.c.6321386>). Programs developed for this project are available in the Saier Lab's software repository (<https://github.com/SaierLaboratory>). Further inquiries can be directed to the corresponding authors.

References

- Altschu SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. . Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25(17):3389–402. [PubMed: 9254694]
- Andolfo I, Alper SL, De Franceschi L, Auriemma C, Russo R, De Falco L, et al. Multiple clinical forms of dehydrated hereditary stomatocytosis arise from mutations in PIEZO1. *Blood.* 2013;121(19):3925–35, S1–12. [PubMed: 23479567]
- Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. *Nucleic Acids Res.* 2015; 43(W1):W39–49. [PubMed: 25953851]
- Baker JA, Wong WC, Eisenhaber B, Warwicker J, Eisenhaber F. Charged residues next to transmembrane regions revisited: “Positive-inside rule” is complemented by the “negative inside depletion/outside enrichment rule. *BMC Biol.* 2017;15(1):66. [PubMed: 28738801]

- Bartoli F, Debant M, Chuntharpursat-Bon E, Evans EL, Musialowski KE, Parsonage G, et al. Endothelial Piezo1 sustains muscle capillary density and contributes to physical activity. *J Clin Invest*. 2022;132(5): e141775.
- Bass RB, Strop P, Barclay M, Rees DC. Crystal structure of *Escherichia coli* MscS, a voltage-modulated and mechanosensitive channel. *Science*. 2002;298(5598):1582–7. [PubMed: 12446901]
- Beech DJ, Kalli AC. Force sensing by piezo channels in cardiovascular Health and disease. *Arterioscler Thromb Vasc Biol*. 2019;39(11): 2228–39. [PubMed: 31533470]
- Brohawn SG, del Marmol J, MacKinnon R. Crystal structure of the human K2P TRAAK, a lipid- and mechano-sensitive K⁺ ion channel. *Science*. 2012;335(6067):436–41. [PubMed: 22282805]
- Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25(15):1972–3. [PubMed: 19505945]
- Chang G, Spencer RH, Lee AT, Barclay MT, Rees DC. Structure of the MscL homolog from *Mycobacterium tuberculosis*: a gated mechanosensitive ion channel. *Science*. 1998; 282(5397):2220–6. [PubMed: 9856938]
- Chen Y, Su Y, Wang F. The Piezo1 ion channel in glaucoma: a new perspective on mechanical stress. *Hum Cell*. 2022;35(5):1307–22. [PubMed: 35767143]
- Coste B, Houge G, Murray MF, Stitzel N, Bandell M, Giovanni MA, et al. Gain-of-function mutations in the mechanically activated ion channel PIEZO2 cause a subtype of Distal Arthrogryposis. *Proc Natl Acad Sci U S A*. 2013;110(12):4667–72. [PubMed: 23487782]
- Coste B, Mathur J, Schmidt M, Earley TJ, Ranade S, Petrus MJ, et al. Piezo1 and Piezo2 are essential components of distinct mechanically activated cation channels. *Science*. 2010; 330(6000):55–60. [PubMed: 20813920]
- Coste B, Xiao B, Santos JS, Syeda R, Grandl J, Spencer KS, et al. Piezo proteins are pore-forming subunits of mechanically activated channels. *Nature*. 2012;483(7388):176–81. [PubMed: 22343900]
- Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*. 2004;14(6):1188–90. [PubMed: 15173120]
- de Meira Oliveira P, Balan A, Muto NH, Cervato MC, Fonseca GHH, Sukanuma LM, et al. Heterogeneous phenotype of Hereditary Xerocytosis in association with PIEZO1 variants. *Blood Cells Mol Dis*. 2020;82:102413.
- Delle Vedove A, Storbeck M, Heller R, Holker I, Hebbar M, Shukla A, et al. Biallelic loss of proprioception-related PIEZO2 causes muscular atrophy with perinatal respiratory distress, arthrogryposis, and scoliosis. *Am J Hum Genet*. 2016;99(6):1406–8. [PubMed: 27912047]
- Fan C, Sukomon N, Flood E, Rheinberger J, Allen TW, Nimigeon CM. Ball-and-chain inactivation in a calcium-gated potassium channel. *Nature*. 2020;580(7802):288–93. [PubMed: 32269335]
- Faucherre A, Nargeot J, Mangoni ME, Jopling C. piezo2b regulates vertebrate light touch response. *J Neurosci*. 2013;33(43):17089–94. [PubMed: 24155313]
- Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 2012;28(23):3150–2. [PubMed: 23060610]
- Ge J, Li W, Zhao Q, Li N, Chen M, Zhi P, et al. Architecture of the mammalian mechanosensitive Piezo1 channel. *Nature*. 2015; 527(7576):64–9. [PubMed: 26390154]
- Haliloglu G, Becker K, Temucin C, Talim B, Küçük ahin N, Pergande M, et al. Recessive PIEZO2 stop mutation causes distal arthrogryposis with distal muscle weakness, scoliosis and proprioception defects. *J Hum Genet*. 2017;62(4):497–501. [PubMed: 27974811]
- Ikeda R, Cha M, Ling J, Jia Z, Coyle D, Gu JG. Merkel cells transduce and encode tactile stimuli to drive A β -afferent impulses. *Cell*. 2014;157(3):664–75. [PubMed: 24746027]
- Jiang Y, Yang X, Jiang J, Xiao B. Structural designs and mechanogating mechanisms of the mechanosensitive piezo channels. *Trends Biochem Sci*. 2021;46(6):472–88. [PubMed: 33610426]
- Karlsson M, Zhang C, Mear L, Zhong W, Digre A, Katona B, et al. A single-cell type transcriptomics map of human tissues. *Sci Adv*. 2021; 7(31):eabh2169.
- Kawate T, Michel JC, Birdsong WT, Gouaux E. Crystal structure of the ATP-gated P2X(4) ion channel in the closed state. *Nature*. 2009; 460(7255):592–8. [PubMed: 19641588]

- Kim SE, Coste B, Chadha A, Cook B, Patapoutian A. The role of *Drosophila* Piezo in mechanical nociception. *Nature*. 2012;483(7388):209–12. [PubMed: 22343891]
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*. 2001; 305(3):567–80. [PubMed: 11152613]
- Kung C, Martinac B, Sukharev S. Mechanosensitive channels in microbes. *Annu Rev Microbiol*. 2010;64:313–29. [PubMed: 20825352]
- Lee PM. *Bayesian Statistics: An Introduction*. Chichester, West Sussex; Hoboken, N.J.: John Wiley & Sons; 2012.
- Li J, Hou B, Tumova S, Muraki K, Bruns A, Ludlow MJ, et al. Piezo1 integration of vascular architecture with physiological force. *Nature*. 2014;515(7526):279–82. [PubMed: 25119035]
- Li S, You Y, Gao J, Mao B, Cao Y, Zhao X, et al. Novel mutations in TPM2 and PIEZO2 are responsible for distal arthrogryposis (DA) 2B and mild DA in two Chinese families. *BMC Med Genet*. 2018;19(1):179. [PubMed: 30285720]
- Liang X, Howard J. Structural biology: piezo senses tension through curvature. *Curr Biol*. 2018;28(8):R357–9. [PubMed: 29689211]
- Liao M, Cao E, Julius D, Cheng Y. Structure of the TRPV1 ion channel determined by electron cryomicroscopy. *Nature*. 2013;504(7478):107–12. [PubMed: 24305160]
- Liu Z, Gandhi CS, Rees DC. Structure of a tetrameric MscL in an expanded intermediate state. *Nature*. 2009;461(7260):120–4. [PubMed: 19701184]
- Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res*. 2012; 40(Database issue):D370–6. [PubMed: 21890895]
- Ma Y, Zhao Y, Cai Z, Hao X. Mutations in PIEZO2 contribute to Gordon syndrome, Marden-Walker syndrome and distal arthrogryposis: a bioinformatics analysis of mechanisms. *Exp Ther Med*. 2019;17(5): 3518–24. [PubMed: 30988732]
- Maksimovic S, Nakatani M, Baba Y, Nelson AM, Marshall KL, Wellnitz SA, et al. Epidermal Merkel cells are mechanosensory cells that tune mammalian touch receptors. *Nature*. 2014;509(7502):617–21. [PubMed: 24717432]
- Matias-Perez D, Garcia-Montano LA, Cruz-Aguilar M, Garcia-Montalvo IA, Nava-Valdez J, Barragan-Arevalo T, et al. Identification of novel pathogenic variants and novel gene-phenotype correlations in Mexican subjects with microphthalmia and/or anophthalmia by next-generation sequencing. *J Hum Genet*. 2018;63(11):1169–80. [PubMed: 30181649]
- Medrano-Soto A, Moreno-Hagelsieb G, McLaughlin D, Ye ZS, Hendargo KJ, Saier MH Jr. Bioinformatic characterization of the Anoctamin Superfamily of Ca²⁺-activated ion channels and lipid scramblases. *PLoS One*. 2018;13(3):e0192851.
- Moroni M, Servin-Vences MR, Fleischer R, Sanchez-Carranza O, Lewin GR. Voltage gating of mechanosensitive PIEZO channels. *Nat Commun*. 2018;9(1):1096. [PubMed: 29545531]
- Nonomura K, Woo SH, Chang RB, Gillich A, Qiu Z, Francisco AG, et al. Piezo2 senses airway stretch and mediates lung inflation-induced apnoea. *Nature*. 2017;541(7636):176–81. [PubMed: 28002412]
- Paulsen CE, Armache JP, Gao Y, Cheng Y, Julius D. Structure of the TRPA1 ion channel suggests regulatory mechanisms. *Nature*. 2015; 525(7570):552–7.
- Payandeh J, Scheuer T, Zheng N, Catterall WA. The crystal structure of a voltage-gated sodium channel. *Nature*. 2011;475(7356): 353–8. [PubMed: 21743477]
- Pearson WR. Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms. *Genomics*. 1991; 11(3):635–50. [PubMed: 1774068]
- Pearson WR. Empirical statistical estimates for sequence similarity searches. *J Mol Biol*. 1998; 276(1):71–84. [PubMed: 9514730]
- Ranade SS, Qiu Z, Woo SH, Hur SS, Murthy SE, Cahalan SM, et al. Piezo1, a mechanically activated ion channel, is required for vascular development in mice. *Proc Natl Acad Sci U S A*. 2014a;111(28):10347–52. [PubMed: 24958852]

- Ranade SS, Woo SH, Dubin AE, Moshourab RA, Wetzel C, Petrus M, et al. Piezo2 is the major transducer of mechanical forces for touch sensation in mice. *Nature*. 2014b;516(7529): 121–5. [PubMed: 25471886]
- Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Res*. 2019;47(W1):W5–W10. [PubMed: 31062021]
- Saier MH, Reddy VS, Moreno-Hagelsieb G, Hendargo KJ, Zhang Y, Iddamsetty V, et al. The transporter classification database (TCDB): 2021 update. *Nucleic Acids Res*. 2021; 49(D1):D461–7. [PubMed: 33170213]
- Saotome K, Murthy SE, Kefauver JM, Whitwam T, Patapoutian A, Ward AB. Structure of the mechanically activated ion channel Piezo1. *Nature*. 2018;554(7693):481–6. [PubMed: 29261642]
- Shah V, Patel S, Shah J. Emerging role of Piezo ion channels in cardiovascular development. *Dev Dyn*. 2022;251(2):276–86. [PubMed: 34255896]
- Steinegger M, Meier M, Mirdita M, Vohringer H, Haunsberger SJ, Soding J. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics*. 2019; 20(1):473. [PubMed: 31521110]
- Tusnady GE, Simon I. Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J Mol Biol*. 1998;283(2):489–506. [PubMed: 9769220]
- Tusnady GE, Simon I. The HMMTOP transmembrane topology prediction server. *Bioinformatics*. 2001;17(9):849–50. [PubMed: 11590105]
- Volkers L, Mechoukhi Y, Coste B. Piezo channels: from structure to function. *Pflugers Arch*. 2015;467(1):95–9. [PubMed: 25037583]
- Wang J, Jiang J, Yang X, Zhou G, Wang L, Xiao B. Tethering Piezo channels to the actin cytoskeleton for mechanogating via the cadherin-beta-catenin mechanotransduction complex. *Cell Rep*. 2022;38(6):110342.
- Wang L, Zhou H, Zhang M, Liu W, Deng T, Zhao Q, et al. Structure and mechanogating of the mammalian tactile channel PIEZO2. *Nature*. 2019;573(7773):225–9. [PubMed: 31435011]
- Wong WC, Maurer-Stroh S, Eisenhaber F. More than 1,001 problems with protein domain databases: transmembrane regions, signal peptides and the issue of sequence homology. *PLoS Comput Biol*. 2010;6(7):e1000867.
- Wong WC, Maurer-Stroh S, Eisenhaber F. Not all transmembrane helices are born equal: towards the extension of the sequence homology concept to membrane proteins. *Biol Direct*. 2011;6:57. [PubMed: 22024092]
- Wong WC, Maurer-Stroh S, Schneider G, Eisenhaber F. Transmembrane helix: simple or complex. *Nucleic Acids Res*. 2012;40(Web Server issue):W370–5. [PubMed: 22564899]
- Woo SH, Lukacs V, de Nooij JC, Zaytseva D, Criddle CR, Francisco A, et al. Piezo2 is the principal mechanotransduction channel for proprioception. *Nat Neurosci*. 2015;18(12): 1756–62. [PubMed: 26551544]
- Woo SH, Ranade S, Weyer AD, Dubin AE, Baba Y, Qiu Z, et al. Piezo2 is required for Merkel-cell mechanotransduction. *Nature*. 2014; 509(7502):622–6. [PubMed: 24717433]
- Wu J, Lewis AH, Grandl J. Touch, tension, and transduction - the function and regulation of piezo ion channels. *Trends Biochem Sci*. 2017;42(1):57–71. [PubMed: 27743844]
- Yarishkin O, Phuong TTT, Baumann JM, De Ieso ML, Vazquez-Chona F, Rudzitis CN, et al. Piezo1 channels mediate trabecular meshwork mechanotransduction and promote aqueous fluid outflow. *J Physiol*. 2021; 599(2):571–92. [PubMed: 33226641]
- Zhai Y, Saier MH Jr. A web-based program for the prediction of average hydrophobicity, average amphipathicity and average similarity of multiply aligned homologous proteins. *J Mol Microbiol Biotechnol*. 2001;3(2):285–6. [PubMed: 11321584]
- Zhang X, Ren W, DeCaen P, Yan C, Tao X, Tang L, et al. Crystal structure of an orthologue of the NaChBac voltage-gated sodium channel. *Nature*. 2012;486(7401):130–4. [PubMed: 22678295]
- Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res*. 2005;33(7):2302–9. [PubMed: 15849316]
- Zhao Q, Wu K, Geng J, Chi S, Wang Y, Zhi P, et al. Ion permeation and mechanotransduction mechanisms of mechanosensitive piezo channels. *Neuron*. 2016;89(6):1248–63. [PubMed: 26924440]

Zhao Q, Zhou H, Chi S, Wang Y, Wang J, Geng J, et al. Structure and mechanogating mechanism of the Piezo1 channel. *Nature*. 2018; 554(7693):487–92. [PubMed: 29469092]

Zhao Q, Zhou H, Li X, Xiao B. The mechano-sensitive Piezo1 channel: a three-bladed propeller-like structure and a lever-like mechanogating mechanism. *FEBS J*. 2019; 286(13):2461–70. [PubMed: 30500111]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

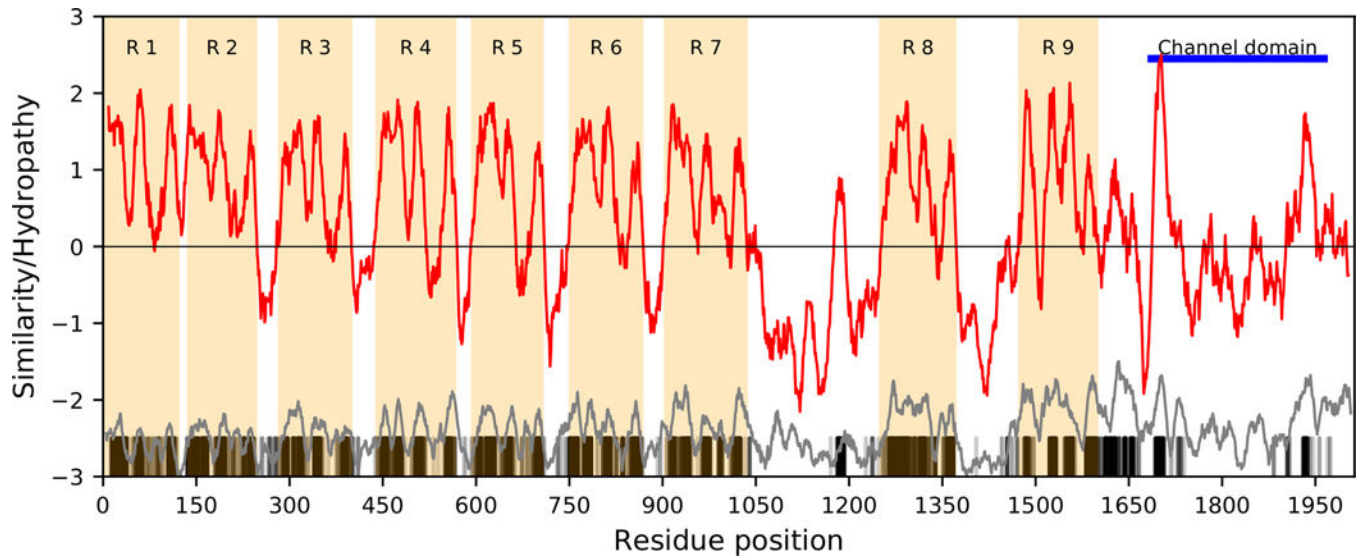


Fig. 1.

Average hydropathy profile for members of the Piezo family. The red curve indicates average hydropathy, the gray curve indicates average similarity, and the vertical thin black bars on the X-axis indicate residues predicted to be parts of TMSs using HMMTOP [Tusnady and Simon, 2001]. Each repeat is highlighted with tan-colored boxes. For clarity, the 4 TMSs within each repeat are not highlighted. For the visualization of TMSs relative to their sequence logos and hydropathy curves, see online suppl. Figures S1, S2. The figure was generated with the AveHAS program [Zhai and Saier, 2001] as described in Methods.

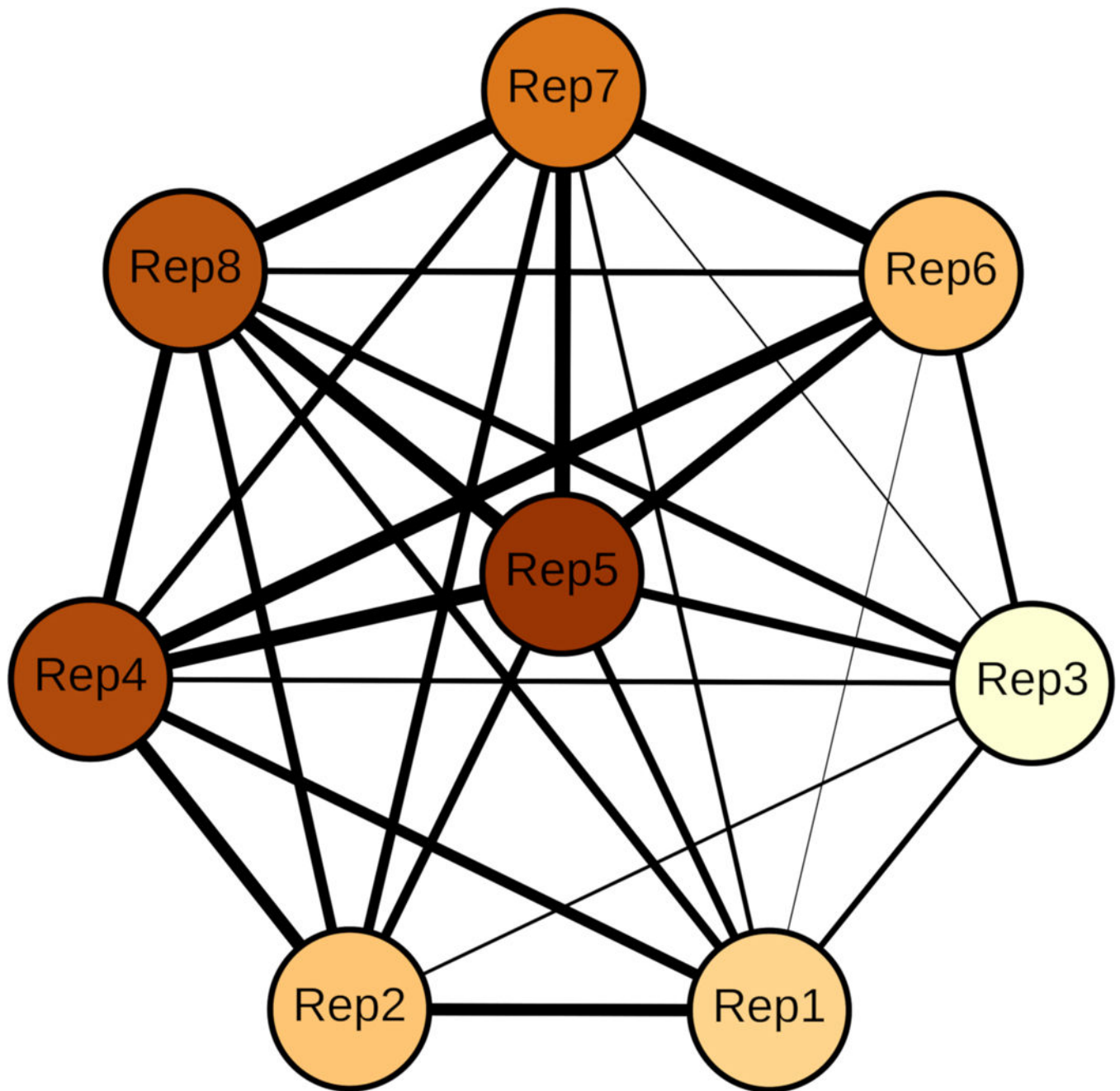


Fig. 2. Network of connections among repeats in the Piezo family. Nodes represent each repeat. The confidence level of the relationship between pairs of repeats (see Methods) is expressed with the thickness levels of the edges connecting pairs of nodes, where thicker lines correspond to higher levels of confidence. The darker the color of a node, the more significant the connections to other nodes. The length of the edges connecting nodes is not meaningful (see Methods).

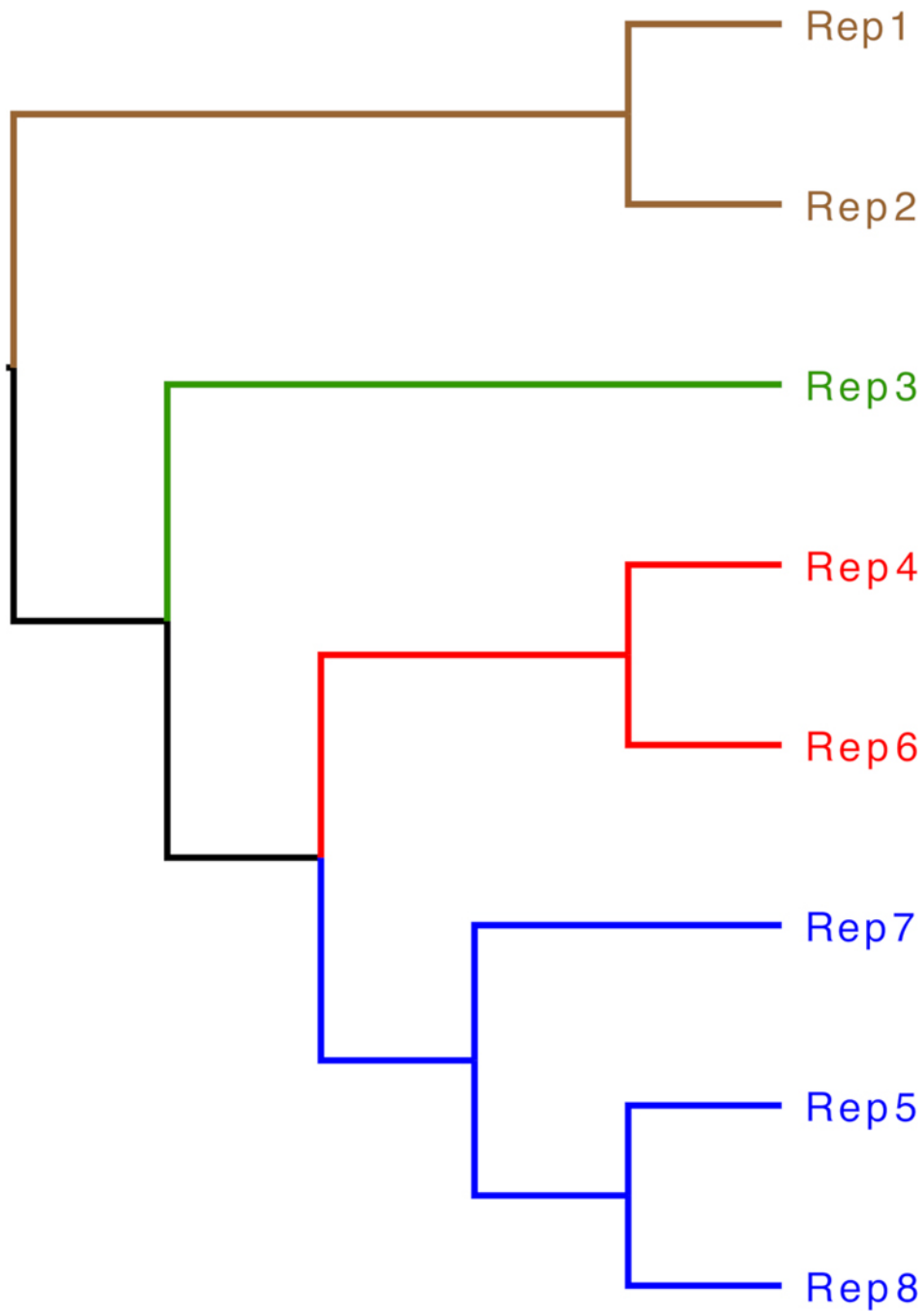
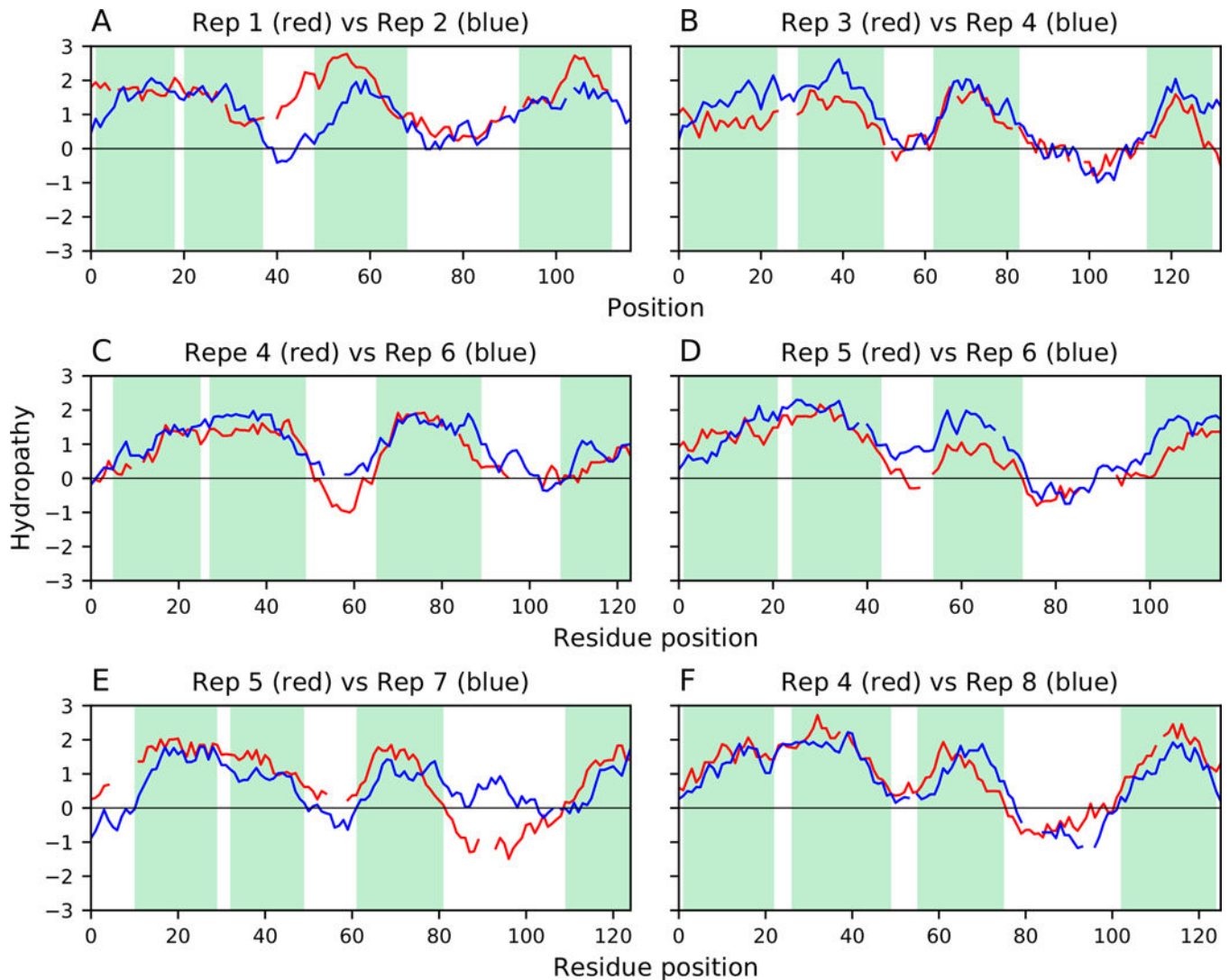


Fig. 3. Rectangular dendrogram of HMM profile-profile similarities among Piezo repeats. The dendrogram was constructed based on coverage-weighted HHalign scores of pairwise HMM alignments using the Ward algorithm (agglomerative coefficient: 0.75; see Methods). The colors of the labels indicate the four groups in which the repeats are organized according to their similarities. Coverage-weighted HHalign scores are not necessarily proxies of phylogenetic distances. Thus, only the topology of the dendrogram was considered meaningful, and the scale bar was disregarded.

**Fig. 4.**

Pairwise sequence alignments between Piezo repeats. The figure shows hydropathy curves for six examples of pairwise sequence alignments between different repeats of Piezo homologs. TMSs are highlighted with green bars. The Smith-Waterman (S-W) E-value is provided as determined by ssearch36 [Pearson, 1991] for each pair of repeats. Two additional scores are provided for each alignment, where we used 10,000 shuffles to calculate the p value based on the extreme value distribution (EVD) of raw S-W scores and the corresponding upper limit (max) p value, with a 0.999 degree of confidence, based on coverage-weighted S-W scores (see Methods). **a** Repeat 1 of RQP00136 (red) is aligned to repeat 2 of XP_013404042 (blue) with alignment scores: S-W E-value: 2.3×10^{-04} , EVD p value: 3.8×10^{-06} (online suppl. Fig. S5A), and max p value: 6.90×10^{-04} . **b** Repeat 3 of XP_013104300 (red) versus repeat 4 of XP_013766249 (blue) with alignment scores: S-W E-value: 5.6×10^{-06} , EVD p value = 4.66×10^{-06} (online suppl. Fig. S6A), and max p value: 6.90×10^{-04} (online suppl. Fig. S6). **c** Repeat 4 of PWA88896 (red) versus repeat 6 of PAA91575 (blue) with alignment scores: S-W E-value: 6.5×10^{-05} , EVD p value: $1.87 \times$

10^{-05} , and max p value: 6.90×10^{-04} . **d** Repeat 5 of XP_024369685 (red) versus repeat 6 of XP_027097338 (blue) with alignment scores: S-W E-value: 5.7×10^{-06} , EVD p value: 4.44×10^{-06} , and max p value: 6.90×10^{-04} . **e** Repeat 5 of KFB42311 (red) versus repeat 7 of XP_024369685 (blue) with alignment scores: S-W E-value: 1.6×10^{-07} , EVD p value: 4.18×10^{-07} , and max p value: 6.90×10^{-04} . **f** Repeat 4 of OWK56671 (red) versus repeat 8 of XP_026690886 (blue) with alignment scores: S-W E-value: 1.5×10^{-05} , EVD p value: 1.78×10^{-07} , and max p value: 6.90×10^{-04} . Both EVD and max p values support the reliability of the alignments. The max p values are all the same in panels (**a–f**) because in every case, the reference score was better than all 10,000 scores derived from shuffled alignments. In panels (**b–f**), the reference score adjusted by coverage is so distant from the distribution of shuffled scores, that the Gaussian Kernel Density function returns a p value of zero due to the lack of data points (shuffled scores) near the reference score (online suppl. Fig. S5B, S6B). In these cases, we infer that the maximal p value $< 6.90 \times 10^{-04}$ with 0.999 confidence (see Methods).

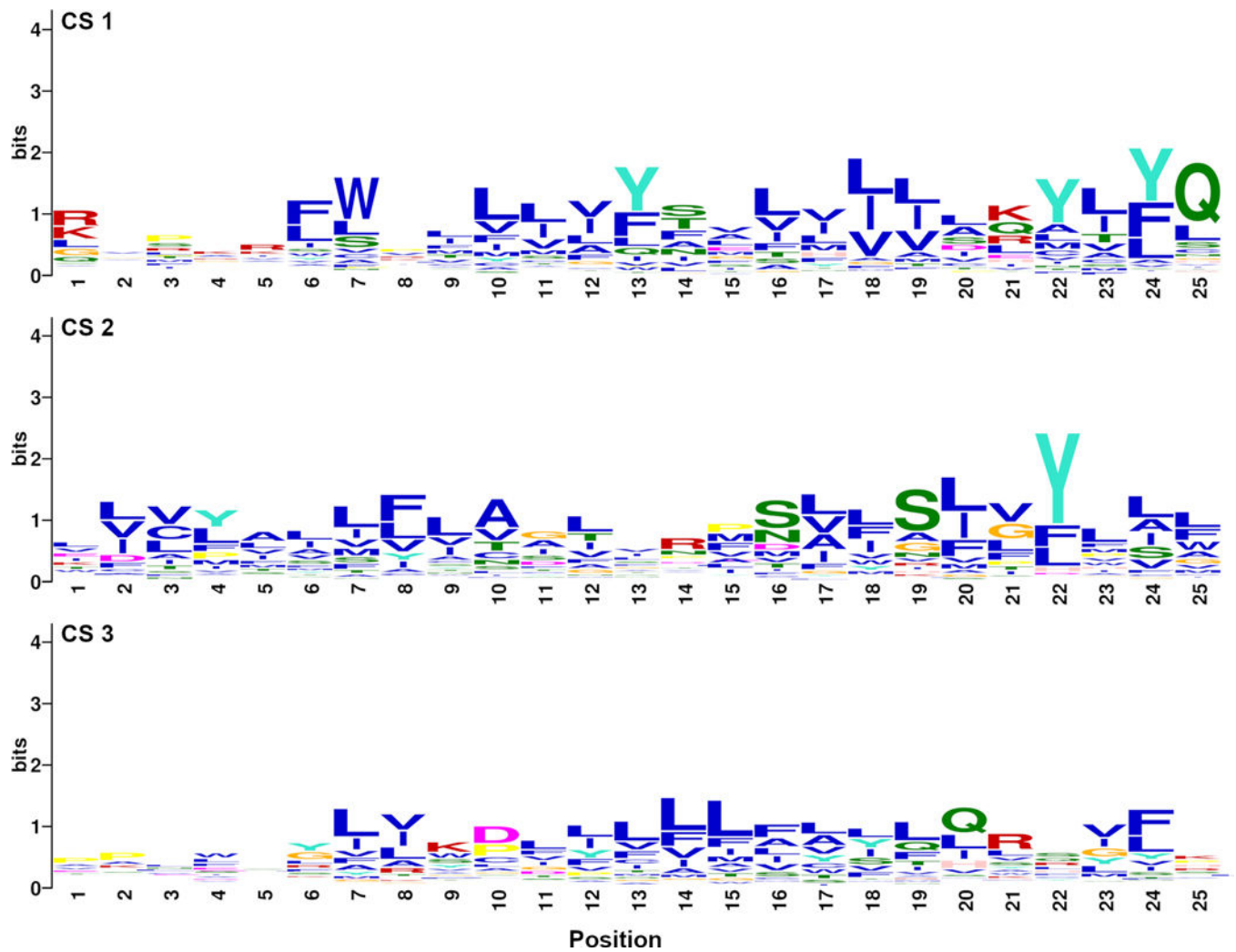


Fig. 5. Sequence logos of conserved signature sequences (CSs). Three 25 aa long CSs were identified with MEME E-value $<10^{-10}$ (see Methods). **a** CS 1 (E-value: $1.6 \times 10^{-1,143}$) matches TMS 3. **b** CS 2 (E-value: 3.3×10^{-863}) matches the central region between TMSs 1–2. **c** CS 3 (E-value: 1.8×10^{-461}) matches TMS 4. Figure S7 illustrates where the CSs match relative to the hydrophathy profile of two representative Piezo members with solved structures: E2JF22 (TC: 1.A.75.1.14; PDB: 5Z10) and Q8CD54 (TC: 1.A.75.1.2; PDB: 6KG7). See text for discussion.

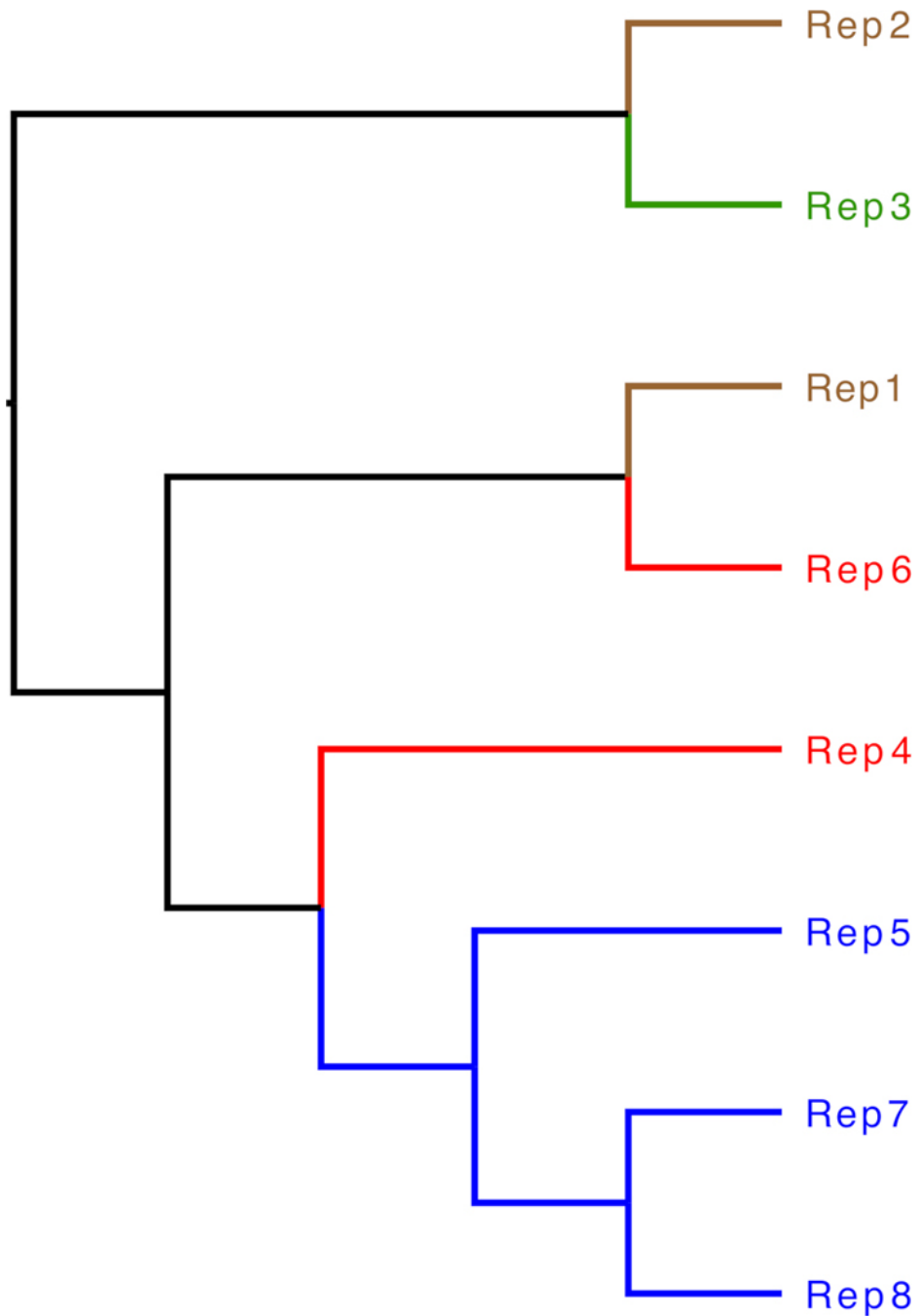


Fig. 6. Structural similarities among Piezo repeats. Similarities are measured with the RMSD values of superpositions between pairs of repeats. The rectangular dendrogram was constructed with the Ward algorithm and has weak clustering structure (agglomerative coefficient: 0.34; see Methods). To facilitate comparisons, the coloring format of branches and labels is the same as in Figure 3. RMSD scores are not necessarily proxies of phylogenetic distances.

Thus, only the topology of the dendrogram is considered meaningful, and the scale bar was disregarded. See text for discussion.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

Pairwise comparison of HMM-profiles for all repeats.

Repeat A	Repeat B	E-value	Prob	MaxCov	WScore
4	6	6.9×10^{-08}	94.6	0.95	89.9
5	8	6×10^{-10}	96.6	0.93	89.8
7	8	1.3×10^{-08}	95.6	0.92	88.0
6	7	1.9×10^{-07}	93.7	0.92	86.2
4	5	6.5×10^{-07}	92.0	0.94	86.5
5	7	8.8×10^{-07}	91.5	0.94	86.0
2	4	2.4×10^{-07}	93.4	0.81	75.7
5	6	2.7×10^{-06}	88.6	0.85	75.3
4	8	6.7×10^{-07}	92.0	0.79	72.7
1	2	4.4×10^{-08}	94.9	0.73	69.3
1	4	2.1×10^{-05}	76.7	0.85	65.2
2	8	1.2×10^{-06}	90.9	0.68	61.8
2	7	1.6×10^{-06}	90.0	0.68	61.2
3	8	1.1×10^{-05}	81.6	0.75	61.2
3	5	1.1×10^{-05}	81.5	0.72	58.7
2	5	4.6×10^{-07}	92.6	0.62	57.4
4	7	4.1×10^{-07}	92.8	0.61	56.6
1	8	1.4×10^{-05}	80.0	0.70	56.0
1	5	1.1×10^{-05}	81.5	0.65	53.0
3	6	3.4×10^{-06}	87.8	0.54	47.4
6	8	1.3×10^{-04}	53.2	0.79	42.0
1	3	4.3×10^{-05}	69.3	0.60	41.6
3	4	5.3×10^{-07}	92.4	0.40	37.0
1	7	1.2×10^{-04}	55.0	0.66	36.3
2	3	2.8×10^{-04}	41.1	0.69	28.4
3	7	3.7×10^{-04}	36.2	0.61	22.1
1	6	7.8×10^{-04}	24.9	0.74	18.4

Columns 1–2 present the pairs of repeats being compared, columns 3–4 show the E-values and probabilities of the alignments, respectively, as reported by hhalgn [Steinegger et al., 2019] (see Methods). Column 5 is the coverage of the alignment relative to the shorter HMM. Column 6 presents a weighted score (WScore) or confidence level where the probability of the alignment reported by hhalgn is multiplied by the coverage. Only pairs of repeats with hhalgn E-value $< 10^{-3}$ are tabulated; no alignments with repeat 9 passed this filter.

Table 2.

Conserved collocated and similarly oriented residue classes in Piezo1 and Piezo2 structures.

Helix	Residue Class	Repeats	Notes
1	Cys	4, 6, 8	Intrarepeat, points to helix 4.
2	Tyr	4, 6, 7	Intrarepeat, points to helix 3.
2	Cys	4, 6, 7*	Lipid-exposed, orientation difference in repeat 7.
2	Trp	6, 8, 5*	Intrarepeat to helix 3, instance in repeat 5 is off by approximately 90 degrees.
3	Trp	4, 6, 7, 8	Intrarepeat, points to core.
4	Ser	6, 7, 8	Inter-repeat, buried, points to previous helix 2, positions differ but orientations are comparable.
4	Tyr	5, 9, 4*	Inter-repeat, points to previous repeat or solvent exposed.
4	Asp	6, 7, 8	Intrarepeat, buried, points to core.
4	Cys	4, 6, 7, 9	Scattered orientations but remarkable proximity (± 1 turn).

The table presents 4 columns: (1) the helix number within repeats; (2) the residue classes in 3-letter code present in Piezo1 (5z10) and Piezo2 (6kg7) on the same TMS and having similar locations/orientations in at least three repeats; (3) specific repeats where residue classes are conserved; and (4) an explanatory note.

* Instances in this repeat are marginally different in position/orientation.