Mechanochemical evolution of the giant muscle protein titin as inferred from resurrected proteins

A. Manteca[1], J. Schönfelder[1], A. Alonso-Caballero[1], M. J. Fertin[1], N. Barruetabeña[1], B. F. Faria[2], E. Herrero-Galán[3], J. Alegre-Cebollada[3], D. De Sancho[1,4] & R. Perez-Jimenez[1,4,5]

[1]CIC nanoGUNE, San Sebastian, Spain.
[2]Laboratory of Molecular Modeling, Federal University of São João del-Rei, Brazil.
[3]Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), Madrid, Spain
[4]IKERBASQUE, Basque Foundation for Science, Bilbao, Spain.
[5]Evolution and Genomics Technologies, S. L. (Evolgene), San Sebastian, Spain.

Address correspondence to: R. P.-J: r.perezjimenez@nanogune.eu

Abstract

The sarcomere-based structure of muscles is conserved among vertebrates; however, vertebrate muscle physiology is extremely diverse. A molecular explanation for this muscle diversity and its evolution has not been proposed. We use phylogenetic analysis and single-molecule force spectroscopy (smFS) to investigate the mechanochemical evolution of titin, a giant protein responsible for the elasticity of muscle filaments. We bring back to life eight-domain fragments of titin corresponding to ancestors to mammals, sauropsids, and tetrapods, that lived 105-356 Myr, and compare them with some of their modern descendants. We demonstrate that resurrected titin molecules are rich in disulfide bonds and display high mechanical stability. These mechanochemical elements have changed over time creating a paleomechanical trend that seems to correlate with animal body size, allowing us to estimate the size of extinct species. We hypothesize that mechanical adjustments in titin contributed to physiological changes that allowed the muscular development and diversity of modern tetrapods.

Titin is a micrometer-long muscle protein composed of hundreds of individually folded domains[1]. Titin is present in all vertebrates, being one of the main components of the sarcomere together with actin and myosin. The main constituents of titin are immunoglobulin (Ig) domains, fibronectin type III domains and the unstructured PEVK region[2]. In the sarcomere, titin connects the Z disc to the M line (**Fig. 1a**). The main function of titin is providing passive elasticity to the muscle by acting like a spring. In addition, recent studies support an important role of titin refolding during contraction[3]. Although titin has been studied for decades, there is yet much to be explored regarding the correlation between muscle physiological diversity in animals and the biochemistry and nanomechanics of titin. Given the morphological and locomotor diversity in vertebrates, titin may hold the key for some phenotypes displayed by animals in terms of muscle physiology. In this respect, it can be hypothesized that the evolution of titin has been central to the acquisition of muscle diversity in animals. However, the role of titin in evolution, for instance, during the huge physiological outbreak after the Cambrian explosion 542 Myr[4], remains unexplored.

In the past two decades, our knowledge of the mechanical properties of titin has increased dramatically due to the use of single-molecule force spectroscopy (smFS) techniques, which make it possible to apply calibrated mechanical forces to titin domains[5,6]. In addition, the increasing amount of genetic data offers new avenues for comparative biology to better understand biological systems. In this regard, phylogenetic methods applied to genomic information have made it possible to establish evolutionary relationships among different living organisms including the possibility of inferring the putative sequences of the genes of their already extinct ancestors[7,8]. Ancestral sequence reconstruction allows us to track the evolutionary history of genes and proteins to obtain information about extinct species. This information relates to physiological and metabolic features[9,10], but also to the environmental conditions that hosted ancestral organisms[11,12].

Here, we have combined smFS and ancestral sequence reconstruction to investigate the evolution of the mechanical and biochemical properties of titin. We have used phylogeny to travel back in time and reconstruct a fragment of titin from different extinct species, including the last common ancestors of tetrapods, sauropsids and mammals. We expressed the ancient proteins in the laboratory and measured their mechanical properties using smFS. We found that differences in mechanical stability and disulfide bond occurrence between titin from living species and their ancestors appear as key elements that have driven the mechanical evolution of titin. These differences illustrate a paleomechanical trend, that is, a mechanical trend from the ancestral to modern animals, that allow us to establish physiological correlations purely based on the nanomechanical properties of titin. These correlations allow us predicting body size from ancestral species, comparable to those found in fossil records. Our experiments shed light on some of the molecular features that have driven the diversification and speciation of titin and possibly muscle physiology in vertebrates.

Results

Reconstructing ancestral titin molecules

We used 33 protein sequences of titin from modern vertebrate species, most of them corresponding to the complete protein sequence composed of more than 30,000 residues. The amino acid sequences were retrieved from Uniprot and GenomeNet databases containing titin from mammals, sauropsids, amphibians and ray-finned fishes (see Methods). Using these sequences we generated a sequence alignment and

constructed a phylogenetic chronogram using Bayesian inference as well as maximum parsimony (**Fig. 1b and Supplementary Fig. 1a and 1b**). From this tree we sampled several internal nodes for reconstruction of the most probable ancestral sequence. In particular, we chose the tree node corresponding to the last tetrapod common ancestor (LTCA) that lived in the early Carboniferous Period around 356 Myr; the one corresponding to the last sauropsid common ancestor (LSCA) that is thought to have lived in the Permian Period ~278 Myr; the last mammal common ancestor (LMCA) that lived in our planet in the Jurassic ~179 Myr; and finally, the last placental mammals common ancestor (LPMCA) from the mid-Cretaceous ~105 Myr (**Fig. 1b**). We used maximum likelihood to infer the most probable sequence of the ancestral nodes following well described procedures[9,11,12] (see Methods**)**. Posterior probability distributions across all sites for ancestral sequences are reported in **Supplementary Fig. 2**. The overall posterior probability of the sites lies between 0.90 and 0.99. Of note, the determination of ancestral sequences is not free from uncertainty. However, in experimental ASR the goal is determining the phenotypes and characteristics of proteins rather than the exact and accurate sequence. There is no one single exact reconstructed sequence that represents the true ancestor of a particular set of organisms, but the phenotype and characteristics of those sequences must be unique. Interestingly, ASR is able to capture those phenotypes. Over the past years, numerous studies have been carried out in order to assess the robustness of ASR[11,13-15].

We dated the nodes using multiple sources from the Time Tree Of Life[16] (TTOL), as well as paleontological data[17]. We selected for resurrection and laboratory testing an eight-domain fragment of titin encompassing domains I65 to I72 in the canonical human titin sequence (Uniprot Q8W2Z42), and the homologous in other species. This fragment is a good proxy of the elastic I-band region, located in the proximal tandem-Ig region of N2A skeletal titin and N2BA cardiac titin isoforms[18]. Part of this fragment has been characterized in terms of structure and mechanics[19,20]. Also, the alignment of this fragment is well resolved suggesting that it is structurally conserved.

Sequence conservation and the role of cysteines in sarcomeric proteins

The comparison of sequences of the inferred ancestral titin's I65-I72 with their modern counterparts yielded amino acid identities ranging from ~76 to 90% (**Supplementary Table 1**). We also compared titin sequence diversity levels with those of myosin II and actin, the other two most abundant proteins in the sarcomere (**Supplementary Table 1**). Actin is an extremely conserved protein that shows almost 99% identity across most living vertebrates, suggesting that it had very little influence on the molecular diversification of the sarcomere. In the case of myosin, identities of modern forms reach over 90%. We reconstructed a phylogenetic tree of myosin II from a similar pool of vertebrates as in the case of titin (**Supplementary Fig. 3**) and inferred ancient myosin sequences. Using ancestral and modern sequences, we observed that titin's I65-I72 fragment has around twice the mutation rate of myosin II (**Supplementary Fig. 4a and 4b**). This suggests that titin has contributed to the molecular diversification of the sarcomere more extensively than myosin. Furthermore, we determined the pattern of amino acid replacement in I65-I72 in the transition from LTCA to human. Amino acid replacements weighted by their relative mutability[21] show that Cys residues display mutability different than expected during the evolution of titin I65-I72 from LTCA to human. Human titin contains significantly fewer Cys residues than LTCA (**Supplementary Fig. 4c to 4f**). Assignment of Cys residues in the ancestral sequences are supported by values of posterior probability close to 1. This is significant given than

Cys is one of the least mutable residues in proteins, second only to tryptophan[21]. Cys residues are rarely lost once acquired[22]. We have carried out a similar analysis for a titin fragment from the distal region of the elastic I-band (I88-I95) and have observed a similar behavior for Cys (**Supplementary Fig. 4c to 4f**). In contrast, the same analysis in a fragment from the rigid A-band titin (I126-Fn90-94-I128) shows the opposite behavior in the mutability of Cys. This tendency is similar to that found in myosin. Altogether, these observations suggest that Cys residues played a crucial role in the molecular evolution of the elastic I-band of titin. Interestingly, cysteines are involved in the formation of disulfide bonds, modulating very efficiently the mechanical properties of the parent protein[23,24].

The existence of disulfides in titin was first proposed following identification of a disulfide bond in the crystal structure of domain I1, and the observation that many domains in titin contain proximal cysteines that can engage in disulfide bonds[25]. The crystal structures of different titin domains (pdb: 3b34, 2rik, 2rjm, 1g1c) have shown the existence of such disulfides. Indeed, studies of disulfide bonds in titin have been limited by deficient disulfide formation in recombinant expression systems, sometimes leading to contradictory results[26]. Nevertheless, experimental evidence of disulfide formation in native titin is still lacking. Given the fundamental role of disulfide bonds in the mechanics of proteins, we hypothesize that disulfides may be related to the mechanochemical evolution of titin. The genes encoding the four ancestral I65-I72 fragments were synthesized, the gene products expressed in *E. coli* and purified under equal oxidative stress conditions to overcome limited disulfide bond formation in the host. We also expressed and purified I65-I72 from representative modern amniote species covering different clades in our tree: human, brown rat, orca, chicken and zebra finch.

smFS reveals different paleomechanical trends for birds and mammals

To investigate the mechanical properties of the titin variants we performed smFS by mechanically stretching the proteins at a constant speed of 400 nm/s (**Fig. 2a, see Methods**). The stretching of titin domains leads to sawtooth patterns in force vs extension recordings, in which each peak represents the mechanical unfolding of an individual domain (**Fig. 2b**). The analysis of the traces using well established procedures[27], allows us determining the mechanical stability and contour length of the constituent domains. For all the variants tested we observed two distinct populations of peaks. The first one has contour lengths of ~30 nm and the second one displays lengths of 5 to 20 nm (**Supplementary Fig. 5**). We believe that the first population corresponds to fully extended domains of about 90 residues, whereas the second population represents disulfide-containing domains that make the contour length of the extended peptide shorter, as expected by the different position of the cysteines in the different domains[28]. This is quite significant since disulfide bonds in titin have been suggested to participate in titin mechanical regulation[24]. We have constructed cumulative histograms of mechanical stability for both populations. In the case of fully extended domains, we estimate an average unfolding force ranging from 180 to 218 pN (**Fig. 2c**), depending on the variant, being the lowest stability for orca titin and the highest for LTCA titin domains. Results for this first population of peaks are in agreement with previous mechanical characterization of I65-I70 purified in reducing conditions where no disulfide bonds can be established[20]. For the collection of domains that contains disulfide bonds, we determine average stability ranging from 134 to 182 pN depending on the variant (**Fig. 2d**), with orca titin having the lowest and zebra finch the highest. We plotted stabilities for both types of domains versus the age of each titin fragment

both for bird (**Fig. 3a and 3b**) and mammalian lineages (**Fig. 3c and 3d**). In the case of birds the stability has slightly changed for zebra finch and chicken with respect to LTCA and LSCA, but in opposite directions. In the case of mammals, we see a clear paleomechanical trend by which stability has decreased over time. Interestingly, in both birds and mammals, we observe that animals with larger body sizes have lower mechanical stabilities of titin domains with and without disulfide bridges. By counting the number of domains that contain disulfide bonds and comparing it with those that fully extend, we observed that in general ancestral proteins LTCA and LSCA have the highest proportion of S-S bonded domains (**Fig. 3e and 3f**).

Force clamp analysis confirms the evolutionary decay of disulfide bonds

The content of disulfide bonds seems to have decreased for modern species. In fact, the content of disulfide bonds seems to be related to the mechanical stability. More stable titin forms imply more disulfide bonded domains (**Supplementary Fig. 5**). To better identify disulfide bonds we performed single-molecule experiments in the force-clamp modality in order to capture reduction of disulfide bonds. In this mode, the force applied to the protein can be controlled. The mechanical unfolding is monitored as an increment in length versus time. We have used force-clamp techniques in previous work to demonstrate the kinetics of disulfide bond reduction under force by thioredoxin enzymes (Trx)[29,30]. Thioredoxin is an enzyme that controls the redox balance in cells by reducing disulfide bonds. The reduction of disulfide bonds by Trx is a force-dependent reaction that can be readily monitored and quantified[30]. By applying force to a disulfide-bonded titin domain we can trigger the unfolding of the domain up to the disulfide bond that becomes exposed to the solution. If Trx is present in the solution the disulfide bond can be reduced and the sequestered residues behind the disulfide bond are released giving rise to an extra extension of the polypeptide chain (**Supplementary Fig. 6a**). With this assay we can quantify disulfide bonds that are cryptic, i.e. that require exposure to be reduced. Non-cryptic disulfide bonds will be reduced without the need of mechanical exposure. We have applied this test to LSCA and human titin fragments in the presence of Trx. We first apply a pulse of force of 135 pN during 0.5 s that triggers unfolding of all the domains (**Fig. 4a, 4c**). The unfolding of the domains is monitored as a staircase of ~27 nm per step for reduced domains, and shorter steps between 5-20 nm for disulfide-bonded domains (**Fig. 4b, 4d**). The expected length of extended disulfide-containing domains varies due to the different position of the cysteines in the sequences. After the unfolding force pulse we quench the force to 80 pN for 20 s to monitor disulfide bond reductions as steps within the range 5-20 nm for LSCA titin and 15-20 nm for human titin (**Fig. 4a, 4c**). Again, the length attributed to reduction events will be different for each domain. Also, some domains have more than two cysteines which may also imply the possibility of isomerizations[31]. All possible disulfide bond combinations have been estimated and are shown in **Supplementary Table 2**. A histogram of the observed reduction events demonstrates that in LSCA it is common to observe up to 4 reduction events (**Fig. 4a**), whereas for human titin it is common to observe only one reduction event (**Fig. 4c**). In the absence of Trx no reduction events are observed (**Supplementary Fig. 6b to 6e**). Thus, force-clamp experiments confirm that LSCA titin contains more disulfide bonds than the human titin, supporting conclusions obtained using force extension for all I65-I72 fragments.

To provide independent measurements of cysteine oxidation, we used a biochemical assay that detects oxidized cysteines[32] by labeling them with the fluorophore monobromobimane (mBBr). In this assay, reduced thiols in the protein are alkylated with an excess of N-ethylmaleimide in denaturing conditions. After running

polyacrylamide electrophoresis, the gel is used as a reaction chamber to reduce oxidized thiols by incubation with dithiothreitol (DTT) and subsequent reaction of the newly reduced thiols with mBBr. The resulting fluorescence signal is proportional to the amount of oxidized thiols in the sample. Results show that LTCA and LSCA proteins contain more oxidized cysteines than rat and human proteins (**Fig. 4e**). In our biochemical assays, we also included a control $(I91\text{-}32/75)_8$ protein (formerly I27) that can be produced in reduced (no disulfides) or oxidized (1 disulfide per domain for a total of 8 disulfides) forms and whose oxidation status can be determined unambiguously by smFS[23]. Since the size of the control protein and the I65-I72 fragments is comparable, we used the normalized fluorescence signals to estimate the number of disulfides in LTCA (6), LSCA (5), human and rat (3-4). This independent method of determining the population of disulfide bonds confirms the trend observed in smFS experiments for more disulfides in the ancestral proteins (**Fig. 4e**). There are, however, small discrepancies between the exact number of disulfides from the different methods that probably reflect contributions of terminal cysteines needed for attachment in smFS experiments or other forms of oxidation different from disulfides, such as sulfenylation induced by treatment with H_20_2.

Discussion

Our data suggests that small animals have titin domains with greater mechanical resistance than large animals. In fact, this is consistent with the fact that in small mammal hearts the isoform N2A, which is stiffer than N2BA, is more abundant [18]. We wondered if mechanical stability is related to animal size. This correlation may be related to titin mechanics as small animals have faster muscle contraction and shivering frequency[33]. To probe this idea, we plotted the mechanical stability of domains, with and without disulfide bonds, versus body mass. We observed that the dependence of the average unfolding force with the body mass can be fitted to a power law, suggesting an evolutionary allometric relationship (**Fig. 5a and 5b**). Body mass has been shown to follow allometric scaling with other physiological traits such as metabolic rate, speed, arterial pressure or heat production[34]. In fact, it is common to observe allometric scaling in biological systems, and even enzyme activity has been suggested to show allometry[35]. However, to the best of our knowledge, this is the first observation showing allometric scaling between a physiological feature and molecular-level parameters. From these correlations body mass of extinct species could be predicted. Inasmuch as we have two allometric scaling plots (**Fig. 5a and 5b**), we determined a range of body mass for each extinct species that is determined by the minimum and maximum value of mass obtained. This is 8-70 gr for LTCA, 14-16 gr for LSCA, 14-70 gr for LMCA and 95-116 gr for LPMCA. These weights are typical of small animals, perhaps between 10 and 40 cm in length. These estimates compare surprisingly well with sizes from fossils that could be related to these extinct animals[36-41]. Thus, the observed correlation between mechanical stability and body mass allows us predicting the size of extinct species. However, it must be noted that this correlations seem to be more sensitive for high mechanical stability of titin molecules.

Our results show that the evolution of muscle physiology seems to be linked to the molecular evolution of titin in tetrapods. The reconstruction of ancient forms of titin demonstrates that titin domains from the small zebra finch are similar to those of its ancestors, in terms of mechanical properties. However, titin domains from modern mammals have experienced more drastic changes leading to proteins that have lower mechanical stability and fewer disulfide bonds compared to those in their oldest ancestor. This is quite remarkable given that mutation rates for titin in zebra finch and

living mammals are quite similar. Most likely, these changes are related to morphological and physiological consequences that derived in the vast diversity of physical and locomotor features found in mammals. This raises the question of whether the relation of titin mechanics with muscle contraction differs across species with different physiological characters. A possible interpretation is that small amniotes with fast muscle contractions rely upon the mechanical response of titin domains. Under physiological forces, titin domains have been shown to unfold and refold during muscle contraction[3]. The presence of disulfide bonds prevents overstretching of titin and increases the recoiling speed of the domains[28], probably increasing also the speed of muscle contraction. Thus, we hypothesize that the balance between mechanical stability and disulfide bonds may be a key factor in titin mechanical regulation and its evolution. Another possible interpretation relates to the fact that chemical modifications such as S-Glutathionylation, or oxidative stress have been suggested to occur in titin and both involve cysteine residues[6]. But, even these phenomena are still related to titin mechanics. Although other reasons might be related to the mechanochemical differences observed, it is hard to speculate with an interpretation other than the purely mechanical, given the clear elastic character of titin.

The size that we obtain for ancient species from the allometric correlations may be explained under the light of fossil remains. It has been demonstrated that after the mass extinction occurred in the late Devonian, the so-called Hangenberg event (359 Ma), the majority of taxa found in fossil records were under 40 cm[41]. This is consistent with a global shrinkage process during the early Carboniferus that lasted around 40 Myr. Our predicted data for LTCA, which supposedly lived after the Hangenberg event, are within the range of the sizes reported. Fossils of early amniotes including sauropsids and mammalian-like reptiles from the Carboniferous and Permian, are also within that range[39,42]. In the case of mammals, numerous findings have shown that early mammals, including placentals, were small rodent-like animals[36-38,40,43], which is in line with the sizes that we predict from nanomechanical information of titin. Nevertheless, establishing direct comparison between our estimations and any known fossils is difficult, because it is unlikely that a single fossil could be unambiguously labeled as LTCA, LSCA, LMCA or LPMCA, representing the true common ancestor of different taxonomic groups.

The eight-domain fragments of titin that we have studied are significant because they belong to the critical elastic region of titin. Studying additional segments of titin as well as comparing features such as folding kinetics of ancient and modern domains will be interesting in order to gain a complete understanding of the molecular elements that have driven the molecular evolution of titin and its connection to muscle physiology.

Acknowledgements

Research has been supported by Ministry of Economy and Competitiveness (MINECO) grant BIO2016-77390-R, BFU2015-71964 to R. P-J.; BIO2014-54768-P and RYC-2014-16604 to J.A-C., and CTQ2015-65320-R to D. DS.; European Commission grant CIG Marie Curie Reintegration program FP7-PEOPLE-2014 to R. P-J. A. A-C. is funded by the predoctoral program of the Basque Government. R. P-J., D. DS., thank CIC nanoGUNE and Ikerbasque Foundation for Science for financial support. CNIC is supported by the Spanish Ministry of Economy and Competitiveness (MINECO) and the Pro-CNIC Foundation, and is a Severo Ochoa Center of Excellence (MINECO award SEV-2015-0505). Plasmid pQE80-$(I91\text{-}32/75)_8$ was a kind gift from Professor J. Fernández (Columbia University, NY, USA). We thank Professor R. Zardoya (National Museum of Natural Sciences Madrid, Spain) for helpful discussions and comments. The

authors acknowledge technical support provided by IZO-SGI SGIker of UPV/EHU and European funding (ERDF and ESF) for the use of the Arina HPC cluster, and the assistance provided by T. Mercero and E. Ogando.

Author contributions

R. P-J. designed research. A. M., B. F. N. B., D. DS. and R. P-J. conducted phylogenetic analysis. A. M. and M. F. cloned and expressed proteins. A. M., A. A-C., J.S., D. DS. and R. P-J. performed AFM experiments and data analysis. E. H-G. and J. A.-C. performed biochemical determination of disulfides in titin fragments. R.P-J. drafted the paper and all authors contributed in revising and editing the manuscript.

Competing financial interests

The authors declare no competing financial interest.

Author information

Reprints and permissions information is available at http://www.nature.com/nsmb.

Correspondence and requests for materials should be addressed to R. P.-J: r.perezjimenez@nanogune.eu

References

1. Furst, D.O., Osborn, M., Nave, R. & Weber, K. The organization of titin filaments in the half-sarcomere revealed by monoclonal antibodies in immunoelectron microscopy: a map of ten nonrepetitive epitopes starting at the Z line extends close to the M line. *J. Cell. Biol.* **106**, 1563-72 (1988).
2. Labeit, S. & Kolmerer, B. Titins: giant proteins in charge of muscle ultrastructure and elasticity. *Science* **270**, 293-6 (1995).
3. Rivas-Pardo, J.A. et al. Work Done by Titin Protein Folding Assists Muscle Contraction. *Cell Rep.* **14**, 1339-47 (2016).
4. Erwin, D.H. et al. The Cambrian conundrum: early divergence and later ecological success in the early history of animals. *Science* **334**, 1091-7 (2011).
5. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J.M. & Gaub, H.E. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **276**, 1109-12 (1997).
6. Alegre-Cebollada, J. et al. S-glutathionylation of cryptic cysteines enhances titin elasticity by blocking protein folding. *Cell* **156**, 1235-46 (2014).
7. Hall, B.G. Simple and accurate estimation of ancestral protein sequences. *Proc. Natl. Acad. Sci. USA* **103**, 5431-6 (2006).
8. Merkl, R. & Sterner, R. Ancestral protein reconstruction: techniques and applications. *Biol. Chem.* **397**, 1-21 (2016).
9. Kratzer, J.T. et al. Evolutionary history and metabolic insights of ancient mammalian uricases. *Proc. Natl. Acad. Sci. USA* **111**, 3763-8 (2014).
10. Zakas, P.M. et al. Enhancing the pharmaceutical properties of protein drugs by ancestral sequence reconstruction. *Nat. Biotechnol.* (2016).
11. Gaucher, E.A., Govindarajan, S. & Ganesh, O.K. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* **451**, 704-7 (2008).
12. Perez-Jimenez, R. et al. Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. *Nat. Struct. Mol. Biol.***18**, 592-6 (2011).
13. Eick, G.N., Bridgham, J.T., Anderson, D.P., Harms, M.J. & Thornton, J.W. Robustness of Reconstructed Ancestral Protein Functions to Statistical Uncertainty. *Mol. Biol. Evol.* **34**, 247-261 (2017).
14. Hanson-Smith, V., Kolaczkowski, B. & Thornton, J.W. Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. *Mol. Biol. Evol.* **27**, 1988-99 (2010).
15. Randall, R.N., Radford, C.E., Roof, K.A., Natarajan, D.K. & Gaucher, E.A. An experimental phylogeny to benchmark ancestral sequence reconstruction. *Nat. Commun.* **7**, 12847 (2016).
16. Hedges, S.B., Marin, J., Suleski, M., Paymer, M. & Kumar, S. Tree of life reveals clock-like speciation and diversification. *Mol. Biol. Evol.* **32**, 835-45 (2015).

17. Benton, M.J. et al. Constraints on the timescale of animal evolutionary history. *Palaeontol. Electron.* **18**, 1-106 (2015).
18. Neagoe, C., Opitz, C.A., Makarenko, I. & Linke, W.A. Gigantic variety: expression patterns of titin isoforms in striated muscles and consequences for myofibrillar passive stiffness. *J. Muscle Res. Cell. Motil.* **24**, 175-89 (2003).
19. von Castelmur, E. et al. A regular pattern of Ig super-motifs defines segmental flexibility as the elastic mechanism of the titin chain. *Proc. Natl. Acad. Sci. USA* **105**, 1186-91 (2008).
20. Watanabe, K., Muhle-Goll, C., Kellermayer, M.S., Labeit, S. & Granzier, H. Different molecular mechanics displayed by titin's constitutively and differentially expressed tandem Ig segments. *J. Struct. Biol.* **137**, 248-58 (2002).
21. Dayhoff M.O., S.R.M., Orcutt B.C. *Amino acid scale: Relative mutability of amino acids (Ala=100)*, (National Biomedical Research Foundation, Silver Spring, MD, 1978).
22. Wong, J.W., Ho, S.Y. & Hogg, P.J. Disulfide bond acquisition through eukaryotic protein evolution. *Mol. Bio.l Evol.* **28**, 327-34 (2011).
23. Kosuri, P. et al. Protein folding drives disulfide formation. *Cell* **151**, 794-806 (2012).
24. Grutzner, A. et al. Modulation of titin-based stiffness by disulfide bonding in the cardiac titin N2-B unique sequence. *Biophys. J.* **97**, 825-34 (2009).
25. Mayans, O., Wuerges, J., Canela, S., Gautel, M. & Wilmanns, M. Structural evidence for a possible role of reversible disulphide bridge formation in the elasticity of the muscle protein titin. *Structure* **9**, 331-40 (2001).
26. Li, H. & Fernandez, J.M. Mechanical design of the first proximal Ig domain of human cardiac titin revealed by single molecule force spectroscopy. *J. Mol. Biol.* **334**, 75-86 (2003).
27. Perez-Jimenez, R., Garcia-Manyes, S., Ainavarapu, S.R. & Fernandez, J.M. Mechanical unfolding pathways of the enhanced yellow fluorescent protein revealed by single molecule force spectroscopy. *J. Biol. Chem.* **281**, 40010-4 (2006).
28. Ainavarapu, S.R. et al. Contour length and refolding rate of a small protein controlled by engineered disulfide bonds. *Biophys. J.* **92**, 225-33 (2007).
29. Perez-Jimenez, R. et al. Diversity of chemical mechanisms in thioredoxin catalysis revealed by single-molecule force spectroscopy. *Nat. Struct. Mol. Biol.* **16**, 890-6 (2009).
30. Wiita, A.P. et al. Probing the chemistry of thioredoxin catalysis with force. *Nature* **450**, 124-7 (2007).
31. Alegre-Cebollada, J., Kosuri, P., Rivas-Pardo, J.A. & Fernandez, J.M. Direct observation of disulfide isomerization in a single protein. *Nat. Chem.* **3**, 882-7 (2011).
32. Rogers, L.K., Leinweber, B.L. & Smith, C.V. Detection of reversible protein thiol modifications in tissues. *Anal. Biochem.* **358**, 171-84 (2006).
33. Legendre, L.J., Guenard, G., Botha-Brink, J. & Cubo, J. Palaeohistological Evidence for Ancestral High Metabolic Rate in Archosaurs. *Syst. Biol.* (2016).
34. Lindstedt, S.L. & Schaeffer, P.J. Use of allometry in predicting anatomical and physiological parameters of mammals. *Lab. Anim.* **36**, 1-19 (2002).
35. Burness, G.P., Leary, S.C., Hochachka, P.W. & Moyes, C.D. Allometric scaling of RNA, DNA, and enzyme levels: an intraspecific study. *Am. J. Physiol.* **277**, R1164-70 (1999).
36. Luo, Z.X. et al. Mammalian evolution. Evolutionary development in basal mammaliaforms as revealed by a docodontan. *Science* **347**, 760-4 (2015).
37. Luo, Z.X., Yuan, C.X., Meng, Q.J. & Ji, Q. A Jurassic eutherian mammal and divergence of marsupials and placentals. *Nature* **476**, 442-5 (2011).
38. Martin, T. et al. A Cretaceous eutriconodont and integument evolution in early mammals. *Nature* **526**, 380-4 (2015).
39. Carroll, R.L. A Middle Pennsylvanian captorhinomorph, and the interrelationships of primitive reptiles. *J. Paleo.* **43**, 151-170 (1969).
40. Bi, S., Wang, Y., Guan, J., Sheng, X. & Meng, J. Three new Jurassic euharamiyidan species reinforce early divergence of mammals. *Nature* **514**, 579-84 (2014).
41. Sallan, L. & Galimberti, A.K. Body-size reduction in vertebrates following the end-Devonian mass extinction. *Science* **350**, 812-5 (2015).
42. Paton, R.L., Smithson, T.R. & Clack, J.A. An amniote-like skeleton from the Early Carboniferous of Scotland. *Nature* **398**, 508-513 (1999).
43. O'Leary, M.A. et al. The Placental Mammal Ancestor and the Post–K-Pg Radiation of Placentals. *Science* **339**, 662-667 (2013).

Figure legends

Figure 1. Reconstruction of ancestral titin fragments. (**a**) Scheme of one half of the sarcomere from Z disk to M line. The three main sarcomeric proteins actin, myosin and titin are shown. The segment encompassing domains I65 to I72 from the elastic part of titin has been selected for ancestral reconstruction and testing. (**b**) Uncorrelated lognormal relaxed-clock chronogram of titin with geological time inferred with Bayesian inference. A total of 33 titin genes were used. The modern species studied are indicated by the animal outlines: zebra finch, chicken, orca, rat and human. The internal nodes LTCA, LSCA, LMCA and LPMCA were selected for resurrection and laboratory testing. They represent the last common ancestor of tetrapods (356±11 Myr), sauropsids (278±14 Myr), mammals (179±38 Myr) and placental mammals (105±17 Myr), respectively. Posterior probabilities for branch support are shown in the nodes. Geological times are shown in the upper bar. Outlines were retrieved from www.phylopic.org.

Figure 2. Single-molecule experiments of titin. (**a**) Schematic representation of a single-molecule experiment using the smFS (not to scale). Although I65-I72 contains 8 domains, only four of them are represented for clarity. Disulfide bonded domains are shown in red/grey with the cysteine highlighted in yellow. Non-disulfide bonded domains are shown in blue. The protein is mechanically stretched between a cantilever tip and a gold-coated surface. (**b**) Representative experimental traces of the polyprotein I65-I72. The unfolding of domains is monitored as a sawtooth pattern of force versus extension peaks. The worm-like chain model was used to fit the data. Fully extended domains (blue lines) show extension of about 30 nm, whereas disulfide bonded domains (red lines) show shorter contour lengths of 5-20 nm. (**c**) Cumulative histogram of mechanical unfolding force for domains that do not contain disulfide bonds and (**d**) domains that are disulfide bonded.

Figure 3. Mechanical stability versus geological time. (**a**) Unfolding forces of non-disulfide bonded domains for birds and ancestors and, (**b**) for mammals and ancestors. (**c**) Unfolding forces of disulfide-bonded domains for birds and ancestors and, (**d**) for mammals and ancestors. Error bars indicate the 95% confidence intervals for the sample mean. Percentage of disulfide bonded domains detected in force-extension traces for mammals (**e**) and sauropsids (**f**). LTCA and LSCA titin fragments are the ones presenting higher percentage of disulfide bonds in both, mammals and birds. In general, modern animals display fewer disulfide bonds than their ancestors. Data collection for each protein is n=374 for LTCA, n=614 for LSCA, n=407 for LMCA, n=366 for LPMCA, n=409 for zebra finch, n=375 for chicken, n=263 for orca, n=341 for rat and n=347 for human titin fragment.

Figure 4. Force-clamp experiments for detection of disulfide bond reductions catalyzed by thioredoxin enzymes. (**a**) Experimental force-clamp trace of LSCA titin. We first apply a pulse of force of 135 pN during 2s that triggers unfolding of non-disulfide bonded domains, indicated in the inset with arrows; and disulfide bonded domains up to their S-S bond, indicated in the inset with asterisks. The disulfide bonds can be reduced by Trx enzymes present at 10 μM concentration. The reduction events are monitored at a force of 80 pN, indicated in the green line. A histogram of the number of reduction events per trace is shown. (**b**) Histogram of step size for unfolding events (grey) and disulfide bond reductions by Trx (green) in LSCA titin (n=372). (**c**)

Experimental force-clamp trace for human titin. We rarely observe more than one reduction. (**d**) Step size histogram for unfolding events (grey) and reduction events (green) (n=267). (**e**) In-gel determination of oxidized thiols for LTCA, LSCA, rat (R) and human (H) I65-I72. Fluorescent bands resulting from the labeling of oxidized thiols with mBBr were normalized by the total quantity of protein as assessed by Coomassie staining and densitometry. The oxidized (Ox) and reduced (Red) versions of $(I91\text{-}32/75)_8$ were used as controls. Mean values ± S.D. of n=3 independent experiments are represented (for human titin, n=4)

Figure 5. Correlation of mechanical stability with animal body mass for: (**a**) disulfide bonded domains; (**b**) non-disulfide bonded domains. Stability versus body mass follows a power law correlation; for S-S bonded domains: $F\text{-}F_0 = 73*M^{-0.17}$; for domains without disulfides: $F\text{-}F_0 = 65*M^{-0.26}$. Modern species are represented in grey circles. The values of body mass for ancient species LTCA, LSCA, LMCA and LPMCA can be interpolated from the different fittings and are represented in black squares.

Online Methods

Phylogenetic analysis and ancestral sequence reconstruction: A set of 33 titin sequences were used from which 28 correspond to the full sequence of titin with over 30K residues. The sequences represent five different classes of vertebrate animals, mammals, amphibians, reptiles, birds and bony fishes, and were retrieved from the UniProt and GenomeNet databases. All sequence ID numbers are listed in Supplementary Information. The sequences were aligned using the MUSCLE software and further edited manually. We tested the alignment for best model of protein evolution using ProTest[44], resulting the Jones-Taylor-Thornton (JTT) with gamma distribution model as the best evolution model. A second set of sequences was used containing new sequences that were added to databases at a later stage during the course of this study. We decided to include these new sequences to test the robustness of our tree as most additions were in the Sauropsida and Amphibia clades that were less represented in our initial set. Two phylogenies were performed; first, using Bayesian inference using Markov Chain Monte Carlo (MCMC); and second, by the maximum parsimony criterion, with identical results in tree topology in all cases. For Bayesian inference we used BEAST v1.8.2 package software[45] incorporating the BEAGLE library for parallel processing. We set monophyletic groups for primates, rodentia, carnivora, chiroptera, cetartiodactyla, archosaurs, testudines, squamata and fishes, being the latest the selected outgroup. We set the JTT model with 8 categories gamma distribution, Yule model for speciation and length chain of 25 million generations, sampling every 1000 generations. Calculations were run for 12 days in a single node of an HPC cluster of Intel Xeon 2680v2 processors, using 16-cores at 2.6 GHz and 64 GB of memory. A 12-core iMac computer was used for smaller trees such as myosin. We discarded the initial 30% of trees as burn-in. All nodes were supported by posterior probabilities above 0.63 with most of them nearly 1. The myosin tree was performed using only BEAST. Tree Annotator was used to estimate a maximum clade credibility tree removing 30% of initial trees as burn-in. FigTree v1.4.2 was used for tree representation and editing. For parsimony we used PAUP* 4.0 software[46] using the heuristic search option and performing 2,000 bootstrap replicates. All bifurcations showed high bootstrap support, with most of them around 100% and a minimum of 67%. Divergence times were collected from different sources using both molecular clocks as well as paleontological records[17]. Finally, ancestral sequence reconstruction is performed by maximum likelihood using PAML 4.8[47], incorporating a gamma distribution for variable replacement rates across sites and the JTT model. The ancestral sequences are listed in the Supplementary Information. Posterior probabilities were calculated for all 20 amino acids. In each site, the residue with the highest posterior probability was selected. We have resurrected titin I65-I72 titin sequences that belong to the last tetrapod common ancestor (LTCA) which lived around 356±11 Ma; the last sauropsida common ancestor LSCA, 278±14 Ma; the last mammal common ancestor LMCA, 179±38 Ma; the last placental mammal common ancestor LPMCA, 105±17 Ma.

Protein expression and purification: Genes encoding the ancestral and extant titin proteins were synthesized and codon-optimized for expression in *E. coli* cells (Life Technologies). The genes were cloned into pQE80L vector (Qiagen) and transformed onto *E. coli* Origami2 cells with enhanced disulfide bond formation machinery (Merck Millipore). Bacteria were incubated overnight in LB medium at 37 °C, and 1 mM IPTG was add after reaching O.D of 0.5 to induce protein expression. After centrifugation, cell pellets were lysed with French pressure cell press and the His_6-tagged proteins were

loaded onto His GraviTrap affinity column (GE Healthcare). Oxidation was triggered by addition of 0.5% H_2O_2 overnight at room temperature. The proteins were then further purified by size exclusion chromatography using a Superdex 200HR column (GE Healthcare). The buffer used was 10 mM HEPES, pH 7.2, 150 mM NaCl, 1 mM EDTA at pH 7.0. The purified proteins were finally verified by SDS-PAGE. Trx and oxidized and reduced $(I91\text{-}32/75)_8$ were purified as described before[29,48].

Biochemical determination of oxidized cysteines:A protocol for in-gel determination of oxidized thiols was adapted and optimizedfrom previous reports[32] . 1 µg of each protein was incubated with 10 mM N-Ethylmaleimide in HEPES buffer in the presence of 3% w/v SDS for 30 min at 60 ºC in order to block all initially reduced thiols by irreversible alkylation. Samples were subsequently run on a 12% SDS-PAGE gel, and oxidized thiols were then reduced by incubation of the gel with 10 mM DTT for 1 hour at 60 ºC. After three washes with 50 mMTris-HCl, 10 mM EDTA, 3% w/v SDS, pH 6.8, the gel was incubated with a 5 mMmBBr solution in the same buffer for 2 h in the dark. Excess mBBr was removed by destaining the gel with 40% ethanol, 10% acetic acid overnight (3 changes). Fluorescent bands resulting from the reaction of the newly reduced thiols with mBBr were visualized on a Gel-Doc with UV excitation using standard filters for ethidium bromide emission. Quantification of the bands was performed by densitometry using the Quantity One software. The amount of protein in each well was later assessed by Coomassie staining and densitometry, and was used to normalize fluorescence signals. The fluorescent background of a replicate gel that was not reduced with DTT was subtracted. Oxidized and reduced $(I91\text{-}32/75)_8$ control proteins were also included in the experiment. Using force-clamp measurements[23], we estimated that 99% of I91-32/75 domains of the oxidized sample were oxidized, while 95% of the domains in the reduced sample were reduced. These experimental values were used to estimate number of disulfides in the I65-I72 samples.

Single-molecule Force Spectroscopy experiments: We performed the experiments on a commercial Atomic Force Spectrometer (AFS) from Luigs & Neumann. Cantilever models MLCT and OBL-10 of silicon nitride were used (Bruker). We calibrated the cantilevers using the equipartition theorem giving rise to a typical spring constant of 0.02 N $m^{-1}$for MCLT and 0.006 N m^{-1} for OBL-10. The AFS works in the force-extension mode at a pulling speed of 400 nm/s and amplitude 400 nm. In the force-clamp mode, the length resolution is 0.5 nm and the piezo feedback response can be as low as 1 ms. The buffer used in the experiment was 10 mM HEPES, pH 7.2, 150 mMNaCl, 1 mM EDTA and 2 mM NADPH. We added eukaryotic Trx enzyme to a final concentration of 10 µM. The buffer also contains eukaryotic Trx reductase (50 nM) to keep Trx enzyme in their reduced state. To perform the experiment, we deposited 10-20 µl of substrate ~0.1 mg ml^{-1} on a gold-covered coverslide. A drop of ~100 µl containing the Trx solution was then added. The force-clamp protocol consists of three pulses of force. In the first pulse the cantilever tip was pressed against the surface at 800 pN for 2 s. In the second pulse the attached I65-I72 titin was stretched at 135 pN for 2 s. The third pulse was the test force where the reduction events were captured. This pulse was applied at 80 pN for 20s to capture all possible reduction events. The traces were collected and analyzed using custom-written software in Igor Pro 6.37 (Wavemetrics). All proteins were measured following a double-blind protocol in which three independent researchers measured all samples labeled as A, B, C,… from different expressions, using several cantilevers for each protein and two AFS instruments. All figures were generated using Igor Pro and Adobe Illustrator CS6.

Data analysis

When we represent averages of the experimental observables (unfolding forces for the disulfide bonded and non-disulfide bonded domains) the error bars indicate the standard error of the mean. The percentage of disulfide bonds was estimated as the fraction of unfolding events corresponding to the disulfide bonded (i.e. low extension) population. For the correlation between unfolding forces and mass we use a power law expression including a pre-defined offset to fit the data, $F\text{-}F_0=a\cdot m^b$, where F_0 corresponds to the lowest measured value for the force, reflecting a lower limit in the unfolding force of the titin domains. Least-squares fits were performed using the Levenberg-Marquardt algorithm.

Data availability

The data that support the findings of this study are available from the corresponding author upon request.

44. Abascal, F., Zardoya, R. & Posada, D. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104-5 (2005).
45. Drummond, A.J., Suchard, M.A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969-73 (2012).
46. Wilgenbusch, J.C. & Swofford, D. Inferring evolutionary trees with PAUP*. *Curr Protoc Bioinformatics* **Chapter 6**, Unit 6 4 (2003).
47. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586-91 (2007).
48. Kahn, T.B., Fernandez, J.M. & Perez-Jimenez, R. Monitoring Oxidative Folding of a Single Protein Catalyzed by the Disulfide Oxidoreductase DsbA. *J. Biol. Chem.* **290**, 14518-27 (2015).

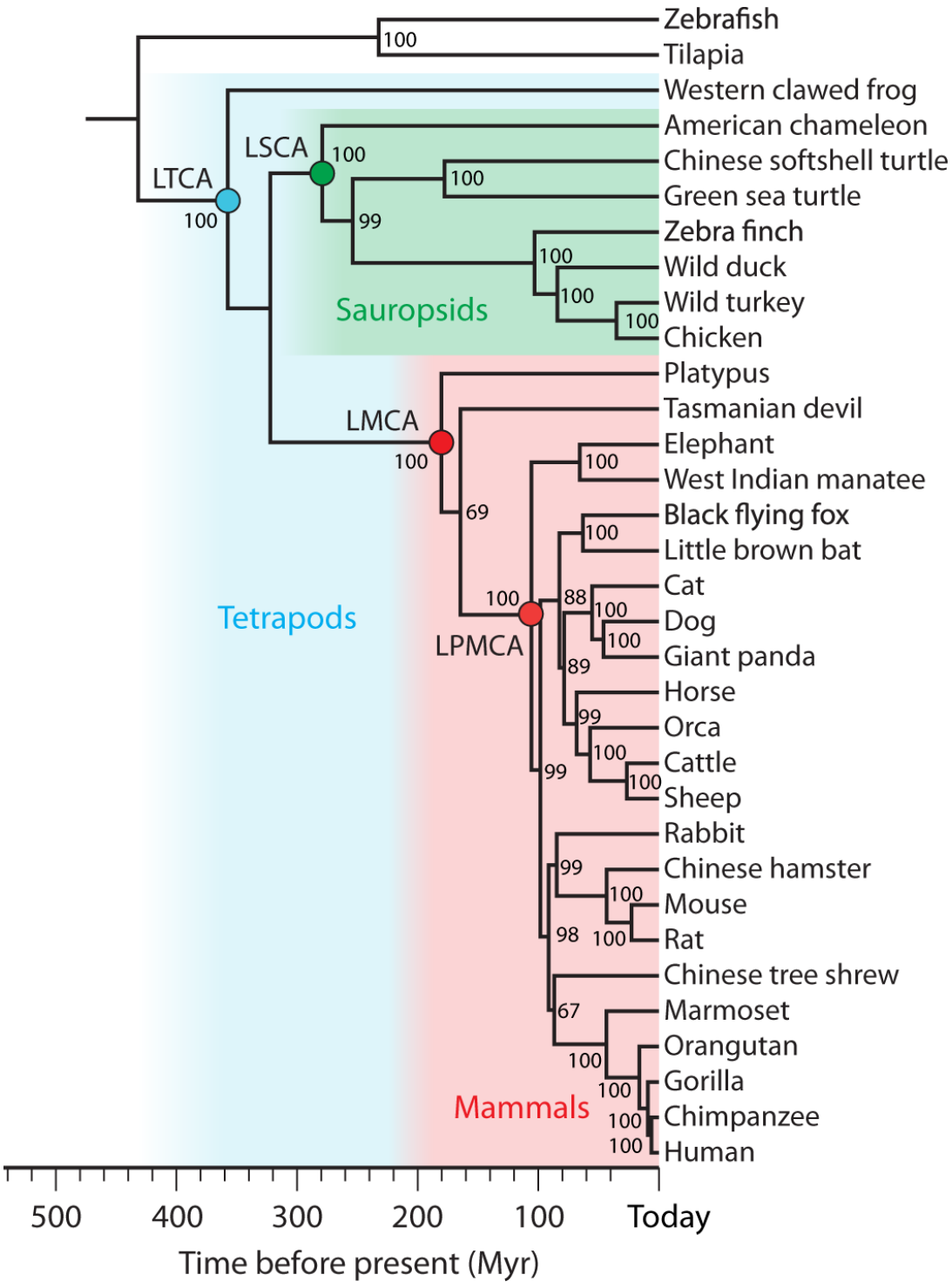
SUPPLEMENTARY INFORMATION

Mechanochemical evolution of the giant muscle protein titin as inferred from resurrected proteins

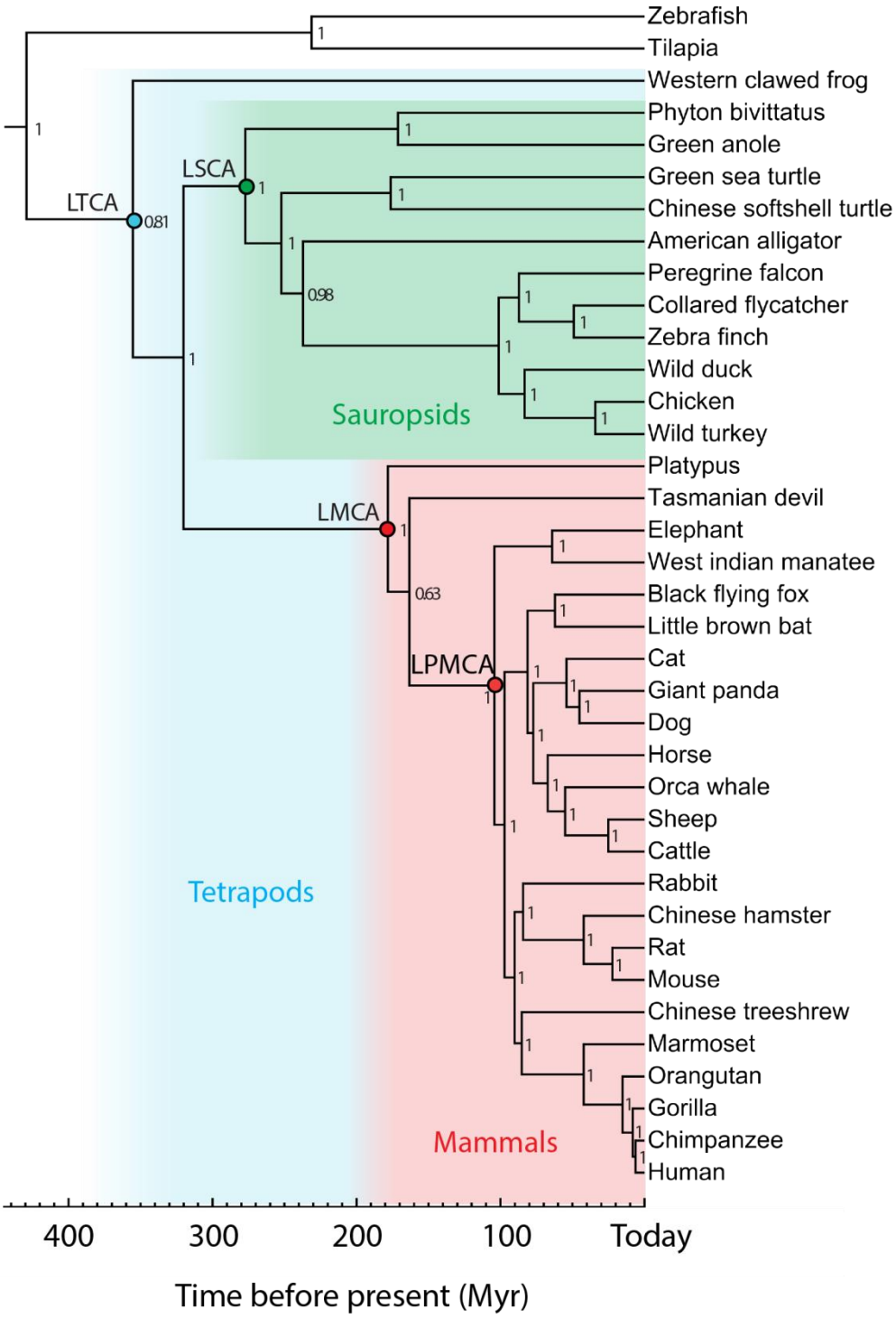
A. Manteca[1], J. Schönfelder[1], A. Alonso-Caballero[1], M. J. Fertin[1], N. Barruetabeña[1], B. F. Faria[2], E. Herrero-Galán[3], J. Alegre-Cebollada[3], D. De Sancho[1,4] & R. Perez-Jimenez[1,4,5]

[1]CIC nanoGUNE, San Sebastian, Spain.
[2]Laboratory of Molecular Modeling, Federal University of São João del-Rei, Brazil.
[3]Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), Madrid, Spain.
[4]IKERBASQUE, Basque Foundation for Science, Bilbao, Spain.
[5]Evolution and Genomics Technologies, S. L. (Evolgene), San Sebastian, Spain.

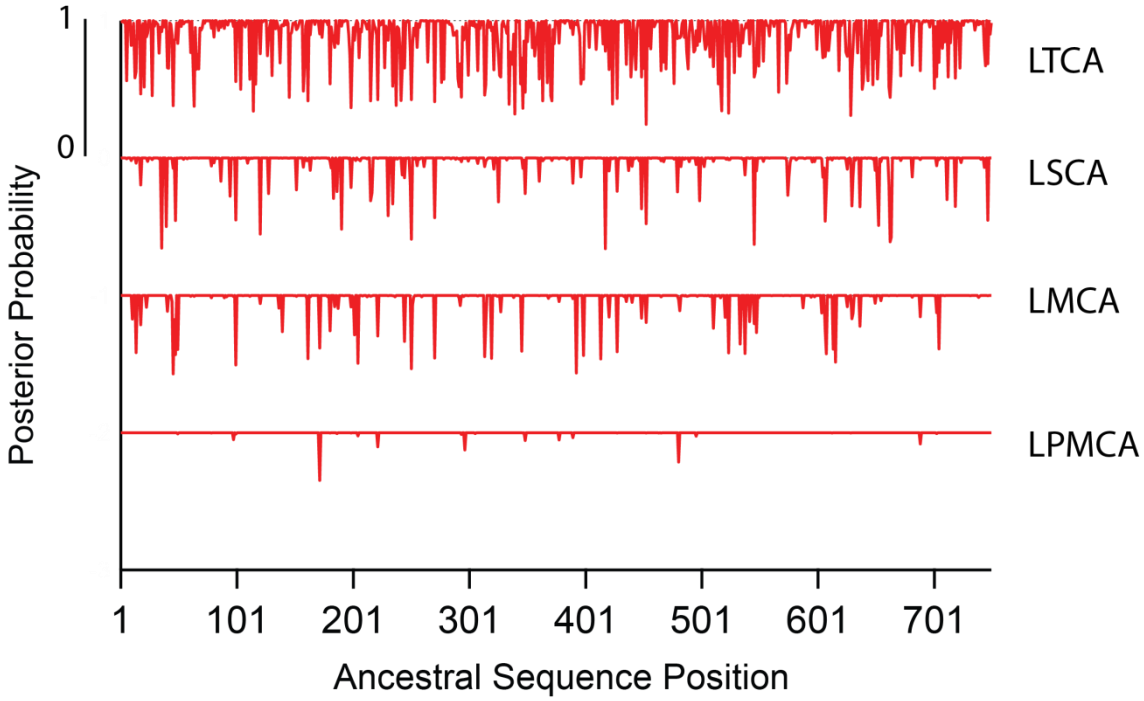
Address correspondence to: R. P.-J: r.perezjimenez@nanogune.eu



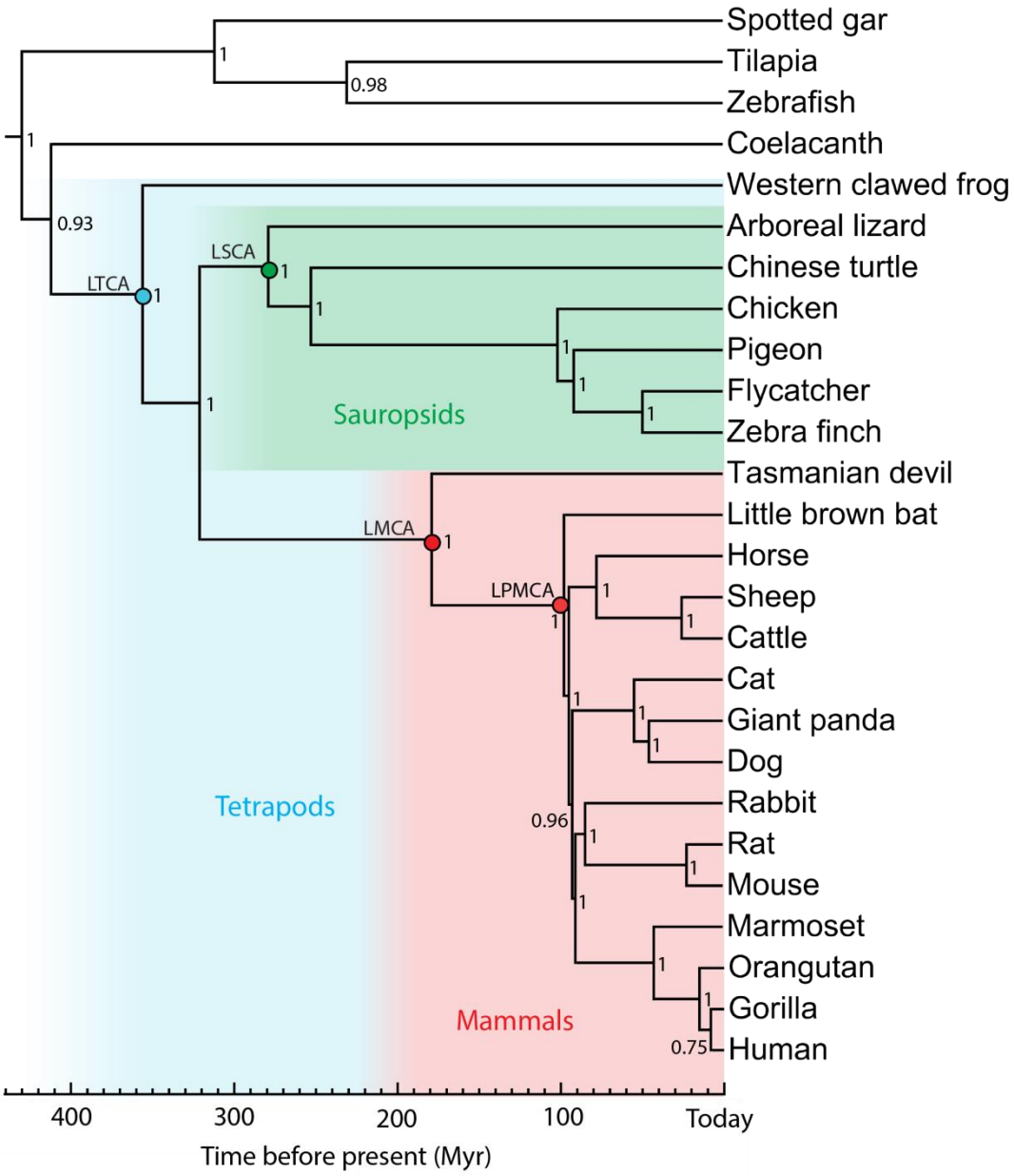
Supplementary Figure 1. Phylogenetic tree of titin using parsimony. A total of 33 titin molecules from different animals were used to reconstruct the phylogentic tree. Bootstrap support for each bifurcation is indicated. Different sources [1,2] were considered for divergence times of ancestral nodes**.**



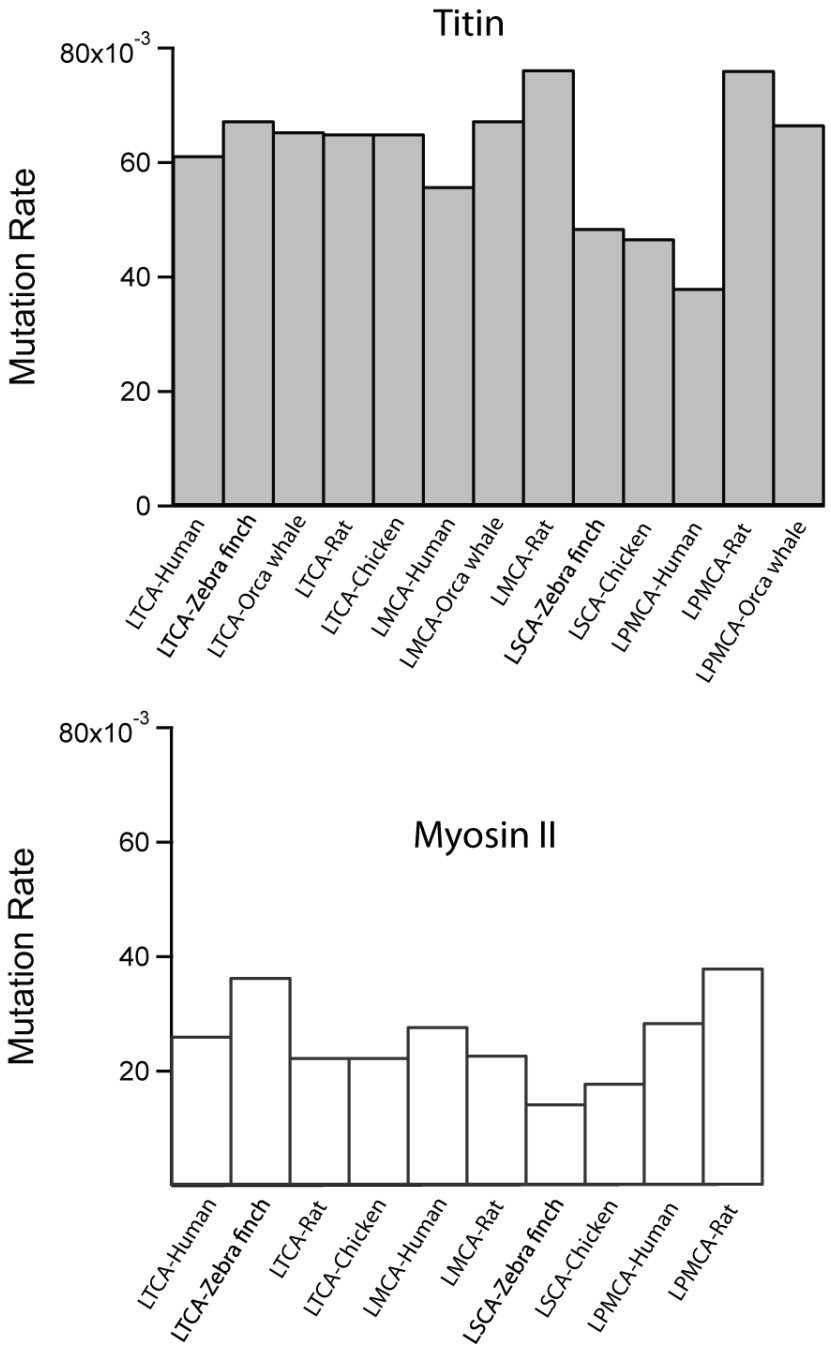
Supplementary Figure 2. Phylogenetic tree of titin using Bayesian inference. This tree contains new and updated sequences that appeared later during this study. A total of 37 titin molecules from different animals were used to reconstruct the phylogentic tree by Bayesian inference using MCMC. Posterior probabilities for each bifurcation are indicated. Different sources [1,2] were considered for divergence times of ancestral nodes.



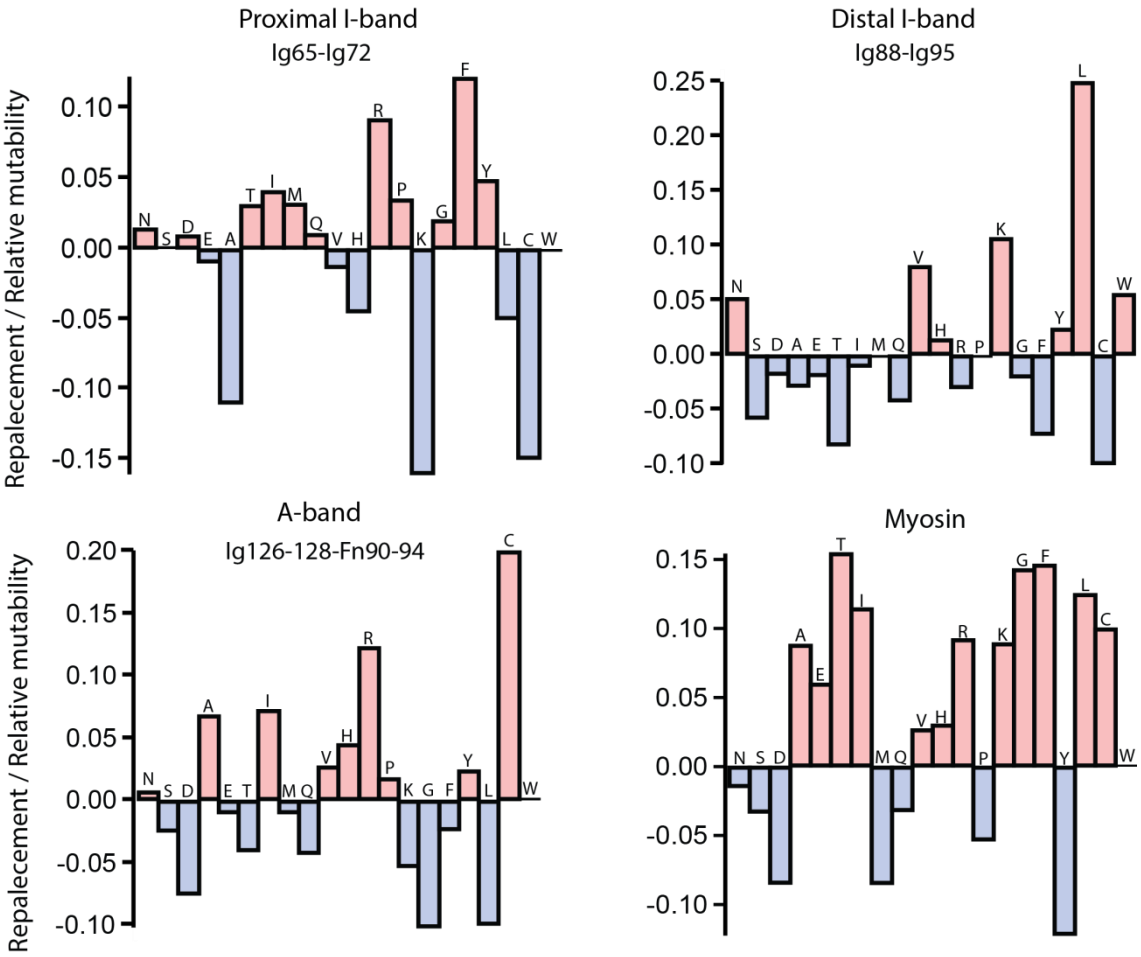
Supplementary Figure 3. Posterior probability distribution for each inferred residue of all ancestral titin fragments. The residue with the highest posterior probability is assigned at each position.



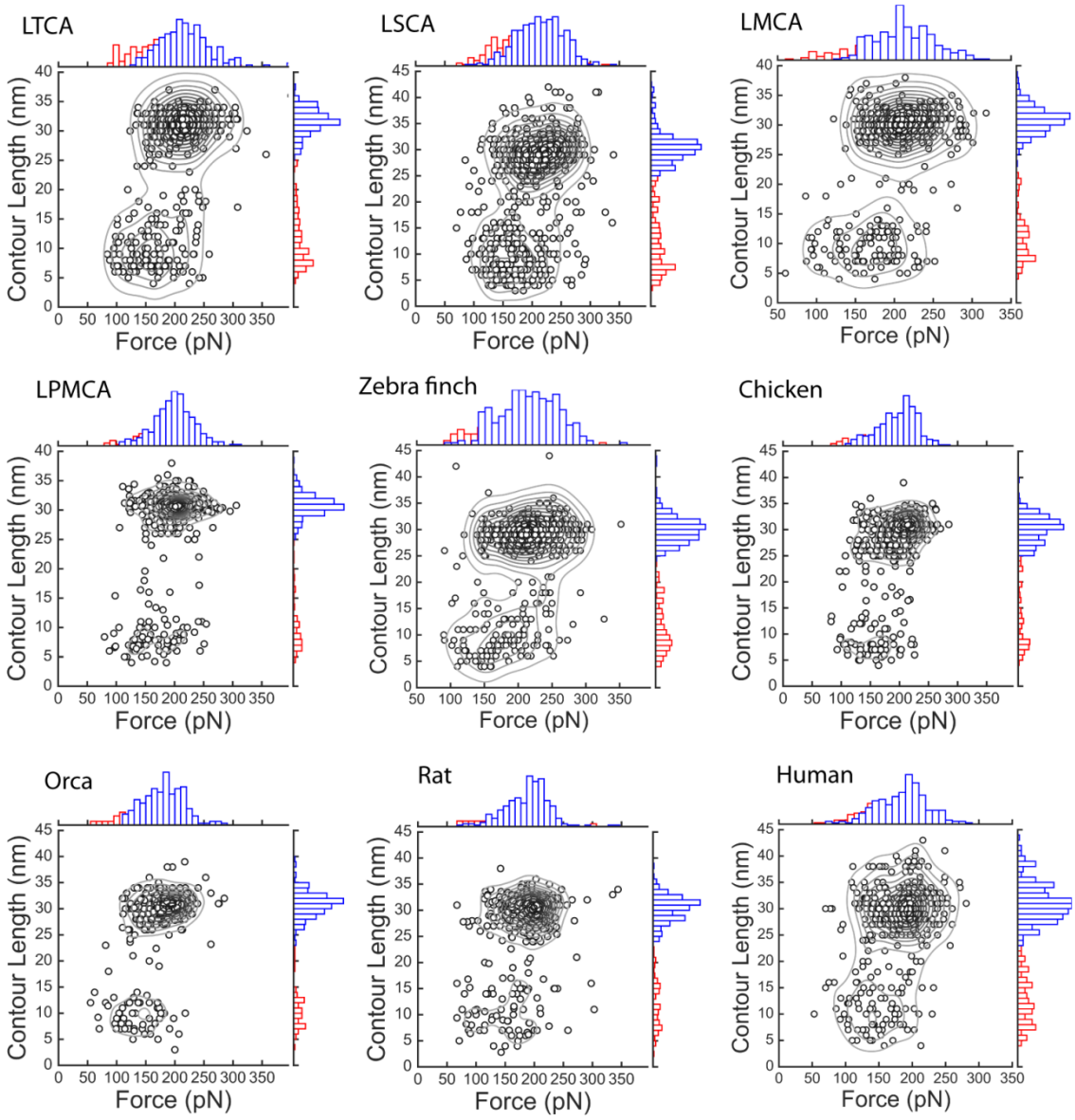
Supplementary Figure4. Phylogenetic tree used for the reconstruction of the ancestral myosin. A total of 26 sequences were used. The tree was built by Bayesian inference using MCMC. The ancestors of interest are displayed with colored dots. We analyzed the sequences from LTCA (Last Tetrapod Common Ancestor), LSCA (Last Sauropod Common Ancestor) and LMCA (Last Mammal Common Ancestor) myosin molecules. Divergence times for ancestral nodes were collected from different sources [1,2]. Posterior probabilities for each bifurcation are indicated.



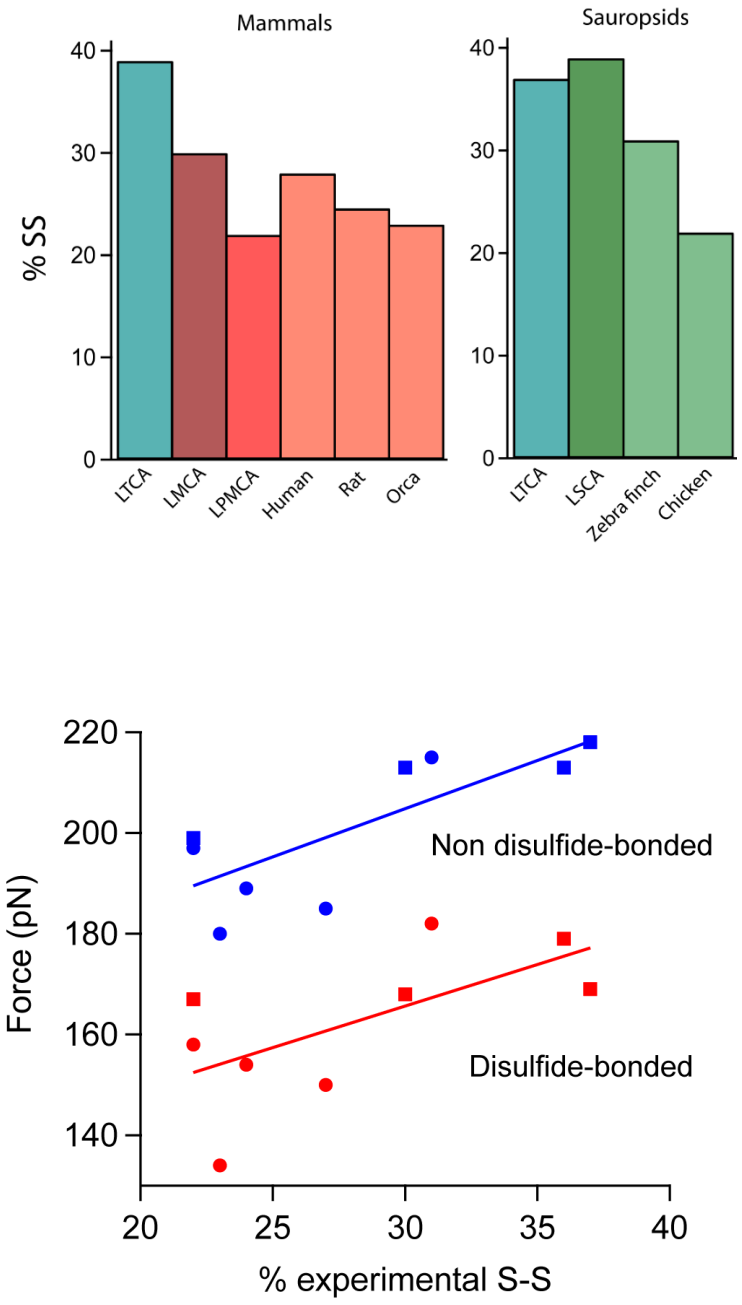
Supplementary Figure 5. Titin and myosin mutation rates between ancestral and extant species. Mutation rates for titin and myosin are estimated as the number of mutations from the ancestral forms to their modern counterparts per 100 residues and per Myr. Overall, mutation rate is double in titin (grey bars) with respect to myosin (white bars).



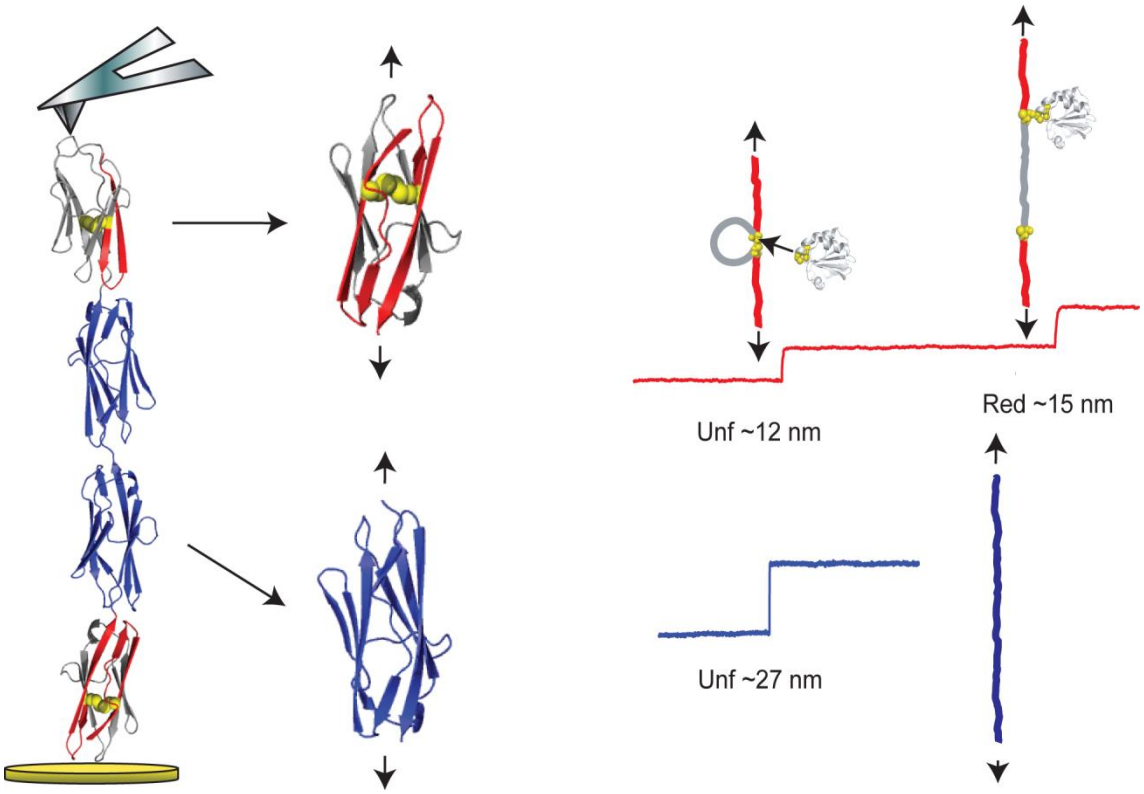
Supplementary Figure 6. Analysis of residue replacement and mutability for LTCA and human titin and myosin. For each residue type we estimate the occurrence in the different titin fragment as well as myosin for human and LTCA titin. The difference is the replacement for each residue type in the transition between LTCA and human titin. We thus estimate the ratio between replacement and relative mutability considering Ala as reference with a value of 100 [3]. In the plot this ratio is represented for each residue. A positive value (red bars) indicate an increasing number of the specific amino acid in the transition from LTCA to human. The blue bars represent a decreasing number of residues. Cysteine residues have decreased their representation in the I-band of human titin much more prominently than any other residue. In the contrary, these residues have increasedtheir number in the A-band of titin and myosin.



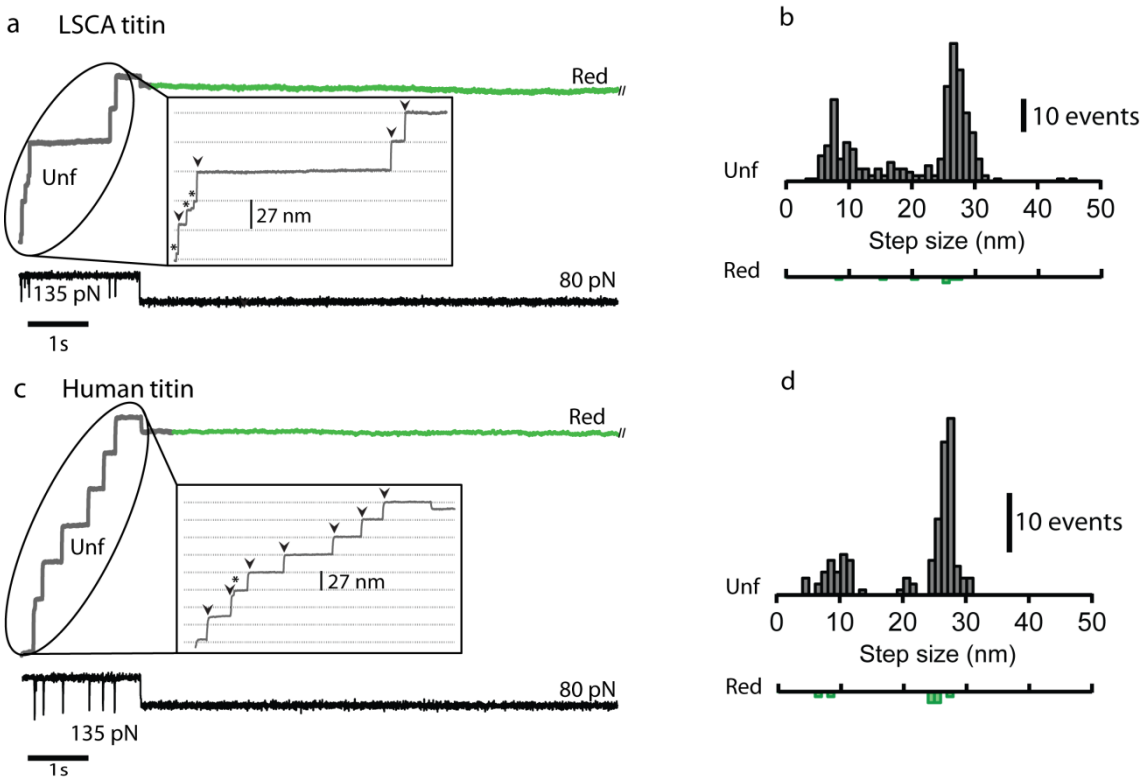
Supplementary Figure 7. Scatter plots of contour lengths and unfolding forces for all the titin fragments studied, with kernel density estimates shown as lines. The histograms are shown on each axis. Data collection for each protein is n=374 for LTCA, n=614 for LSCA, n= 407 for LMCA, n= 366 for LPMCA, n= 409 for zebra finch, n= 375 for chicken, n= 263 for orca, n= 341 for rat and n= 347 for human titin fragment. Blue bars indicate force and length of domains with no disulfide bonds. Red bars represent domains with disulfide bridges.



Supplementary Figure 8. Percentage of disulfide bonded domains detected in force-extension traces. LTCA and LSCA titin fragments are the ones presenting higher percentage of disulfide bonds in both, mammals and birds. In general, modern animals display fewer disulfide bonds than their ancestors. In the lower plot, we represent unfolding force versus observed disulfide bond percentage. Domains that do not contain disulfide bonds and are thus fully extended are represented in blue, whereas those showing disulfide bonds are represented in red and display lower mechanical resistance. Ancestral and extant species are shown with square and circles, respectively. Unfolding forces in both types of domains increase when the percentage of experimental disulfide bonds is higher, following a linear relationship.



Supplementary Figure 9. Schematics of force-clamp experiment for detection of single disulfide reduction events. Disulfide bonded domains (red) show a 2 two-step unfolding pattern. The first step (~12 nm) corresponds to the unfolding of the beta sheets that are not trapped in the disulfide bond while the second step (~15 nm) shows the unfolding of the rest of the protein after the reduction of the disulfide bond cause bythioredoxin. Not disulfide bonded domains have a single-step (~27 nm) unfolding pattern that represents the stretching of the whole domain.



Supplementary Figure 10. Force-clamp experiment for detection of single disulfide reduction events in the absence of thioredoxin enzymes. (a) Experimental trace of LSCA titin with seven steps representing each immunoglobulin domain visualized within the first pulse at 135 pN (grey). The fully unfolded domains are marked with an arrow (~27 nm) whereas disulfide bonded domains are marked with an asterisk (~5-20 nm). No steps are detected in the second stage at 80 pN (green) that was kept for 20 s. **(b)** Two populations of steps can be observed in the histogram. The histogram below shows step captured in the 80 pN pulse where reduction are generally observed in the presence of Trx (Fig. 3 in main text). **(c)** and **(d)** Experimental trace and steps histograms for human titin, respectively. In this case eight unfolding events are shown from which only one shows step size corresponding to a disulfide bonded domain. In the traces from human titin is common to observe 1 or 2 disulfide bonded domains. In both cases the histograms of the unfolding pulse resemble the histograms of the force-extension experiments, as expected.

a. Titin

	Human	Orca	Rat	Zebra finch	Chicken	LTCA	LSCA	LMCA	LPMCA
Human		92	91	73	74	78	80	90	96
Orca	92		88	72	73	77	79	91	93
Rat	91	88		72	73	77	78	86	92
Zebra finch	73	72	72		91	76	86	77	74
Chicken	74	73	73	91		77	87	78	75
LTCA	78	77	77	76	77		86	84	80
LSCA	80	79	78	86	87	86		87	82
LMCA	90	91	86	77	78	84	87		93
LPMCA	96	93	92	74	75	80	92	93	

b. Myosin

	Human	Orca	Rat	Zebra finch	Chicken	LTCA	LSCA	LMCA	LPMCA
Human		-	94	91	92	91	93	95	97
Orca	-		-	-	-	-	-	-	-
Rat	94	-		92	92	92	94	96	96
Zebra finch	91	-	92		97	87	96	94	93
Chicken	92	-	92	97		92	95	94	93
LTCA	91	-	92	87	92		90	89	88
LSCA	93	-	94	96	95	90		96	95
LMCA	95	-	96	94	94	89	96		97
LPMCA	97	-	96	93	93	88	95	97	

c. Actin

	Human	Orca	Rat	Zebra finch	Chicken	LTCA	LSCA	LMCA	LPMCA
Human		-	100	99	100	*	*	*	*
Orca	-		-	-	-	*	*	*	*
Rat	100	-		99	100	*	*	*	*
Zebra finch	99	-	99		99	*	*	*	*
Chicken	100	-	100	99		*	*	*	*
LTCA	*	*	*	*	*		*	*	*
LSCA	*	*	*	*	*	*		*	*
LMCA	*	*	*	*	*	*	*		*
LPMCA	*	*	*	*	*	*	*	*	

Supplemetary Table 1.Titin, myosin and actin sequence identities (%). (a) Titin has lowest identity with values that vary from 72% (zebra finch- rat) to 92% (human-orca whale). **(b)** Myosin identities fluctuate between 87 and 96% whereas actin **(c)** has the highest identity (99-100%). This phenomenon suggests that titin has driven the molecular evolution of muscle sarcomere much more extensively than the other two proteins. Orca myosin and actin sequences are not available at the time of this study. (*) Ancestral reconstruction of actin could not be performed due to the high identity of the sequences.

I65

Species	N aa	L full domain (nm)	L_{S-S} [L_{red}] (nm)		
			23-74	23-85	74-85
LTCA	89	26	9 [17]	5 [21]	23 [3]
LSCA	89	26	9 [17]	-	-
LMCA	89	26	9 [17]	-	-
LPMCA	89	26	9 [17]	-	-
Zebra finch	89	26	9 [17]	-	-
Human	89	26	9 [17]	-	-
Orca	89	26	9 [17]	-	-
Rat	89	26	9 [17]	-	-
Chicken	89	26	9 [17]	-	-

I66

Species	N aa	L full domain (nm)	L_{S-S} [L_{red}] (nm)		
			22-73	22-84	73-84
LTCA	88	26	9 [17]	5 [21]	23 [3]
LSCA	88	26	9 [17]	5 [21]	23 [3]
LMCA	88	26	9 [17]	5 [21]	23 [3]
LPMCA	88	26	9 [17]	5 [21]	23 [3]
Zebra finch	88	26	9 [17]	5 [21]	23 [3]
Human	88	26	9 [17]	5 [21]	23 [3]
Orca	88	26	9 [17]	5 [21]	23 [3]
Rat	88	26	9 [17]	5 [21]	23 [3]
Chicken	88	26	9 [17]	5 [21]	23 [3]

I67

Species	N aa	L full domain (nm)	L_{S-S} [L_{red}] (nm)
			74-85
LTCA	89	26	23 [3]
LSCA	89	26	23 [3]
LMCA	89	26	23 [3]
LPMCA	89	26	23 [3]
Zebra finch	89	26	23 [3]
Human	89	26	23 [3]
Orca	89	26	23 [3]
Rat	89	26	23 [3]
Chicken	89	26	23 [3]

I68

Species	N aa	L full domain (nm)	L_{S-S} [L_{red}] (nm)		
			19-85	23-85	19-23
LTCA	91	27	5 [22]	6 [21]	26 [1]
LSCA	91	27	5 [22]	6 [21]	26 [1]
LMCA	91	27	5 [22]	6 [21]	26 [1]
LPMCA	91	27	5 [22]	6 [21]	26 [1]
Zebra finch	91	27	5 [22]	6 [21]	26 [1]
Human	91	27	5 [22]	6 [21]	26 [1]
Orca	91	27	5 [22]	6 [21]	26 [1]
Rat	91	27	5 [22]	6 [21]	26 [1]
Chicken	91	27	5 [22]	6 [21]	26 [1]

I69

Species	N aa	L full domain (nm)	L_{S-S} [L_{red}] (nm)					
			22-73	22-84	73-84	22-56	56-73	56-84
LTCA	88	26	9 [17]	5 [21]	23 [3]	14 [12]	20 [6]	17 [9]
LSCA	88	26	9 [17]	5 [21]	23 [3]	14 [12]	20 [6]	17 [9]
LMCA	88	26	9 [17]	5 [21]	23 [3]	14 [12]	20 [6]	17 [9]
LPMCA	88	26	9 [17]	5 [21]	23 [3]	14 [12]	20 [6]	17 [9]
Zebra finch	88	26	9 [17]	5 [21]	23 [3]	14 [12]	20 [6]	17 [9]
Human	88	26	9 [17]	5 [21]	23 [3]	14 [12]	20 [6]	17 [9]
Orca	88	26	9 [17]	5 [21]	23 [3]	14 [12]	20 [6]	17 [9]
Rat	88	26	9 [17]	5 [21]	23 [3]	14 [12]	20 [6]	17 [9]
Chicken	88	26	9 [17]	5 [21]	23 [3]	14 [12]	20 [6]	17 [9]

I70

Species	N aa	L full domain (nm)	L_{S-S} [L_{red}] (nm)		
			47-73	47-84	73-84
LTCA	88	26	17 [9]	13 [13]	23 [3]
LSCA	88	26	17 [9]	13 [13]	23 [3]
LMCA	88	26	17 [9]	13 [13]	23 [3]
LPMCA	88	26	17 [9]	13 [13]	23 [3]
Zebra finch	88	26	17 [9]	13 [13]	23 [3]
Human	88	26	17 [9]	13 [13]	23 [3]
Orca	88	26	17 [9]	13 [13]	23 [3]
Rat	88	26	17 [9]	13 [13]	23 [3]
Chicken	88	26	17 [9]	13 [13]	23 [3]

I71

Species	N aa	L full domain (nm)	L_{S-S} [L_{red}] (nm)								
			48-74	48-85	74-85	23-48	23-74	23-85	23-57	48-57	57-74	57-85
LTCA	89	26	17 [9]	14 [12]	23 [3]	17 [9]	9 [17]	5 [21]	15 [11]	23 [3]	20 [6]	17 [19]
LSCA	89	26	17 [9]	14 [12]	23 [3]	17 [9]	9 [17]	5 [21]	-	-	-	-
LMCA	89	26	17 [9]	14 [12]	23 [3]	-	-	-	-	-	-	-
LPMCA	89	26	17 [9]	14 [12]	23 [3]	-	-	-	-	-	-	-
Zebra finch	89	26	17 [9]	14 [12]	23 [3]	18 [9]	9 [17]	5 [21]	-	-	-	-
Human	89	26	17 [9]	14 [12]	23 [3]	-	-	-	-	-	-	-
Orca	89	26	17 [9]	14 [12]	23 [3]	-	-	-	-	-	-	-
Rat	89	26	17 [9]	14 [12]	23 [3]	-	-	-	-	-	-	-
Chicken	89	26	17 [9]	14 [12]	23 [3]	17 [9]	9 [17]	5 [21]	-	-	-	-

I72

Species	N aa	L full domain (nm)	L_{S-S} [L_{red}] (nm)		
			22-73	22-47	47-73
LTCA	88	26	9 [17]	-	-
LSCA	88	26	9 [17]	-	-
LMCA	88	26	9 [17]	-	-
LPMCA	88	26	9 [17]	-	-
Zebra finch	88	26	9 [17]	-	-
Human	88	26	9 [17]	-	-
Orca	88	26	9 [17]	-	-
Rat	88	26	9 [17]	17 [9]	17 [9]
Chicken	88	26	9 [17]	-	-

Supplemetary Table 2. Estimation of force-clamp lengths for unfolding and reductions of the titin domains I65 to I72. The columns indicate species, number of residues of the domain, expected length of full domain and possible pairs of cysteines forming a disulfide bond. The length of the unfolding up to the disulfide bond is indicated as L_{S-S}, and the reduction associated to the particular disulfide bond as L_{red} in square brackets. Lengths have been estimated as reported elsewhere [4].

List of titin proteins from the species used in the construction of phylogenetic trees

The sequences were retrieved from the Uniprot and GenomeNET databases using BLAST. The asterisks indicate sequences that were updated or added during the course of the study and considered for the tree in Supplementary Figure 2:

B0S758	*Danio rerio* (Zebrafish)
gi 348541917	*Oreochromis niloticus* (Tilapia)
F7EAV6	*Xenopus tropicalis* (Western clawed frog)
gi 103063864	*Phyton bivittatus* (Burmese python*)
gi 100560476	*Anolis carolinensis* (American chameleon*)
H9GP88	*Anolis carolinensis* (American chameleon)
K7G060	*Pelodiscus sinensis* (Chinese softshell turtle)
gi 102559759	*Alligator mississippiensis* (American alligator)
gi 465955284	*Chelonia mydas* (Green sea turtle)
gi 449507164	*Taeniopygia guttata* (Zebrafinch)
gi 483520158	*Anas platyrhynchos* (Wild duck)
G1NAX9	*Meleagris gallopavo* (Wild turkey)
gi 100546807	*Meleagris gallopavo* (Wild turkey*)
gi 101921795	*Falco peregrines* (Peregrine falcon*)
gi 101819370	*Ficedula albicollis* (Collared flycatcher*)
gi 363735918	*Gallus gallus* (Chicken)
F6VRV4	*Ornithorhynchus anatinus* (Platypus)
gi 103171044	*Ornithorhynchus anatinus* (Platypus*)
gi 395519871	*Sarcophilus harrisii* (Tasmanian devil)
G3UK67	*Loxodonta africana* (African elephant)
gi 471370017	*Trichechus manatus* (West Indian manatee)
L5K2L4	*Pteropus alecto* (Black flying fox)
G1P5X9	*Myotis lucifugus* (Little brown bat)
M3WG03	*Felis catus* (Cat)
gi 359323893	*Canis lupus* (Dog)
G1L1P3	*Ailuropoda melanoleuca* (Giant panda)
F6VG02	*Equus caballus* (Horse)
gi 465995183	*Orcinus orca* (Orca)
F1N757	*Bos taurus* (Cattle)
gi 426220782	*Ovis aries* (Sheep)
G1U9S3	*Oryctolagus cuniculus* (Rabbit)
G3HAC6	*Cricetulusgriseus* (Chinese hamster)
A2ASS6	*Mus musculus* (Mouse)
gi 392339498	*Rattus norvegicus* (Brown Rat)
L9KLA3	*Tupaia belangeri chinensis* (Chinese tree shrew)
F7IGY9	*Callithrix jacchus* (Marmoset)
H2P803	*Pongo abelii* (Orangutan)
G3QYH8	*Gorilla gorilla* (Gorilla)
H2QJ24	*Pan troglodytes* (Chimpanzee)
D3DPG0	*Homo sapiens* (Human)

Sequences of modern and ancestral titin fragments studied

The nomenclature of the fragment corresponds to the canonical human titin with Uniprot code Q8W242 from residues 8232 to 8976, which correspond to domains I65 to I72. The equivalent domains are used for the rest of the species and ancestral proteins.

Human

EPPRFIKKLEPSRIVKQDEFTRYECKIGGSPEIKVLWYKDETEIQESSKFRMSFVDSVAVL
EMHNLSVEDSGDYTCEAHNAAGSASSSTSLKVKEPPIFRKKPHPIETLKGADVHLECEL
QGTPPFHVSWYKDKRELRSGKKYKIMSENFLTSIHILNVDAADIGEYQCKATNDVGSD
TCVGSIALKAPPRFVKKLSDISTVVGKEVQLQTTIEGAEPISVVWFKDKGEIVRESDNIWI
SYSENIATLQFSRVEPANAGKYTCQIKNDAGMQECFATLSVLEPATIVEKPESIKVTTGD
TCTLECTVAGTPELSTKWFKDGKELTSDNKYKISFFNKVSGLKIINVAPSDSGVYSFEVQ
NPVGKDSCTASLQVSDRTVPPSFTRKLKETNGLSGSSVVMECKVYGSPPISVSWFHEGN
EISSGRKYQTTLTDNTCALTVNMLEESDSGDYTCIATNMAGSDECSAPLTVREPPSFVQ
KPDPMDVLTGTNVTFTSIVKGTPPFSVSWFKGSSELVPGDRCNVSLEDSVAELELFDVD
TSQSGEYTCIVSNEAGKASCTTHLYIKAPAKFVKRLNDYSIEKGKPLILEGTFTGTPPISV
TWKKNGINVTPSQRCNITTTEKSAILEIPSSTVEDAGQYNCYIENASGKDSCSAQILILEP
PYFVKQLEPVKVSVGDSASLQCQLAGTPEIGVSWYKGDTKLRPTTTYKMHFRNNVATL
VFNQVDINDSGEYICKAENSVGEVSASTFLT

Rat

EPPRFIKKLDQSRIVKQDEYTRYECKIGGSPEIKVLWYKDEVEIQESSKFRMSFEDSVAIL
EMHNLSVEDSGDYTCEARNAAGSASSSTSLKVKEPPVFRKKPFPVETLKGADVHLECE
LQGTPPFQVSWYKDKRELRSGKKYKIMSENLLTSIHILNVDTADIGEYQCKATNDVGSD
TCVGSVTLKAPPQFVKKLSDVSTIIGKEVQLQTTIEGAEPISVAWFKDKGEIVRESDNIWI
SHSENVATLHFSRAEPANAGKYTCQIKNDAGVQECYATLSVLEPATIVEKPESIKVTTG
DTCTLECMVSGTPELSTKWFKDGKELTGDSKYKISFFNKVSGLKIISVAPGDSGVYSFE
VQNPVGKDSCTVSIQVSDRIIPPSFTRKLKETNGLSGSSVVMECKVFGSPPISVLWLHDG
NAISSGRKYQTTLTDNTCALTVNMLEDADAGDYTCIATNVAGSDECSAPLTVREPPSFV
QKPDPMDVLTGSNVTFTSIVKGTPPFTVSWFKGSSELVPGARCNVSLQDSVAELELFDV
DTSQSGDYTCIVSNEAGRASCTTQLFVKAPAIFVKRLNDYSIEKGKPLILEGTFSGTPPIS
VTWKKNGVNVTASQRCNITTTEKSAILEILSSTVEDSGQYNCYIENASGKDSCSAQILIL
EPPYFVKQLEPLKVTVGDSASLQCQLAGTPEIGVSWYKGDTKLRPTTTCKMHFKNNVA
TLVFTQVDSNDSGEYICRAENSVGEVSSSTFLT

Orca

EPPRFIKKLEPSRIMKQGESTRYECKVGGSPEIKVLWYKDETEIQESSKFRMSFHDSVAV
LEMHALSVEDSGDYTCEARNAAGRASSSTTLKVKEPPVFRKKPRPVETLEGADVHLEC
ELQGTPPFQVSWHKDKRELRSGKKYKIMSENFLTSIHILSVSAADVGEYQCKATNDVG
GDTCVGSITLKAPPRFVKKPSDISAIVGEEVRLQAAIEGTEPISMVWFKDKGEMVRESD
NIWISYSENIATLQFSRVETANAGKYTCQIKNDAGMQECFATVSILEPAAIVEKPESIKVT
TGDTCTLECMVTGTPELTTKWFKDAKELTSDSKYKISFFNKISGLKIINVAPSDSGVYSF
EVQNPVGKDSCTASVHVSDRIVPPSFTRKLKETNGLSGSSVVMECKVYGSPPISVSWFH
EGNEISSGRKYQTTLTDNTCALTVNMLEDSDTGDYTCIATNVAGSDECSAPLTVREPPS
FVQKPDPMDVLTGANVTFTSLVKGTPPFSVSWFKGSSELVPGDRCNVSLEDSVAELELF

DVDTSQSGEYTCIVSNEAGKASCTTHLYVKAPAKFVKKLNDYSLEKGKPLILEATYTGT
LPISVTWKKNGTNITPSQRCHITTTEKSAILEIPSSTVEDAGQYNCYIENASGKDSCSAQI
LILEPPYFVKQLEPVKVTIGDSASLQCQLAGTPEIGVSWYKGDTKLRPTTTYKMHFKNN
VATLVFNQVDSNDSGEYICRAENSVGEVSSSTFLT

Zebra Finch

EPPRFIKKLDSSKLVKQHDSTTYECKIGGSPEIKVTWYKGETEIHPSEKYRMSFVDSVAV
IEMHNLSVEDSGDYTCEAQNPAGSASTSTTLRVKAPPIFTKKPHPVETLKGSDIHLECKL
QGTPPFQISWYKDKREIRSSKKYKVMSENYIASLHILSVDTADVGEYHCKAVNDVGSD
SCIGSVTLRAPPTFVKKLSDLTVVVGESIELQAAVQGSQPISVLWLKDKGEIIRESDNLWI
SYSENIATMQIGNAEPTNAGKYICQIKNDAGIQECFAMLKVLEPAVIVEKPGPVKVTAG
DSCTLECTVDGTPELTARWFKDGNELSTDHKYKISFFNKVSGLKILNATLEDSGEYTFE
VKNSVGKSSCTASVHVSDRIIPPSFTRKLKETYGQLGSSAVLECKVYGSPPILVSWFHDG
QEITSGEKYQATLTDNTCSLKVNGLQESDMGTYLCTATNAAGSDECSAFLSVIESPSFV
KKPEPLDVLSGANITFTSIVKGSPPLEVKWFRGSVELVPGPRCNITLQDSVAELELFDVH
PLESGDYTCQVSNEAGKISCTTHLFVKEPAKFVKKVNDLSVEKGKNLILECTYMGTPPI
SVTWKKNGVKIMHSEKCSITTTDTSAILEIPNSKLEDQGQYSCHIENDSGKDTCHGTITIL
EPPYFIRSLEPVQVTVGDSASLQCQVAGTPEMIVSWYKGDTKLRGTATMKMHFRNQIA
TLVFSQVDGSDSGEYICKVENSVGEASSSSLLT

Chicken

EPPRFIKKLDSSRLVKQHDSTRYECKVGGSPEIKVTWYKGETEIHPSEKYSMSFVDSVA
VLEMHNLSVEDSGDYSCEAQNPAGSASTSTSLKVKAPPAFTKKPHPVQTLKGSDVHLE
CELQGTPPFQISWYKDKREIRSSKKYKVMSENYLASIHILNVDTADVGEYHCKAVNDV
GSDSCIGSVTLRAPPTFVKKLSDVTVVVGETIELQAAVEGAQPISVLWLKDKGEIIRESE
NLWISYSENVASLKIGNAEPTNAGKYICQIKNDAGFQECFAKLTVLEPAVIVEKPGPVK
VTAGDSCTLECTVDGTPELTARWFKDGNELSTDHKYKISFFNKVSGLKILNAGLEDSGE
YTFEVKNSVGKSSCTASLQVSDRIMPPSFTRKLKETYGQLGSSAVLECKVYGSPPILVS
WFHDGQEITSGDKYQATLTDNTCSLKVNGLQESDMGTYSCTATNVAGSDECSAFLSVR
EPPSFVKKPEPFNVLSGENITFTSIVKGSPPLEVKWFRGSIELAPGHKCNITLQDSVAELE
LFDVQPLQSGDYTCQVSNEAGKISCTTHLFVKEPAKFVMKVNDLSVEKGKNLILECTY
TGTPPISVTWKKNGVILKHSEKCSITTTETSAILEIPNSKLEDQGQYSCHIENDSGQDNCH
GAITILEPPYFVTPLEPVQVTVGDSASLQCQVAGTPEMIVSWYKGDTKLRGTATVKMH
FKNQVATLVFSQVDSDDSGEYICKVENTVGEATSSSLLT

LTCA

EPPRFVKKLESSKVVKQGDSTRFECKISGSPEIRVLWYKNDAEIQHGGKYRMSFVDSVA
VLEISNASVEDSGDYTCEAHNDAGSASCSTSLKVKEPPVFIKKPHPVETLKGSDVSLQCE
LKGTPPFQVSWYKDKREIKSSKKYKIMSENYLASIHILKVDAADIGEYQCKAVNDVGS
DTCLGSIKLKEPPRFVKKLSDASAVVGEPVELQATVEGAQPISVTWLKDKEEIVRESENI
WISFSDNVATLQFLNAEPANAGKYTCQIKNDAGVQECFATLSVLEPAVIVEKPESMKVT
SGDTCTLECTVSGTPELSAKWFKDGKELSSDHKYKISFHNKVSGLKILNAAPNDSGEYT
FEVKNNVGKDSCAMSVQVSDRIIPPSFTRKLKETHGLLGSSVVLECKVSGSPPISVSWFH
NGNEITSGGKYQATLTDNTCTLTVSALETSDAGKYSCTATNVAGSDECSAALTVKEPPS
FVEKPEPLEVLPGATVTFTAIIKGTPPFKVKWFRGSTELVPGRRCNISLEDSVAVLELYN

VDTSQSGDYTCQITNDAGKDSCTTHLFVKEPAKFVKKLNDYSIEKGKPLILECTYTGTP
PISVTWKKNGVEITQSEKCSITTTEKSCILEIPNSTMEDAGQYTCHVENASGHDTCQATIS
ILEPPYFIKPLEPVEVTAGDAASLQCQIAGTPEIKVSWYKGDTKLRATATSKMHFKNNV
ATLVFAQVDSNDSGEYICKAENSVGEASSSTSLT

LSCA

EPPRFIKKLDSSRVVKQHDSTRYECKIGGSPEIKVIWYKNETEIHPSSKYRMSFVDSVAVI
EMHNLSVEDSGDYTCEAQNAAGSASSSTSLKVKEPPVFSKKPHPVETLKGSDVHLECE
LKGTPPFQVSWYKDKREIRSSKKYKIMSENYLASIHILNVDAADIGEYHCKAVNDVGSD
TCIGSITLKAPPRFVKKLSDLTAVVGEPVELQATVEGAQPISVLWLKDKGEIVRESDNL
WISYSENVATLQIGNAEPANAGKYTCQIKNDAGVQECFATLSVLEPAVIVEKPGPMKV
TSGDSCTLECTVAGTPELTARWFKDGNELSTDHKYKISFFNKVSGLKILNAGPEDSGEY
TFEVKNSVGKSSCTASVHVSDRIIPPSFTRKLKETHGLLGSSVVLECKVYGSPPISVSWF
HDGQEITSGDKYQATLTDNTCSLTVNALEESDAGNYSCTATNVAGSDECSAYLTVREP
PSFVKKPEPLQVLSGANITFTSIIKGTPPFDVKWFRGSVELVPGHRCNISLEDSVAELELF
DVHPLQSGDYTCLVTNEAGKISCTTHLFVKEPAKFVKKLNDFSVEKGKPLILECTYTGT
PPISVTWKKNGVKITQSEKCSITTTETSAILEIPNSKMEDAGQYTCHIENDSGQDNCHATI
SILEPPYFVRPLEPVQVTVGDSASLQCQVAGTPEIIVSWYKGDTKLRATATSKMHFKNN
VATLVFNQVDSNDSGEYICKAENSVGEASSSALLS

LMCA

EPPRFIKKLESSRVVKQHDSTRYECKIGGSPEIKVLWYKNETEIQSSSKFRMSFVDSVAVI
EMHNLSVEDSGDYTCEAHNAAGSASSSTSLKVKEPPVFSKKPHPVETLKGSDVHLECE
LRGTPPFQVSWHKDKREIRSGKKYKIMSENFLTSIHILNVDASDIGEYQCKAVNDVGSD
TCVGSITLKAPPRFVKKLSDLSAVVGDQVQLQATIEGAEPISVVWFKDKGEIVRESDNI
WISYSENVATLQFANAEPANAGKYTCQIKNDAGMQECFATLSVLEPAVIVEKPESMKV
TSGDTCTLECTVAGTPELSTKWFKDGKELTSDSKYKISFFNKVSGLKIINVAPNDSGVYT
FEVQNSVGKDSCTASVQVSDRIVPPSFTRKLKETNGLFGSSVVLECKVYGSPPISVSWFH
EGNEITSGRKYQATLTDNTCSLTVNALEESDAGDYTCVATNVAGSDECSAALTVREPPS
FVQKPDPLDVLTGTNVTFTSIIKGTPPFSVSWFKGSSELVPGDRCNISLDDSVAELELFDV
DTSQSGDYTCVVTNEAGKASCTTHLYVKAPAKFVKKLNDYSIEKGKPLILEGTYTGTPP
ISVTWKKNGLNITPSQKCSITTTEKSAILEIPSSTVEDAGQYTCYIENASGKDSCHAQILIL
EPPYFVKQLEPVKVTVGDSASLQCQVAGTPEIAVSWYKGDTKLRATATSKMHFRNNV
ATLVFNQVDSNDSGEYICKAENSVGEVSSSTFLT

LPMCA

EPPRFIKKLEPSRIVKQDEYTRYECKIGGSPEIKVLWYKDETEIQESSKFRMSFVDSVAVL
EMHNLSVEDSGDYTCEAHNAAGSASSSTSLKVKEPPIFRKKPHPVETLKGADVHLECEL
QGTPPFQVSWHKDKRELRSGKKYKIMSENFLTSIHILNVDAADIGEYQCKATNDVGSD
TCVGSITLKAPPRFVKKLSDISTIVGEEVQLQTTIEGAEPISVVWFKDKGEIVRESDNIWIS
YSENIATLQFSRAEPANAGKYTCQIKNDAGMQECFATLSVLEPAAIVEKPESIKVTSGDT
CTLECTVTGTPELSTKWFKDGKELTSDSKYKISFFNKVSGLKIINVAPNDSGVYSFEVQN
PVGKDSCTASVQVSDRIVPPSFTRKLKETNGLSGSSVVMECKVYGSPPISVSWFHEGNEI
SSGRKYQTTLTDNTCALTVNMLEESDAGDYTCVATNVAGSDECSAPLTVREPPSFVQK
PDPMDVLTGTNVTFTSIIKGTPPFSVSWFKGSSELVPGDRCNVSLEDSVAELELFDVDTS
QSGEYTCIVSNEAGKASCTTHLYVKAPAKFVKRLNDYSIEKGKPLILEGTYTGTPPISVT

WKKNGINITPSQRCNITTTEKSAILEIPSSTVEDAGQYNCYIENASGKDSCSAQILILEPPY
FVKQLEPVKVTVGDSASLQCQLAGTPEIAVSWYKGDTKLRPTATYKMHFRNNVATLV
FNQVDSNDSGEYICRAENSVGEVSSSTFLT

References

1 Hedges, S. B., Marin, J., Suleski, M., Paymer, M. & Kumar, S. Tree of life reveals clock-like speciation and diversification. *Mol Biol Evol* **32**, 835-845, doi:10.1093/molbev/msv037 (2015).

2 Benton, M. J. *et al.* Constraints on the timescale of animal evolutionary history. *Palaeontologia Electronica* **18**, 1-106 (2015).

3 Dayhoff M.O., S. R. M., Orcutt B.C. *Amino acid scale: Relative mutability of amino acids (Ala=100)*. Vol. 5 (National Biomedical Research Foundation, Silver Spring, 1978).

4 Alegre-Cebollada, J., Kosuri, P., Rivas-Pardo, J. A. & Fernandez, J. M. Direct observation of disulfide isomerization in a single protein. *Nat Chem* **3**, 882-887, doi:10.1038/nchem.1155 (2011).