# Science Advances

## Supplementary Materials for

## Mechanical gating of the auditory transduction channel TMC1 involves the fourth and sixth transmembrane helices

Nurunisa Akyuz et al.

Corresponding author: Nurunisa Akyuz, nneyzi@gmail.com

*Sci. Adv.* **8**, eabo1126 (2022) DOI: 10.1126/sciadv.abo1126

The PDF file includes:

Figs. S1 to S6 Tables S1 to S4 Legend for movie S1 References

#### Other Supplementary Material for this manuscript includes the following:

Movie S1



**Supplementary Fig. 1: Topology of TMC1. A)** Schematic representation of the 10-TM topology of TMC1. Parts of the N-terminus (residues 100-130 in mouse TMC1) as well as the intracellular loop between TM helices 2-3 (residues 300-350) have been implicated in binding to the auxiliary channel subunit CIB2 <sup>20,46</sup>. **B)** Schematic for the predicted arrangement of the helices in the dimeric structural models. Models based on X-ray and cryo-EM structures of TMEM16 proteins<sup>5, 3</sup> suggest that the subunit interface is near TM10 and may be at least partly filled with lipid or perhaps by auxiliary proteins. A top to down view of the dimeric channel is shown with each TM helix represented as a circle and colored as in panel A. Subunit B shows TM helix numbers. The approximate position of the pore cavity is marked with a **+** sign in each subunit.



Supplementary Fig. 2 Confidence metrics for the AlphaFold2 prediction. A) AlphaFold2-generated per-residue confidence scores (pIDDT) between 0 and 100 are plotted as a function of residue number. TM domains show pIDDT values >80. B) The expected position error map (PAE) from AlphaFold2 also indicates uncertainty in the N-terminal and loop structures but confidence in the fold of the transmembrane domains C) Overlay of the TM2-3 loop (residues 300-350) from the published crystal structure (in cyan, PDB id: 6WUD) with the predicted fold of this domain (in magenta).



4 and TM6

TM4

5

Supplementary Fig. 3. Open-like and closed-like predicted structural models. A) The permeation pathway model generated with the HOLE program for the predicted open-like (pink) and closed-like (blue) structures. The estimated pore radius is plotted against the pore length Z, from outside to inside. The approximate position of the constriction site in the open-like structure, about 10 Å in, corresponds to approximately N404 on TM4 and L524 on TM6. Loops and termini (including residues 1-174, 212-266, 294-351, 466-511, 545-566, 617-625, 655-696, 725-756) are not shown for simplicity. B) A thin cross section of the protein parallel to the membrane at the Z-level of approximately N404 on TM4 and L524 on TM6 for the open-like and closed-like structural predictions (pink and blue, respectively). C) Contact maps for residues 165-680 in TMC1 predicted by EVcoupling analysis (black dots) are mapped onto contact maps from open-like and closed-like predicted structures (pink and blue, respectively). Evolutionarily coupled (EC) pairs that are found in proximity only in the closed state are indicated to the right: they connect TM4 to TM6 (notably S408-T531 and F413-Y533), and the TM1-2 loop to TM4 (Q245 on TM1-2 loop and M403 on TM4). The transmembrane domains corresponding to the residues along the axes are numbered 1-9 (TM10 not shown). The EVcoupling analysis suggests that these residue pairs are likely to interact; the AlphaFold2 structures also predict interaction but only in the closed-like conformation. EC 528-601 between TM6 and TM8, which is the second strongest coupling out of all possible analyzed pairs, is circled in red. D) nhTMEM16 transition from closed (blue) to intermediate (cyan) state involves TM4 moving away from TM5, while nhTMEM16 transition to the open state (wheat color) involves TM4 moving away from TM6. Predicted TMC1 states are shown for comparison.



**Supplementary Fig. 4. Effects of G539A/A544G mutation. A**) The average maximum transduction currents for hair cells virally expressing WT-TMC1, TMC1-G518A, TMC1-G539A, TMC1-G539I and TMC1-G539A/A544G. Amplitudes are all similar. **B**) Sequence alignment of the mouse TM6 helix in orthologs TMC1, 2 and 3. TMC1 and TMC2 have glycine at position 539; TMC3 does not but has a glycine at position 544. **C**) Representative currents from hair cells expressing TMC1-G539I/A544G. **D**) Normalized open probability curve for G539I/A544G compared to the WT control, G539A and G539I. **E**) Fitting parameters for the four TMC variants. None are significantly different (p =0.1, 0.4 and 0.8 for X<sub>0</sub>, and 0.7, 0.6 and 0.7 for Z, relative to WT).



**Supplementary Fig. 5. Location and interactions of D528 in the structural model. A)** Electrostatic surface charge distribution of pore-lining TM6 and TM7 helices, viewed from inside the pore. TM6 contacts the lipid, TM7 is within the protein core. WT-TMC1 and TMC1-D528N structures were generated with AlphaFold2. The D528N mutation substantially reduces the negative charge that might enhance passage of cations. B) Extracellular view of the selected residues D528 and R601. The approximate locations of the permeation pathway (+) and the lipid bilayer (black curve) are indicated. In this conformation, the side chain of D528 forms a salt bridge with the side chain of R601. **C)** The position of deafness mutations M412 and D569 are indicated. In the closed-like conformation, these sites would be inaccessible to extracellular solution (Supplementary Fig. 2A), consistent with their limited accessibility in the closed state as assessed with cysteine-modifying reagents<sup>3</sup>.



Supplementary Fig. 6. Average maximum transduction currents for hair cells virally expressing WT-TMC1, TMC1-D528A, TMC1-D528N, TMC1-D528E, TMC1-R601A and TMC1-D528N/R601F. There is variability among these variants in whole cell currents among. Variability is expected because most of the mutations are likely to alter the charge distribution within the permeation pathway.

| Mutation    | Titer (vg/ml) | Location |
|-------------|---------------|----------|
| G518A       | 1.19E+13      | TM6      |
| G411A       | 3.04E+14      | TM4      |
| G539A       | 7.65E+13      | TM6      |
| G539I/A543G | 8.64E+13      | TM6      |
| G539I       | 1.76E+14      | TM6      |
| D528A       | 4.83E+13      | TM6      |
| D528N       | 3.69E+13      | TM6      |
| D528E       | 1.48E+14      | TM6      |
| R601A       | 6.36E+13      | TM8      |
| R601H       | 4.50E+14      | TM8      |
| R601F/D528N | 2.17E+14      | TM8/6    |
| R523H       | 1.66E+14      | TM6      |
| WT-TMC1-HA  | 3.12E+13      | N/A      |

## Supplementary Table 1: TMC1 Mutants Packaged in AAV9-PHP.B

## Supplementary Table 2: Fitting Parameters. $X_{0}$ and Z values

|            | WT           | G411A        | G518A        | G539A        | G539I        |
|------------|--------------|--------------|--------------|--------------|--------------|
| <b>X</b> 0 | 374±33       | 507±28       | 493±19       | 518±30       | 478±25       |
| Z          | 0.027±0.0027 | 0.017±0.0014 | 0.021±0.0023 | 0.024±0.0023 | 0.024±0.0020 |

|            | WT           | D528A        | D528N        | D528E        | R601A        | R601F/D528N  |
|------------|--------------|--------------|--------------|--------------|--------------|--------------|
| <b>X</b> 0 | 374±33       | 581±73       | 340±25       | 510±29       | 580±25       | 510±46       |
| Z          | 0.027±0.0027 | 0.017±0.0022 | 0.016±0.0018 | 0.020±0.0011 | 0.025±0.0030 | 0.014±0.0015 |

#### Supplementary Table 3: Occurrence of the indicated amino acids at the positions

**corresponding to 518, 539 and 544.** The amino acids found in the mouse TMC1 sequence are underlined. The frequencies represent those occurring in a multiple sequence alignment of 5,050 TMC sequences (see Methods).

|     | G518       | G539       | A544       |
|-----|------------|------------|------------|
| А   |            |            | <u>27%</u> |
| G   | <u>98%</u> | <u>50%</u> | 11%        |
| R   |            |            | 34%        |
| K   |            |            | 19%        |
| I/V |            | 41%        |            |

Supplementary Table 4: Frequency of finding the indicated amino acids at the positions corresponding to 523/528/601 within the mouse TMC1 sequence. The frequencies represent those occurring in a multiple sequence alignment of 5,050 TMC sequences (see Methods).

| 523/528/601 |            |
|-------------|------------|
| K/D/K       | 30%        |
| R/D/K       | 20%        |
| R/D/R       | <u>17%</u> |
| K/D/R       | 7%         |
| K/N/F       | 6%         |
| other       | 20%        |

**Supplementary Video 1: A proposed gating trajectory**. The trajectory is based on the two groups of AlphaFold2 structures possibly representing closed and open states of TMC1.

#### **REFERENCES AND NOTES**

- D. P. Corey, A. J. Hudspeth, Kinetics of the receptor current in bullfrog saccular hair cells. *J. Neurosci.* 3, 962–976 (1983).
- B. Pan, G. S. Geleoc, Y. Asai, G. C. Horwitz, K. Kurima, K. Ishikawa, Y. Kawashima, A. J. Griffith, J. R. Holt, TMC1 and TMC2 are components of the mechanotransduction channel in hair cells of the mammalian inner ear. *Neuron* 79, 504–515 (2013).
- B. Pan, N. Akyuz, X. P. Liu, Y. Asai, C. Nist-Lund, K. Kurima, B. H. Derfler, B. Gyorgy, W. Limapichat, S. Walujkar, L. N. Wimalasena, M. Sotomayor, D. P. Corey, J. R. Holt, TMC1 forms the pore of mechanosensory transduction channels in vertebrate inner ear hair cells. *Neuron* 99, 736–753.e6 (2018).
- 4. A. Medrano-Soto, G. Moreno-Hagelsieb, D. McLaughlin, Z. S. Ye, K. J. Hendargo, M. H. Saier Jr., Bioinformatic characterization of the anoctamin superfamily of Ca<sup>2+</sup>-activated ion channels and lipid scramblases. *PLOS ONE* **13**, e0192851 (2018).
- 5. A. Ballesteros, C. Fenollar-Ferrer, K. J. Swartz, Structural relationship between the putative hair cell mechanotransduction channel TMC1 and TMEM16 proteins. *eLife* **7**, e38433 (2018).
- 6. Y. Hahn, D. S. Kim, I. H. Pastan, B. Lee, Anoctamin and transmembrane channel-like proteins are evolutionarily related. *Int. J. Mol. Med.* **24**, 51–55 (2009).
- K. Kunzelmann, I. Cabrita, P. Wanitchakool, J. Ousingsawat, L. Sirianant, R. Benedetto, R. Schreiber, Modulating Ca<sup>2+</sup> signals: A common theme for TMEM16, Ist2, and TMC. *Pflugers Arch.* 468, 475– 490 (2016).
- S. Walujkar, J. M. Lotthammer, C. R. Nisler, J. C. Sudar, A. Ballesteros, M. Sotomayor, In silico electrophysiology of inner-ear mechanotransduction channel TMC1 models. bioRxiv 2021.09.17.460860 [Preprint]. 18 September 2021. https://doi.org/10.1101/2021.09.17.460860.
- Y. Jia, Y. Zhao, T. Kusakizako, Y. Wang, C. Pan, Y. Zhang, O. Nureki, M. Hattori, Z. Yan, TMC1 and TMC2 proteins are pore-forming subunits of mechanosensitive ion channels. *Neuron* 105, 310–321.e3 (2020).
- 10. C. Paulino, V. Kalienkova, A. K. M. Lam, Y. Neldner, R. Dutzler, Activation mechanism of the calcium-activated chloride channel TMEM16A revealed by cryo-EM. *Nature* **552**, 421–425 (2017).
- 11. M. E. Falzone, J. Rheinberger, B. C. Lee, T. Peyear, L. Sasset, A. M. Raczkowski, E. T. Eng, A. Di Lorenzo, O. S. Andersen, C. M. Nimigean, A. Accardi, Structural basis of Ca<sup>2+</sup>-dependent activation and lipid transport by a TMEM16 scramblase. *eLife* 8, e43229 (2019).
- 12. A. K. M. Lam, R. Dutzler, Mechanism of pore opening in the calcium-activated chloride channel TMEM16A. *Nat. Commun.* **12**, 786 (2021).
- G. Khelashvili, M. E. Falzone, X. Cheng, B. C. Lee, A. Accardi, H. Weinstein, Dynamic modulation of the lipid translocation groove generates a conductive ion channel in Ca<sup>2+</sup>-bound nhTMEM16. *Nat. Commun.* 10, 4972 (2019).
- 14. A. A. Indzhykulian, S. L. Johnson, G. S. Géléoc, Electrophysiological recordings of voltage-dependent and mechanosensitive currents in sensory hair cells of the auditory and vestibular organs of the mouse, in *Developmental*, *Physiological*, *and Functional Neurobiology of the Inner Ear* (Humana, 2022).

- 15. H. Yang, S. F. Baker, M. E. Gonzalez, D. J. Topham, L. Martinez-Sobrido, M. Zand, J. Holden-Wiltse, H. Wu, An improved method for estimating antibody titers in microneutralization assay using green fluorescent protein. *J. Biopharm. Stat.* 26, 409–420 (2016).
- M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch, R. D. Schaeffer, C. Millan, H. Park, C. Adams, C. R. Glassman, A. DeGiovanni, J. H. Pereira, A. V. Rodrigues, A. A. van Dijk, A. C. Ebrecht, D. J. Opperman, T. Sagmeister, C. Buhlheller, T. Pavkov-Keller, M. K. Rathinaswamy, U. Dalwadi, C. K. Yip, J. E. Burke, K. C. Garcia, N. V. Grishin, P. D. Adams, R. J. Read, D. Baker, Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, **373**, 871–879 (2021).
- 17. C. A. Rohl, C. E. Strauss, K. M. Misura, D. Baker, Protein structure prediction using Rosetta. *Methods Enzymol.* **383**, 66–93 (2004).
- 18. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
- 19. J. F. Moult, A. Kryshtafovych, T. Schwede, M. Topf, vol. CASP 14 Abstract Book.
- 20. X. Liang, X. Qiu, G. Dionne, C. L. Cunningham, M. L. Pucak, G. Peng, Y. H. Kim, A. Lauer, L. Shapiro, U. Muller, CIB2 and CIB3 are auxiliary subunits of the mechanotransduction channel of hair cells. *Neuron* 109, 2131–2149.e15 (2021).
- 21. D. P. Corey, A. J. Hudspeth, Ionic basis of the receptor potential in a vertebrate hair cell. *Nature* **281**, 675–677 (1979).
- 22. C. J. Kros, A. Rusch, G. P. Richardson, Mechano-electrical transducer currents in hair cells of the cultured neonatal mouse cochlea. *Proc. Biol. Sci.* **249**, 185–193 (1992).
- 23. A. J. Ricci, R. Fettiplace, Calcium permeation of the turtle hair cell mechanotransducer channel and its relation to the composition of endolymph. *J. Physiol.* **506**, 159–173 (1998).
- 24. M. Beurg, L. A. Schimmenti, A. Koleilat, S. S. Amr, A. Oza, A. J. Barlow, A. Ballesteros, R. Fettiplace, New Tmc1 deafness mutations impact mechanotransduction in auditory hair cells. *J. Neurosci.* 41, 4378–4391 (2021).
- 25. M. Beurg, A. Barlow, D. N. Furness, R. Fettiplace, A *Tmc1* mutation reduces calcium permeability and expression of mechanoelectrical transduction channels in cochlear hair cells. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 20743–20749 (2019).
- 26. O. S. Smart, J. G. Neduvelil, X. Wang, B. A. Wallace, M. S. Sansom, HOLE: A program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* **14**, 354–360, (1996).
- 27. V. Kalienkova, V. Clerico Mosina, L. Bryner, G. T. Oostergetel, R. Dutzler, C. Paulino, Stepwise activation mechanism of the scramblase nhTMEM16 revealed by cryo-EM. *eLife* **8**, e44364 (2019).
- 28. H. E. Farris, C. L. LeBlanc, J. Goswami, A. J. Ricci, Probing the pore of the auditory hair cell mechanotransducer channel in turtle. *J. Physiol.* **558**, 769–792 (2004).

- 29. J. E. Chen, C. C. Huang, T. E. Ferrin, RRDistMaps: A UCSF Chimera tool for viewing and comparing protein distance maps. *Bioinformatics* **31**, 1484–1486 (2015).
- T. A. Hopf, C. P. Scharfe, J. P. Rodrigues, A. G. Green, O. Kohlbacher, C. Sander, A. M. Bonvin, D. S. Marks, Sequence co-evolution gives 3D contacts and structures of protein complexes. *eLife* 3, e03430 (2014).
- 31. C. J. Peters, J. M. Gilchrist, J. Tien, N. P. Bethel, L. Qi, T. Chen, L. Wang, Y. N. Jan, M. Grabe, L. Y. Jan, The sixth transmembrane segment Is a major gating component of the TMEM16A calcium-activated chloride channel. *Neuron* 97, 1063–1077.e4 (2018).
- 32. C. Paulino, Y. Neldner, A. K. Lam, V. Kalienkova, J. D. Brunner, S. Schenck, R. Dutzler, Structural basis for anion conduction in the calcium-activated chloride channel TMEM16A. *eLife* **6**, e26232 (2017).
- 33. Y. Kawashima, G. S. Geleoc, K. Kurima, V. Labay, A. Lelli, Y. Asai, T. Makishima, D. K. Wu, C. C. Della Santina, J. R. Holt, A. J. Griffith, Mechanotransduction in mouse inner ear hair cells requires transmembrane channel-like genes. *J. Clin. Invest.* **121**, 4796–4809 (2011).
- 34. C. Askew, C. Rochat, B. Pan, Y. Asai, H. Ahmed, E. Child, B. L. Schneider, P. Aebischer, J. R. Holt, Tmc gene therapy restores auditory function in deaf mice. *Sci. Transl. Med.* **7**, 295ra108 (2015).
- J. Lee, C. Nist-Lund, P. Solanes, H. Goldberg, J. Wu, B. Pan, B. L. Schneider, J. R. Holt, Efficient viral transduction in mouse inner ear hair cells with utricle injection and AAV9-PHP.B. *Hear. Res.* 394, 107882 (2020).
- 36. M. V. Ivanchenko, K. S. Hanlon, M. K. Devine, K. Tenneson, F. Emond, J. F. Lafond, M. A. Kenna, D. P. Corey, C. A. Maguire, Preclinical testing of AAV9-PHP.B for transgene expression in the non-human primate cochlea. *Hear. Res.* **394**, 107930 (2020).
- 37. A. J. Hudspeth, Y. Choe, A. D. Mehta, P. Martin, Putting ion channels to work: mechanoelectrical transduction, adaptation, and amplification by hair cells. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11765–11772 (2000).
- 38. F. J. Sigworth, The variance of sodium current fluctuations at the node of Ranvier. *J. Physiol.* **307**, 97–129 (1980).
- 39. T. Holton, A. J. Hudspeth, The transduction channel of hair cells from the bull-frog characterized by noise analysis. *J. Physiol.* **375**, 195–227 (1986).
- 40. S. Ding, L. Ingleby, C. A. Ahern, R. Horn, Investigating the putative glycine hinge in Shaker potassium channel. *J. Gen. Physiol.* **126**, 213–226 (2005).
- 41. I. Luque, O. L. Mayorga, E. Freire, Structure-based thermodynamic scale of alpha-helix propensities in amino acids. *Biochemistry* **35**, 13681–13688 (1996).
- 42. M. Beurg, J. H. Nam, R. Fettiplace, The speed of the hair cell mechanotransducer channel revealed by fluctuation analysis. *J. Gen. Physiol.* **153**, e202112959 (2021).
- 43. L. F. Corns, S. L. Johnson, C. J. Kros, W. Marcotti, Tmc1 point mutation affects Ca<sup>2+</sup> sensitivity and block by dihydrostreptomycin of the mechanoelectrical transducer current of mouse outer hair cells. *J. Neurosci.* 36, 336–349 (2016).

- 44. S. Jojoa-Cruz, K. Saotome, S. E. Murthy, C. C. A. Tsui, M. S. Sansom, A. Patapoutian, A. B. Ward, Cryo-EM structure of the mechanically activated ion channel OSCA1.2. *eLife* **7**, e41845 (2018).
- 45. B. Zhao, Z. Wu, N. Grillet, L. Yan, W. Xiong, S. Harkins-Perry, U. Muller, TMIE is an essential component of the mechanotransduction machinery of cochlear hair cells. *Neuron* **84**, 954–967 (2014).
- 46. A. P. J. Giese, Y. Q. Tang, G. P. Sinha, M. R. Bowl, A. C. Goldring, A. Parker, M. J. Freeman, S. D. M. Brown, S. Riazuddin, R. Fettiplace, W. R. Schafer, G. I. Frolenkov, Z. M. Ahmed, CIB2 interacts with TMC1 and TMC2 and is essential for mechanotransduction in auditory hair cells. *Nat. Commun.* 8, 43 (2017).
- 47. M. Beurg, W. Xiong, B. Zhao, U. Muller, R. Fettiplace, Subunit determination of the conductance of hair-cell mechanotransducer channels. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 1589–1594 (2015).
- 48. A. J. Hudspeth, How the ear's works work. Nature 341, 397-404 (1989).
- 49. M. Beurg, A. C. Goldring, R. Fettiplace, The effects of Tmc1 Beethoven mutation on mechanotransducer channel function in cochlear hair cells. *J. Gen. Physiol.* **146**, 233–243 (2015).
- 50. W. Marcotti, S. M. van Netten, C. J. Kros, The aminoglycoside antibiotic dihydrostreptomycin rapidly enters mouse outer hair cells through the mechano-electrical transducer channels. *J. Physiol.* **567**, 505–521 (2005).
- 51. V. S. Markin, A. J. Hudspeth, Gating-spring models of mechanoelectrical transduction by hair cells of the internal ear. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 59–83 (1995).
- 52. J. H. Nam, A. W. Peng, A. J. Ricci, Underestimated sensitivity of mammalian cochlear hair cells due to splay between stereociliary columns. *Biophys. J.* **108**, 2633–2647 (2015).
- 53. K. D. Karavitaki, D. P. Corey, Sliding adhesion confers coherent motion to hair cell stereocilia and parallel gating to transduction channels. *J. Neurosci.* **30**, 9051–9063 (2010).
- 54. J. Howard, A. J. Hudspeth, Compliance of the hair bundle associated with gating of mechanoelectrical transduction channels in the bullfrog's saccular hair cell. *Neuron* **1**, 189–199 (1988).
- 55. E. L. Cheung, D. P. Corey, Ca<sup>2+</sup> changes the force sensitivity of the hair-cell transduction channel. *Biophys. J.* **90**, 124–139 (2006).
- 56. A. P. Christensen, D. P. Corey, TRP channels in mechanosensation: Direct or indirect activation? *Nat. Rev. Neurosci.* **8**, 510–521 (2007).
- 57. P. Martin, D. Bozovic, Y. Choe, A. J. Hudspeth, Spontaneous oscillation by hair bundles of the bullfrog's sacculus. *J. Neurosci.* 23, 4533–4548 (2003).
- M. Mirdita, S. Ovchinnikov, M. Steinegger, ColabFold Making protein folding accessible to all. bioRxiv 2021.08.15.456425 [Preprint]. 15 August 2021. https://doi.org/10.1101/2021.08.15.456425.
- 59. D. S. Marks, L. J. Colwell, R. Sheridan, T. A. Hopf, A. Pagnani, R. Zecchina, C. Sander, Protein 3D structure computed from evolutionary sequence variation. *PLOS ONE* **6**, e28766 (2011).
- 60. L. S. Johnson, S. R. Eddy, E. Portugaly, Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11**, 431 (2010).

61. B. E. Suzek, Y. Wang, H. Huang, P. B. McGarvey, C. H. Wu, C. UniProt, UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31**, 926–932 (2015).