

**ARTICLE**

# The number of Z-repeats and super-repeats in nebulin greatly varies across vertebrates and scales with animal size

Jochen Gohlke<sup>1</sup>, Paola Tonino, Johan Lindqvist<sup>1</sup>, John E. Smith III<sup>1</sup>, and Henk Granzier<sup>1</sup>

Nebulin is a skeletal muscle protein that associates with the sarcomeric thin filaments and has functions in regulating the length of the thin filament and the structure of the Z-disk. Here we investigated the nebulin gene in 53 species of birds, fish, amphibians, reptiles, and mammals. In all species, nebulin has a similar domain composition that mostly consists of ~30-residue modules (or simple repeats), each containing an actin-binding site. All species have a large region where simple repeats are organized into seven-module super-repeats, each containing a tropomyosin binding site. The number of super-repeats shows high interspecies variation, ranging from 21 (zebrafish, hummingbird) to 31 (camel, chimpanzee), and, importantly, scales with body size. The higher number of super-repeats in large animals was shown to increase thin filament length, which is expected to increase the sarcomere length for optimal force production, increase the energy efficiency of isometric force production, and lower the shortening velocity of muscle. It has been known since the work of A.V. Hill in 1950 that as species increase in size, the shortening velocity of their muscle is reduced, and the present work shows that nebulin contributes to the mechanistic basis. Finally, we analyzed the differentially spliced simple repeats in nebulin's C terminus, whose inclusion correlates with the width of the Z-disk. The number of Z-repeats greatly varies (from 5 to 18) and correlates with the number of super-repeats. We propose that the resulting increase in the width of the Z-disk in large animals increases the number of contacts between nebulin and structural Z-disk proteins when the Z-disk is stressed for long durations.

## Introduction

Nebulin is a long slender protein that is expressed in skeletal muscle. It associates with the sarcomeric thin filament, with its C terminus anchored in the Z-disk and its N terminus positioned toward the thin filament's pointed end (Kruger et al., 1991; Labeit et al., 1991; Pappas et al., 2011). The giant size of nebulin (the human *NEB* gene has 183 exons that are predicted to encode a protein of ~900 kD; Labeit and Kolmerer, 1995; Donner et al., 2004) has made it challenging to study nebulin, but in recent years progress has accelerated due to the development of animal model systems that target the nebulin gene (Bang et al., 2006; Witt et al., 2006; Ottenheijm et al., 2013; Yamamoto et al., 2013; Li et al., 2019; Laitila et al., 2020; Lindqvist et al., 2020). The importance of nebulin is shown by the many *NEB* mutations that cause nemaline myopathy (NEM; Lehtokari et al., 2014), the most common nondystrophic congenital myopathy (Colombo et al., 2015). NEM results in impaired skeletal muscle function that affects functions such as speech, mobility, and respiration and can lead to early death (Laing and Wallgren-Pettersson,

2009; Wallgren-Pettersson et al., 2011; Lehtokari et al., 2014). NEM can be caused by mutations in multiple genes, but ~50% of patients have mutations in nebulin (Anderson et al., 2004; Laing and Wallgren-Pettersson, 2009; Wallgren-Pettersson et al., 2011; Lehtokari et al., 2014). The size of the nebulin gene is known to vary in the few species that have been studied experimentally so far (Labeit and Kolmerer, 1995; Kazmierski et al., 2003; Donner et al., 2004), most likely because of several independent duplication events of nebulin's super-repeats (Björklund et al., 2010). To gain more in-depth insights into nebulin's sequence variation during vertebrate evolution, and to take advantage of the surge in species that have recently been sequenced, we investigated in the present study the nebulin gene in multiple birds, fish, amphibians, reptiles, and mammals for a total of 53 species. First, we briefly review what is known about nebulin at the protein level.

The protein structure of nebulin is modular and consists of a large number of simple repeats: ~30–35 residue actin-binding

Department of Cellular and Molecular Medicine, University of Arizona, Tucson, AZ.

Correspondence to Henk Granzier: [granzier@arizona.edu](mailto:granzier@arizona.edu)

This work is part of a special collection on myofilament function and disease.

© 2020 Gohlke et al. This article is distributed under the terms of an Attribution–Noncommercial–Share Alike–No Mirror Sites license for the first six months after the publication date (see <http://www.rupress.org/terms/>). After six months it is available under a Creative Commons License (Attribution–Noncommercial–Share Alike 4.0 International license, as described at <https://creativecommons.org/licenses/by-nc-sa/4.0/>).

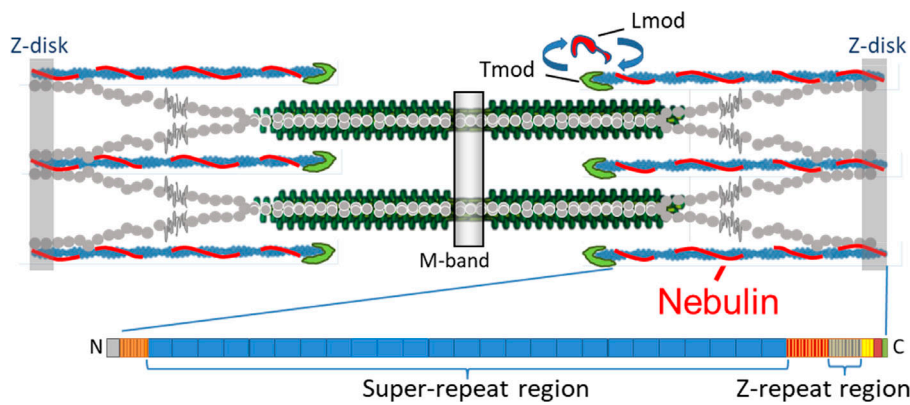


Figure 1. **Schematic of sarcomere.** Nebulin's C terminus is anchored in the Z-disk and its N terminus is near the thin filament pointed end. Shown is the domain structure with super-repeats (blue) and Z-repeats (gray), the 8 simple repeats at the N terminus (orange), 12 simple repeats at the C terminus, linker repeats (orange), the highly conserved triplicate (yellow), serine-rich region (brown), and SH3 at nebulin's C terminus (green).

modules, each containing the actin-binding motif SDxxYK (Fig. 1; Jin and Wang, 1991a, 1991b; Wang et al., 1996). The N terminus contains a unique E-rich segment followed by eight simple repeats (Kazmierski et al., 2003; Donner et al., 2004). These repeats might mediate protein interactions near the thin filament's pointed end (Labeit and Kolmerer, 1995). This is followed by a large region where simple repeats form seven-repeat-containing super-repeats, each with a single tropomyosin binding site, WLKGIGW (Stedman et al., 1988; Jin and Wang, 1991a, 1991b; Labeit and Kolmerer, 1995; Ogut et al., 2003; Donner et al., 2004). Although the individual super-repeats appear similar, recent work showed that their actin-binding affinity varies: super-repeats near both ends of the molecule bind significantly more strongly than the centrally located ones (Laitila et al., 2019). Weak binding of central super-repeats is consistent with structural studies that suggest that nebulin has multiple binding sites on F-actin (Lukoyanova et al., 2002), which indicates that nebulin might move on F-actin, analogous to tropomyosin, and that weak binding makes this possible.

Low-angle x-ray diffraction studies on intact muscle from WT and nebulin-deficient mice have revealed that nebulin's super-repeats stiffen the thin filament (Kiss et al., 2018). This nebulin-induced stiffening of the thin filament might alter thin filament activation and cross-bridge cycling kinetics (Kiss et al., 2018), which is functionally important. Its absence in skeletal muscle in nebulin-deficient mouse muscle (Witt et al., 2006; Bang et al., 2009; Ottenheijm et al., 2013; Li et al., 2015; Joureau et al., 2017; Kawai et al., 2018) and nebulin-based NEM (Ottenheijm et al., 2009, 2010; Lawlor et al., 2011) is likely to contribute to their deficit in force production. Using mouse models in which the number of super-repeats was varied, it was shown recently that nebulin's super-repeat region strictly regulates thin filament length (TFL) in fast muscle; in slow muscle, nebulin collaborates with Lmod2 (leiomodin-2), with Lmod2 regulating the length of a distal nebulin-free thin filament segment (Kiss et al., 2020).

Nebulin's C terminus is poorly understood but is also likely important, considering the many NEM patients with truncating mutations that result in the loss of nebulin's C-terminal domains (Lehtokari et al., 2006, 2014). Nebulin's C terminus is embedded in the Z-disk (Millevoi et al., 1998) and contains simple repeats that do not form super-repeats, followed by a serine-rich region and the terminal SH3 domain (Fig. 1; Kazmierski et al., 2003).

The initial simple repeats that follow the super-repeat region have properties distinct from regular simple repeats (they are somewhat larger than normal repeats, and several have a low isoelectric point), and they are also referred to as linker repeats (Labeit and Kolmerer, 1995). Linker repeats are followed by simple repeats, many of which are alternatively spliced (Millevoi et al., 1998). We refer to them as Z-repeats for their likely localization within the Z-disk region. A recent RNA sequencing study showed that Z-repeats are expressed at low and high levels in extensor digitorum longus (EDL) and soleus muscles of the mouse, respectively (Kiss et al., 2020). These muscle types are rich in type I (soleus) and IIB (EDL) fibers, implying a fiber type-dependent expression of these exons. This conclusion is consistent with an early quantitative PCR study on slow and fast muscle from human and rabbit (Millevoi et al., 1998). The Z-disk width is known to vary in different fiber types, with the widest Z-disk in type I fibers and the narrowest in type IIB fibers (Li et al., 2019). It is generally thought that this reflects the mechanical strength that is required of the Z-disks in the different fiber types (Vigoreaux, 1994; Luther, 2009; Knöll et al., 2011).

The mechanisms that regulate the Z-disk width are unknown. It has been speculated that titin regulates the Z-disk width because titin's N terminus spans the full Z-disk and contains differentially expressed Z-repeats that bind  $\alpha$ -actinin (Young et al., 1998). However, the number of titin Z-repeats and their spacing are incompatible with titin on its own performing this task (Gregorio et al., 1998; Luther and Squire, 2002). Instead, it has been suggested that nebulin contributes to the regulation of the Z-disk width (Witt et al., 2006; Li et al., 2015). This is supported by the correlation between the number of nebulin Z-repeats and the Z-disk width that has been found (Kiss et al., 2020) and the widened Z-disks that locally exist in nebulin-deficient muscle (Witt et al., 2006; Tonino et al., 2010; Li et al., 2015).

To gain more in-depth understanding of nebulin, we analyzed the nebulin gene in a wide range of species. Species were selected based on completeness of the nebulin genomic sequence while aiming for good representation across different animal classes. We found that the number of nebulin super-repeats and nebulin Z-repeats is highly variable, and that the number of super-repeats scales with animal size, with the largest effect in birds. We propose that the scaling between nebulin length and

animal size provides a functional advantage, as it is expected to enhance the efficiency of isometric force production and slow the speed of shortening in large animals. Finally, we focused on nebulin's Z-repeats and discovered a positive correlation between the number of Z-repeats and the number of super-repeats. We reason that this reflects a need for Z-disks to be wide in muscles with a low shortening velocity.

## Materials and methods

### Data

Nebulin gene sequences were downloaded from NCBI for all analyzed species. See Table S1 (Species\_references) for species names, gene IDs, and coordinates of the genomic sequences used. Most gene models are based on NCBI's eukaryotic gene prediction, which uses a combination of homology searching with ab initio modeling. Owing to alternative splicing events in the Z-repeat region and in super-repeat 12, there are often multiple annotated nebulin splice variants. Because gel electrophoresis and gene expression data from mice indicate that the majority of transcripts include all annotated super-repeat and Z-repeat exons (Kiss et al., 2020), the longest isoform was selected for further analysis. Species were chosen based on completeness of the nebulin genomic sequence, as reflected by the absence of gaps that might indicate missing exons in the annotation.

### Multiple sequence alignment

Protein sequences from all 53 species were aligned using Mirage (Nord et al., 2018), an alignment package for alternatively spliced protein isoforms. Because multiple species alignments are traditionally constructed using scoring criteria that prefer alignments with occasional mismatches over alignments with long gaps, they are not ideal for comparing nebulin isoforms with multiple-exon deletions or insertions. Using Mirage, isoforms were aligned to each other by first mapping each protein sequence to its encoding genomic sequence, and then aligning isoforms to one another based on the relative genomic coordinates of their constitutive codons. Alignments were manually corrected by repositioning of sequences based on actin-binding motifs (SDXXYK).

For the multiple sequence analyses presented in Data S1 and Data S2, MEGA7 (Kumar et al., 2016) was used to prepare the alignments using ClustalW and for refinement between close species using Muscle (Edgar, 2004). Because of the differential expansion of Z-repeats between species, individual translated exons were aligned to the Z-repeat region of chimpanzee using the lalign ExPASy portal ([https://embnet.vital-it.ch/software/LALIGN\\_form.html](https://embnet.vital-it.ch/software/LALIGN_form.html)); for nonmammalian species, this was repeated using the Z-repeat region of alligator as the reference. From this information, Z-repeat alignments were manually adjusted. Annotation of simple repeat designations were defined in Labeit and Kolmerer (1995), and sequence Logos were generated using the WebLogo portal (Schneider and Stephens, 1990; Crooks et al., 2004).

### Identification of nebulin domains and super-repeats

Owing to their high degree of homology to human nebulin sequences, protein domains and super-repeats from all species

were identified based on homology. Regions that did not align with human nebulin sequences were found only in the super-repeat and Z-repeat regions and are a result of DNA duplication events.

For definition of super-repeats, start and end positions of the super-repeat regions were identified by matching human super-repeat boundaries. We used an exon-based definition of super-repeats (Björklund et al., 2010) to allow for easy comparison between species. For analysis, multidomain exons were split into individual domains, with each one containing two half actin-binding motifs at its ends. Super-repeats were assigned for each seven-domain unit within the super-repeat region. Afterward, the number of super-repeats in each species was calculated based on the number of seven-domain units.

The starts of Z-repeats were identified by screening for a transition from 35–38-amino acid-long sequences to repetitive units of exactly 31 amino acids. Similar to super-repeats, Z-repeat boundaries were defined by localizing the first and the last 31-amino acid-long units in all species, and Z-repeats within those boundaries were counted. Because this region is subject to a high degree of alternative splicing, some repeats might not be annotated in the genomic sequences that we selected for analysis because of low expression levels in the tissue used for sequencing and gene prediction. To test this hypothesis, genomic sequences were searched for splice sites enclosing 93 nucleotides (AG-X(93)-GT) in the C-terminal region, and sequences without stop or nonsense codons were annotated as Z-repeats. Although there is no proof that these exons are expressed, it is likely that they are transcribed in slow muscle, based on Z-repeat expression observed in mice (Kiss et al., 2020).

### Sequence identity

To determine super-repeat sequence identity versus human/chimpanzee nebulin, super-repeats of all species were consecutively aligned to those from chimpanzee. After determining the percentage identity, this process was repeated for all super-repeats of the analyzed species. To be able to visually compare super-repeat identity between species, despite different numbers of super-repeats, gaps were introduced in heatmaps at positions without matching chimpanzee super-repeats. This also allowed better localization of super-repeat duplications relative to nebulin length because super-repeats that likely originate from the same sequence line up, and a lack of a matching super-repeat presumably reflects the presence of a super-repeat duplication event. For within-species super-repeat identity, a similar procedure was used, except that sequences were aligned to all other super-repeats within the same species. Heatmaps were generated using the R package pheatmap (<https://CRAN.R-project.org/package=pheatmap>) to visualize super-repeat differences by hierarchical clustering. The dendrogram was generated by hierarchical clustering (complete linkage clustering with Euclidian distance).

### Phylogenetic analysis

Nebulin sequences shared between all species (located outside of super-repeat and Z-repeat regions) were used for analysis. Alignments from all 53 species were used to generate a phylogenetic

tree by neighbor-joining using ClustalX. FigTree v1.4.2 was used for tree representation and editing.

### Scaling of nebulin size with body weight

To determine the possible correlation between the number of super-repeats and body weight, we used the allometric equation  $Y = aX^b$ , following the approach previously used to show that shortening velocity of muscle fibers inversely scales with body mass (Pellegrino et al., 2003). In our case,  $Y$  is the number of super-repeats,  $X$  is body mass (in kg), and  $a$  and  $b$  are variables, with  $b$  representing the scaling factor. The  $a$  and  $b$  values were determined by using a double logarithmic plot of  $\log(\text{number of super-repeats})$  versus  $\log(\text{body mass})$  for each of the species (for weight values, see Table 1), converting the above equation to  $\log(\text{SRs}) = b\log(\text{mass}) + \log(a)$ .

### Protein analysis

Flash-frozen muscle tissues were pulverized in liquid nitrogen and then solubilized in urea buffer (8 M urea, 2 M thiourea, 50 mM Tris-HCl, and 75 mM dithiothreitol with 3% SDS and 0.03% bromophenol blue, pH 6.8) and 50% glycerol with protease inhibitors (0.04 mM E64, 0.16 mM leupeptin, and 0.2 mM PMSF) at 60°C for 10 min (Tonino et al., 2017). Solubilized samples were centrifuged at 13,000 rpm for 5 min, aliquoted, flash-frozen in liquid nitrogen, and stored at -80°C. Nebulin expression analysis was performed on solubilized samples using a vertical SDS-agarose gel system (Warren and Greaser, 2003; Lahmers et al., 2004); 1% gels were run at 15 mA per gel for 3:20 h, stained using Coomassie brilliant blue, and scanned using a commercial scanner.

### Immunoelectron microscopy

Immunolabeling of nebulin-N terminus was performed on skinned EDL fiber bundles from WT mice stretched to different degrees from the slack length and processed by the preembedding technique previously described (Tonino et al., 2017). Fiber bundles were washed in relaxing solution before fixation with 3.7% paraformaldehyde in PBS for 30 min at 4°C, then rinsed with PBS containing protease inhibitors. Blocking was performed with 1% BSA in PBS containing protease inhibitors and 0.05% Tween 20, followed by incubation with the anti-nebulin N terminus primary antibody from rabbit (#6969; 0.2 mg/ml; Myomedix) for 48 h. After a rinsing step with PBS containing inhibitors, muscle fiber bundles were incubated with Nanogold (1.4 nm)-Fab' goat anti-rabbit antibody (#2004; 80 µg/ml; Nanoprobes) for 12 h. All incubations were performed in a humidity chamber at 4°C. After labeling, muscle tissues were washed in PBS, fixed with 3% glutaraldehyde in the same buffer, and processed for transmission EM. Muscles were postfixed in 1% OsO<sub>4</sub> in PBS for 30 min at 4°C. After this step, samples were dehydrated in an ethanol graded series, infiltrated with propylene oxide, and transferred to a 1:1 mix of propylene oxide: Araldite 502/Embed 812 (EMS). Subsequently, samples were transferred to a pure Araldite 502/Embed 812 resin and polymerized for 48 h at 60°C. Ultrathin (90-nm) longitudinal sections were obtained with a Reichert-Jung ultramicrotome and contrasted with 1% potassium permanganate and lead citrate.

Table 1. Number of super-repeats and alternatively spliced Z-repeats and predicted maximal length of a nebulin molecule in 53 analyzed species

Species	Weight (log(kg))	Class	Super-repeats	Z-repeats	Length (nm) <sup>a</sup>
Eastern happy	0.06	Fish	21	7	974
Elephant shark	0.84	Fish	22	8	1,018
Northern pike	1.19	Fish	21	5	963
Perch	0.02	Fish	21	7	974
Thorny skate	0.38	Fish	22	10	1,029
Tilapia	0.3	Fish	21	7	974
Trout	1.09	Fish	23	9	1,062
Whale shark	4.0	Fish	24	11	1,111
Zebrafish	-3.60	Fish	21	5	963
Caecilian	0.0	Amphibian	25	10	1,144
Frog	-1.39	Amphibian	22	7	1,012
Alligator	2.56	Reptile	22	15	1,056
Turtle	-0.4	Reptile	26	12	1,194
Viper	1.0	Reptile	29	14	1,320
Chicken	-0.11	Bird	23	18	1,111
Crow	0.15	Bird	24	16	1,139
Falcon	0.04	Bird	23	14	1,089
Golden eagle	0.66	Bird	25	15	1,172
Hummingbird	-2.4	Bird	21	10	990
Kakapo	0.48	Bird	25	14	1,166
Lance tail manakin	-1.74	Bird	21	14	1,012
Zebra finch	-1.93	Bird	22	15	1,056
Camel	3.0	Mammal	31	13	1,392
Cheetah	1.54	Mammal	24	12	1,117
Cow	2.92	Mammal	27	11	1,227
Dromedary	2.66	Mammal	25	13	1,161
Ermine	-0.56	Mammal	25	11	1,150
Horse	2.9	Mammal	27	14	1,243
Koala	0.96	Mammal	25	15	1,172
Little brown bat	-2.04	Mammal	25	11	1,150
Mouse	-1.39	Mammal	25	11	1,150
Pale spear-nose bat	-1.44	Mammal	22	11	1,034
Pig	2.06	Mammal	25	11	1,150
Rabbit	0.51	Mammal	25	14	1,166
Rat	-0.34	Mammal	24	12	1,117
Ryukyu mouse	-1.94	Mammal	22	10	1,029
Shrew mouse	-2.0	Mammal	25	11	1,150
Vaquita	1.64	Mammal	26	9	1,177
Water buffalo	3.0	Mammal	27	11	1,227
Bonobo	1.66	Primate	22	12	1,040

**Table 1. Number of super-repeats and alternatively spliced Z-repeats and predicted maximal length of a nebulin molecule in 53 analyzed species (Continued)**

Species	Weight (log(kg))	Class	Super-repeats	Z-repeats	Length (nm) <sup>a</sup>
Capuchin	0.29	Primate	23	12	1,078
Chimpanzee	1.7	Primate	31	13	1,392
Gelada	1.2	Primate	22	12	1,040
Gibbon	1.04	Primate	27	13	1,238
Gorilla	2.2	Primate	29	13	1,315
Human	1.92	Primate	29	13	1,315
Langur	1.2	Primate	25	14	1,166
Macaque	0.89	Primate	25	13	1,161
Night monkey	-0.23	Primate	25	12	1,155
Orangutan	1.96	Primate	29	13	1,315
Tarsier	-0.86	Primate	22	13	1,045

<sup>a</sup>Length of nebulin (number of simple repeats × 5.5 nm).

Observations were performed in a TECNAI G2 Spirit transmission electron microscope (FEI), and images were acquired with a side-mounted AMT Image Capture Engine V6.02 (4Mpix) digital camera operated at 100 kV. Digital images (1,792 × 1,792 pixels) were saved and calibrated for density plot profile analysis of nebulin N terminus precise location with ImageJ v1.49 (National Institutes of Health).

#### Online supplemental material

Table S1 lists genomic coordinates of analyzed species. Table S2 lists super-repeat sequences. Data S1 shows the sequence alignment of the N-terminal region of nebulin. Data S2 shows the sequence alignment of C-terminal domain of the last super-repeat. Data S3 shows multiple species' alignment of exons 178–180.

## Results

The location of nebulin in the skeletal muscle sarcomere is shown in Fig. 1, top, and the domain composition and its terminology that we will follow in this work in Fig. 1, bottom. We studied multiple birds, fish (bony and cartilaginous), amphibians, reptiles, and mammals (nonprimates and primates) for a total of 53 species. Species were chosen based on completeness of the nebulin genomic sequence. Reference sequences for nebulin were downloaded from NCBI RefSeq, translated, and aligned using a splice-aware multiple alignment software, Mirage (Nord et al., 2018; Table S1). Repeat sequences were determined based on alignments to the human nebulin sequence and the presence of actin-binding motifs (SDxxYK). Exons that did not align with human sequence were named based on their position relative to flanking human nebulin sequences. For determining the number of super-repeats and simple repeats, multidomain exons were split into individual repeats, and super-repeats were assigned in seven-repeat intervals (each containing the conserved tropomyosin binding site WLKIGW

within the third simple repeat) before counting the number of super-repeats and C-terminal simple repeats. Fig. 2 outlines the general methodology that was followed in our study.

Nebulin in all 53 species had a similar domain composition that mostly consists of ~30-residue-repetitive modules referred to as simple repeats, each of which corresponds to one actin-binding site (Fig. 3). All species contain a unique N-terminal sequence of lengths varying between 32 (zebrafish) and 112 (thorny skate) residues (Data S1). This is followed by exactly eight N-terminal simple repeats in all species (Data S1). In the central region of nebulin, simple repeats are always organized into seven-module super-repeats. The number of super-repeats varies between 21 and 31. In the C-terminal region, 12 simple repeats of variable lengths (31–38 amino acids) are followed by varying numbers of simple repeats (5–18) or Z-repeats, many of which are known to display a high degree of alternative splicing (Millevoi et al., 1998; Donner et al., 2004; Kiss et al., 2020) and can be distinguished from the initial repeats by their defined length of 31 amino acids and the presence of a SSVLYK motif (Data S2). At the C terminus, all species contain four simple repeats and serine-rich and Src homology-3 (SH3) domains. See Fig. 3 and Table 1 for details.

The highest degree of species-specific variation in nebulin was found in the number of super-repeats and Z-repeats, with the number of super-repeats ranging from 21 to 31 and the number of alternatively spliced Z-repeats from 5 to 18 (Table 1). The expected maximal length of nebulin was calculated for all analyzed species, assuming 5.5-nm spacing between simple repeats and 38.5 nm between super-repeats (Trinick, 1992; Kiss et al., 2020). Calculations are based on 24 simple repeats that are shared between all species, plus the species-specific number of super-repeats and Z-repeats. Maximal nebulin length is predicted to vary from 963 nm (zebrafish and northern pike) to 1,382 nm (camel and chimpanzee; Table 1).

To gain insights into sequence identity, super-repeat sequences from all species were aligned with chimpanzee super-repeats. Chimpanzee was chosen because it contains the highest number of super-repeats detected in our study. Conclusions regarding sequence identity are equally valid for human nebulin because sequences are >99.6% identical. Percentage sequence identity was plotted as a heatmap for all super-repeats from all species (Fig. 4 A). Gray squares represent positions without a matching super-repeat in the respective species. This allows row-wise comparisons of closely related super-repeats between species. Analysis revealed that species cluster by class. Bony fish have the lowest similarity to humans, followed by cartilaginous fish, amphibians, reptiles, and birds. Mammals, especially primates, have high sequence similarity to chimpanzee (or human) in the super-repeat region. Interestingly, the number of super-repeats does not correlate with sequence identity, indicating that multiple super-repeat duplication events occurred independently during evolution. The size of the super-repeat region can differ vastly between closely related species, while it can also be very similar for evolutionary distant species. This becomes apparent when comparing evolutionary distance with super-repeat counts. See Fig. 4, A and B, for details and Table S2 for super-repeat sequences.

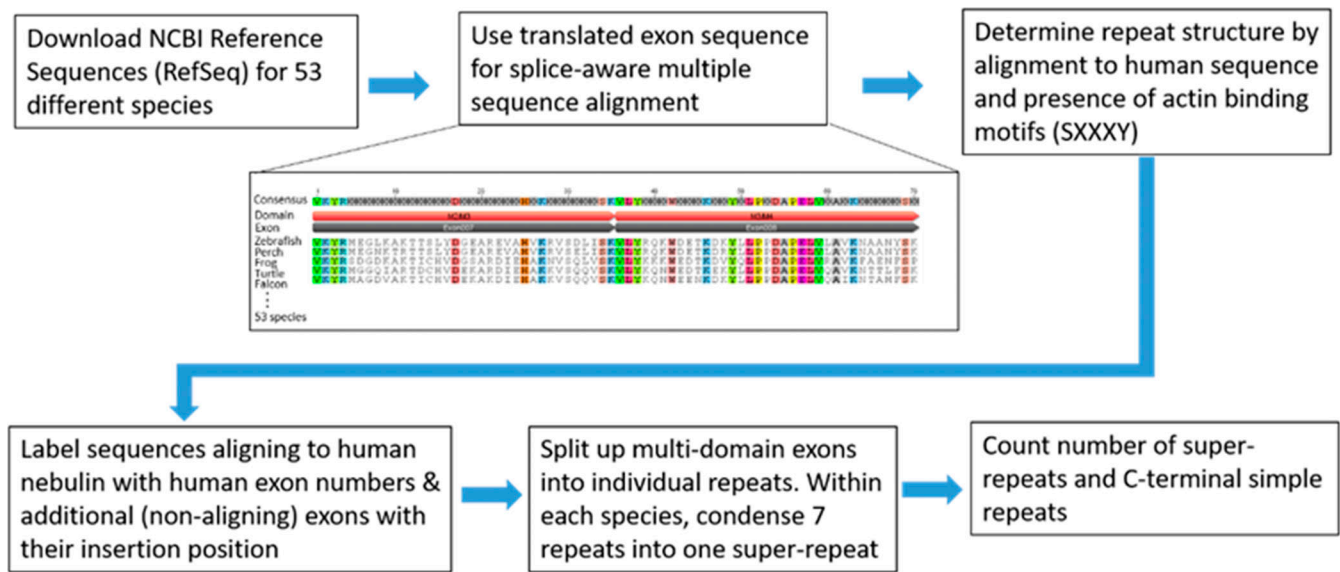


Figure 2. Outline of methods followed for obtaining the full nebulin sequence in 53 species. For details, see text.

Within-species super-repeat identity was determined by aligning individual super-repeats against all other super-repeats of the same species. High sequence similarity is evidence of recent super-repeat duplication events. Most super-repeat expansions occurred in the central part of nebulin (human super-repeats 11–13 and 17–21; Fig. 5). There is evidence for a duplication event of super-repeat 1 in mammals (except for vaquita and cheetah) that is absent in other classes and that must have occurred during relatively recent evolution. In elephant and viper, there has been a unique triplication of super-repeat 2 that is not found in other species. Super-repeat expansion does not seem to occur toward the C-terminal end of nebulin.

Conservation of nebulin exons that are located outside the super-repeat region and that are present in all species was

analyzed by determining pairwise identity across all species that were studied (Fig. 6, Data S1, Data S2, and Data S3). The last 3 simple repeat exons (178–180) are positioned between the last Z-repeat and the serine-rich region and are extraordinarily conserved (92.6%, 86.6%, and 94.4% identity), indicating that they might perform an important function.

It was shown recently that all encoded super-repeats are highly expressed in skeletal muscle of adult WT and genetically modified mice, in which the number of super-repeats (SRs) was altered from 25 (in WT mice) to either 22 (WT – 3 SR) or 28 (WT + 3 SR; Kiss et al., 2020). Whether this might be the case in other species was determined by coelectrophoresis of mouse, human, and carp muscle protein. Assuming that all super-repeats are also expressed in human and carp, the predicted size of nebulin would be highest in human muscle (29 super-repeats), followed

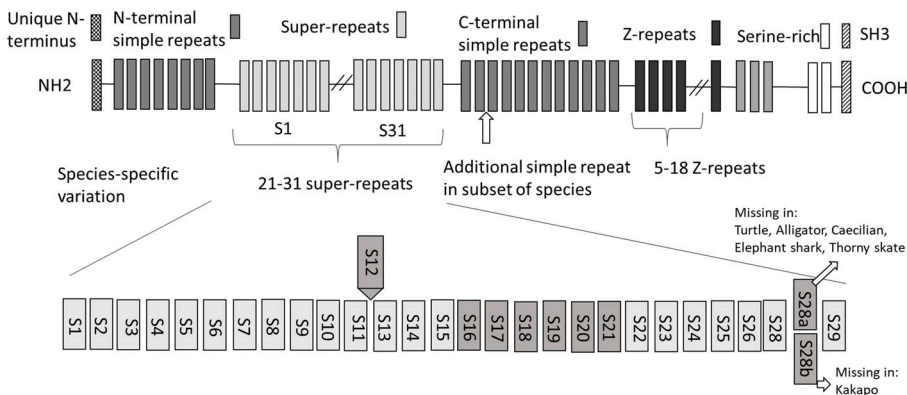


Figure 3. Domain composition of nebulin. In all 53 analyzed species, nebulin contains a unique N-terminal sequence and 8 simple N-terminal repeats, followed by 21–31 super-repeats (each containing 7 simple repeats), 12 constitutive simple repeats, 5–18 Z-repeats, 3 simple repeats, a serine-rich domain, and an SH3 domain. The fifth simple repeat of the penultimate super-repeat is alternatively spliced (Donner et al., 2004; Lindqvist et al., 2020), and both versions of this repeat are present in most species. S12 is known to be alternatively spliced in humans and mice (Donner et al., 2004; Witt et al., 2006; Kiss et al., 2020), and S16–S21 are alternatively spliced in humans (Donner et al., 2004). Most species encode a C-terminal simple repeat shortly after the end of the last super-repeat that had not been annotated in mouse or human nebulin; it contains an actin-binding motif and is present in the majority of species, except for pig, cheetah, trout, and all rodents (ermine, rat, and three mouse species).

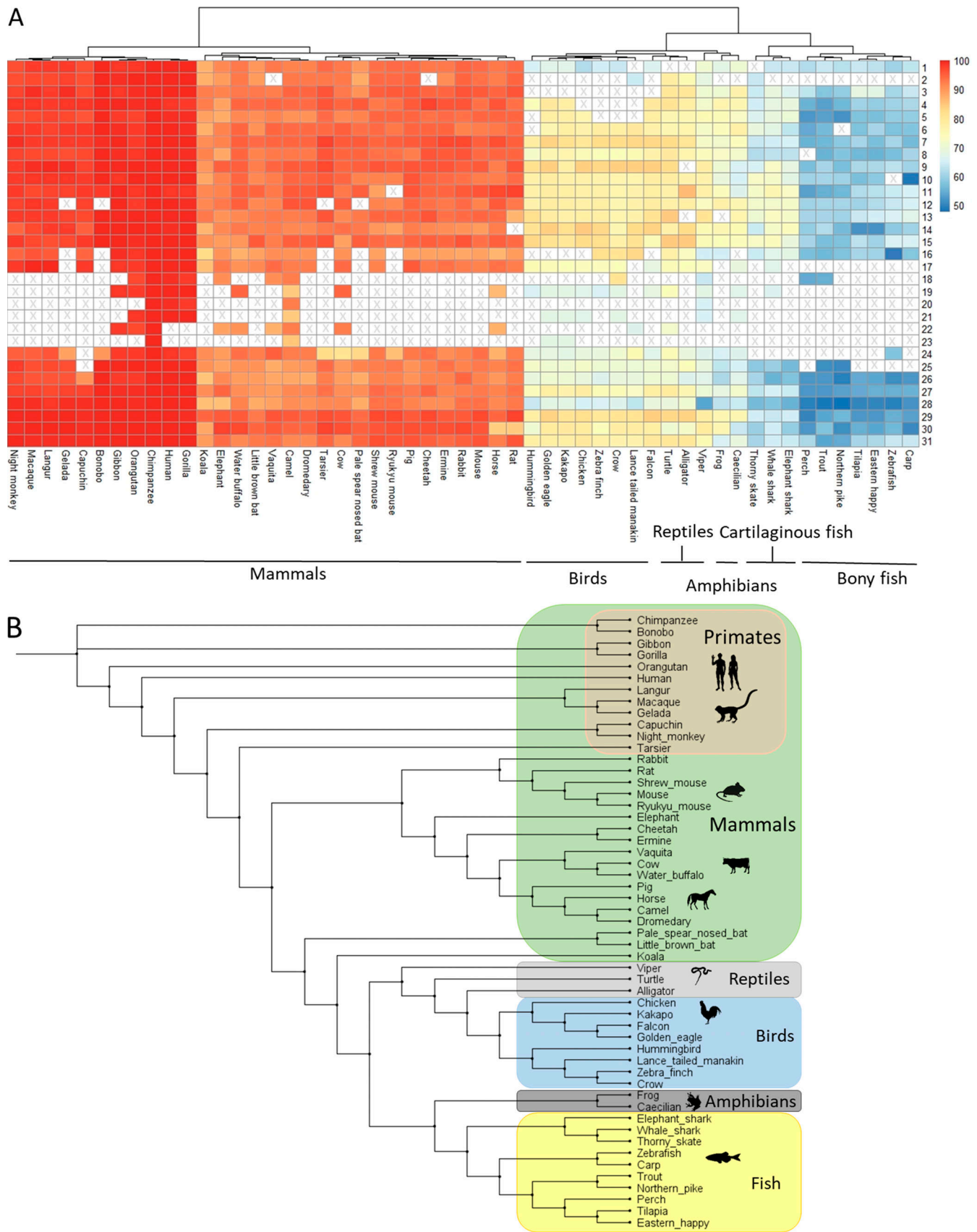


Figure 4. **Sequence identity between species.** (A) Super-repeat sequence identity compared with human/chimpanzee nebulin. Sequence identity is high in primates (red), intermediate in nonprimate mammals (orange), low in birds, reptiles, and amphibians (yellow), and lowest in fish (blue). The dendrogram is based on complete linkage hierarchical clustering with Euclidian distance and is a measure of super-repeat similarity. (B) Phylogenetic analysis of nebulin sequences in 53 analyzed species. Distances are based on nebulin protein sequences outside of the super-repeat/Z-repeat region.

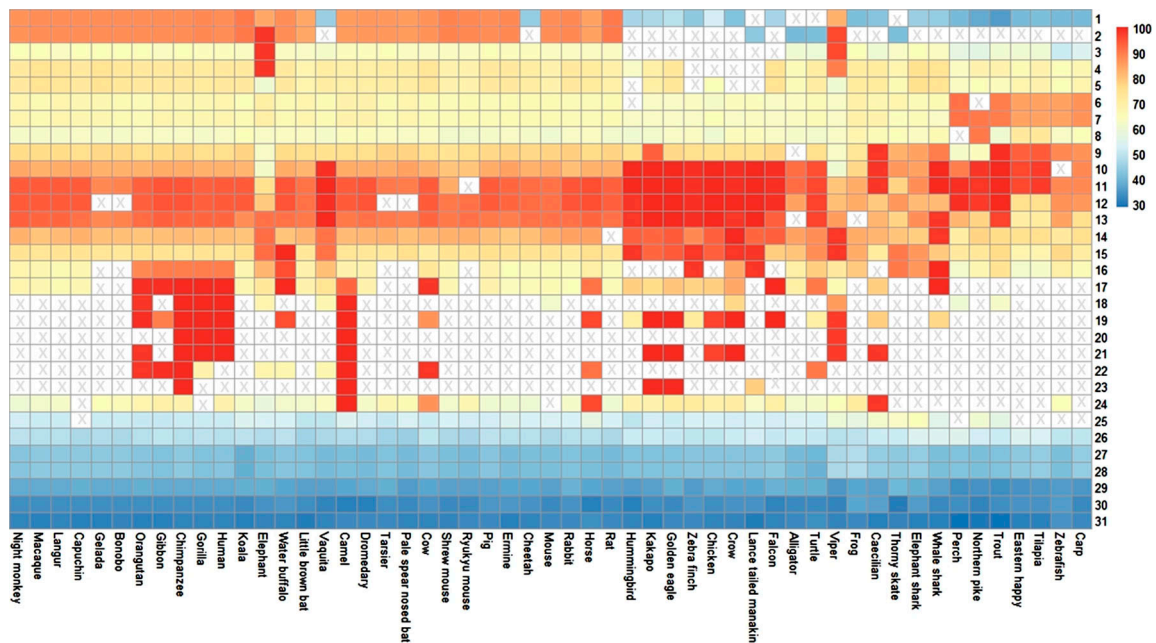


Figure 5. **Within-species super-repeat identity.** High sequence identity of super-repeat 1 in mammals (right) suggests that a duplication event has occurred. Super-repeat expansion also occurred in the central part of nebulin but not toward nebulin’s C terminus. For better comparability of relative positions of duplication events, gaps were inserted at positions without matching chimpanzee super-repeats. See text for details.

by mouse (WT + 3 SR), WT mouse (WT - 3 SR), and carp (21 super-repeats).

We tested this prediction by protein agarose-gel electrophoresis and running lysates prepared from white trunk muscle of carp, EDL muscle from the mouse models, and vastus lateralis muscle from human. Results shown in Fig. 7 reveal that the mobility of nebulin varies according to the number of super-repeats that the nebulin gene contains. This suggests that similar to mouse, carp and human muscles fully incorporate all of their super-repeats and that it is possible to predict nebulin size of diverse species based on their genomic sequence. This experiment also visualizes well the wide range of nebulin sizes found in different species.

Because nebulin functions in fast muscle as a TFL ruler (Kiss et al., 2020), we also plotted the number of super-repeats in a range of species against their known TFL. For this, we selected perch fast muscle, EDL muscle from WT mice (25 repeats), WT - 3 SR mice (22 SRs), WT + 3 SR mice (28 SRs), and human

quadriceps (29 super-repeats). Fig. 8 A shows that there is a linear relationship ( $R^2 = 0.99$ ) between the number of super-repeats and published TFLs from perch (Granzier et al., 1991), mouse (Kiss et al., 2020), and human (Kruger et al., 1991). The slope of the linear regression line is  $38.3 \pm 1.7$  nm, supporting that each super-repeat lengthens the TFL by  $\sim 38.5$  nm. We also determined whether nebulin ends at the thin filament’s pointed end or perhaps stops short, as immunofluorescence studies indicate a  $\sim 50$ -nm gap in fast muscle (Littlefield and Fowler, 2008; Kiss et al., 2020). We performed immunoelectron microscopy using an antibody to nebulin N terminus on mouse fast twitch fibers from the EDL muscle. This revealed that the N terminus is located at the edge of the H-zone (Fig. 8 B), supporting the theory that in fast muscle types nebulin extends to the thin filament’s pointed end. The variation in TFL across species has a pronounced effect on the predicted force-sarcomere length relation. For example, comparing perch and human predicts a shift in the location of the descending limb of  $0.66 \mu\text{m}$  (Fig. 8 C).

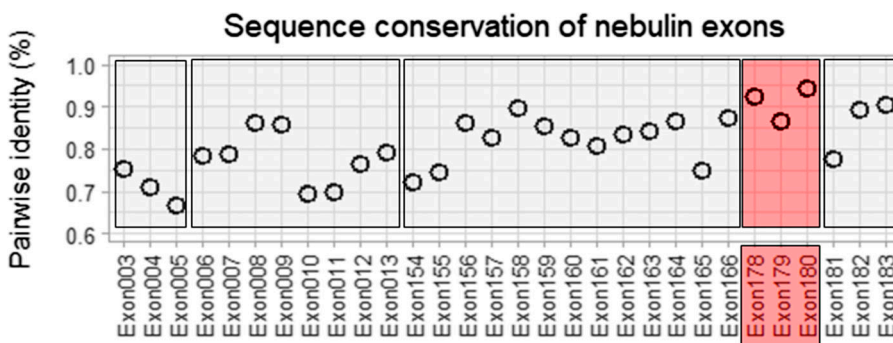


Figure 6. **Conservation of nebulin exons across species.** Displayed are exons outside of the super-repeat region that are present in all species. The last three simple repeat exons (178–180) are highly conserved. For a multiple-species alignment, see Data S3.



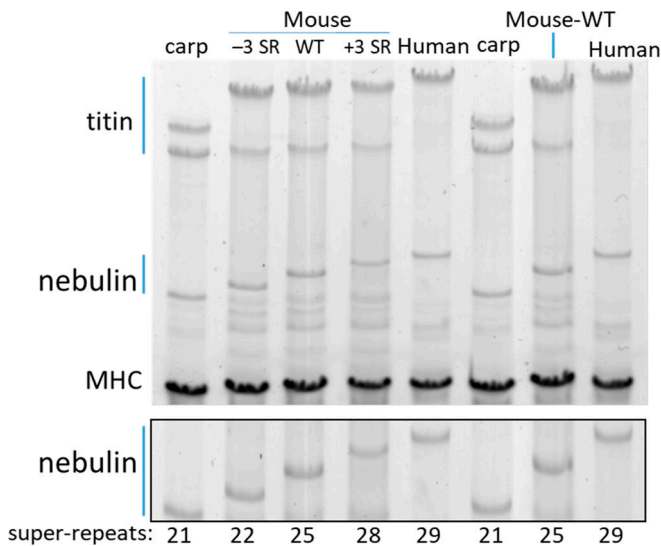


Figure 7. Gel electrophoresis of lysates prepared from carp muscle (white trunk muscle), mouse EDL muscle from WT (25 super-repeats), genetically altered mice expressing 22 or 28 nebulin super-repeats, and human skeletal muscle (vastus lateralis) with 29 nebulin super-repeats. The nebulin region of the gel is also shown on an expanded scale at the bottom. The mobility of nebulin varies inversely with the number of super-repeats.

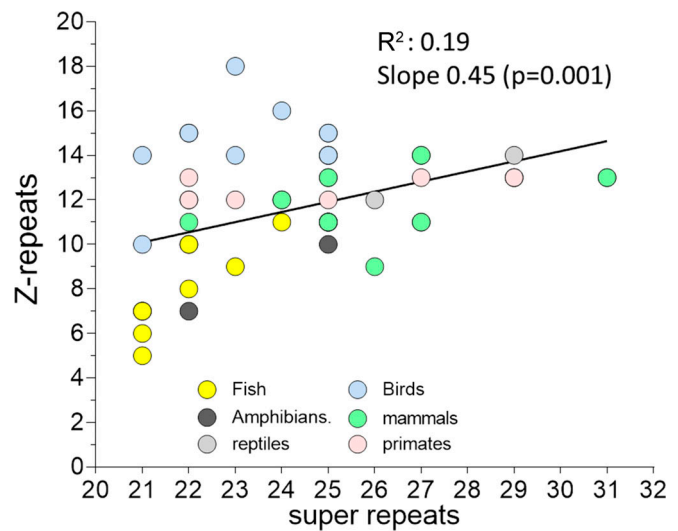


Figure 9. Correlation between the nebulin super-repeats and Z-repeats in the 53 analyzed species. There is a weak but significant correlation (slope is significantly different from 0,  $P = 0.001$ ). Thus, as the number of super-repeats increases, there is a tendency for the number of Z-repeats to increase as well (note that all 53 species were analyzed, but that several of their data points overlap). When analyzing the individual groups separately, fish, mammals, and primates had Z-repeat-versus-super-repeat correlations that were significant).

Because the number of C-terminal alternatively spliced Z-repeats was found to vary significantly between species, their number was plotted against the number of super-repeats for all species included in this study. Fig. 9 shows that although there is a lot of scatter, there is a trend for the number of Z-repeats to increase in species that contain a higher number of super-repeats. Linear regression analysis revealed that the correlation between the number of super-repeats and alternatively spliced repeats is significant ( $P = 0.001$ ), but that the model does not explain much variation of the data ( $R^2 = 0.19$ ). It is interesting

to note that many fish species fall below the linear regression line (i.e., their number of Z-repeats is lower than expected), whereas many bird species are above the linear regression line (i.e., their number of Z-repeats is higher than expected). For possible relevance of these findings, see Discussion.

Finally, to gain further insights into the possible significance of the wide range in the number of super-repeats of the analyzed species, we explored whether the number of super-repeats might scale with body size. We used the allometric equation  $Y = aX^b$  and obtained the scaling factor  $b$  from the slope of

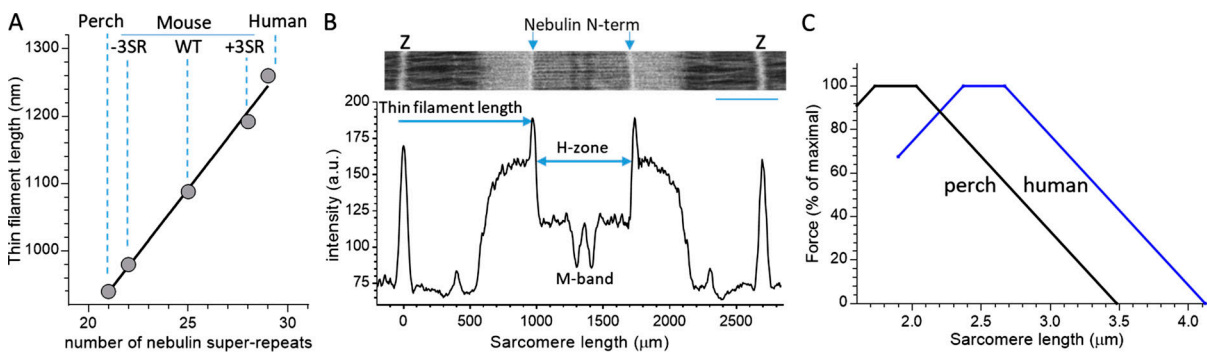


Figure 8. Nebulin super-repeats, TFL, and the force-sarcomere length relation. (A) Correlation between the number of nebulin super-repeats and TFL in a range of species. Data shown are for perch fast muscle (white trunk muscle), which has 21 super-repeats (Table 1) and a TFL of 940 nm (Granzier et al., 1991), EDL muscle from WT mouse (25 repeats, TFL 1,088 nm), WT - 3 SR mouse (22 super-repeats, TFL 980 nm), WT + 3 SR mouse (28 super-repeats, TFL 1,088 nm; Donner et al., 2004; Kiss et al., 2020; Lindqvist et al., 2020), and human quadriceps (29 super-repeats, TFL 1,270 nm; Kruger et al., 1991; Labeit and Kolmerer, 1995). TFL varies with the number of nebulin super-repeats ( $R^2$  of linear regression line = 0.99). (B) Mouse EDL immunolabeled with antibody to nebulin N terminus (inverse image). The antibody is localized at the edge of the overlap zone. If thin filaments were to extend beyond the nebulin epitope by ~50 nm (the approximate Z-disk width), this would be easily detectable. See text for details. Scale bar, 500 nm. (C) Predicted force-sarcomere length relation for perch and human fast muscle fibers (we assumed a constant thick filament length of 1.6  $\mu\text{m}$  and a bare zone width of 0.15  $\mu\text{m}$ ). The descending limbs are horizontally shifted by 0.66  $\mu\text{m}$ .

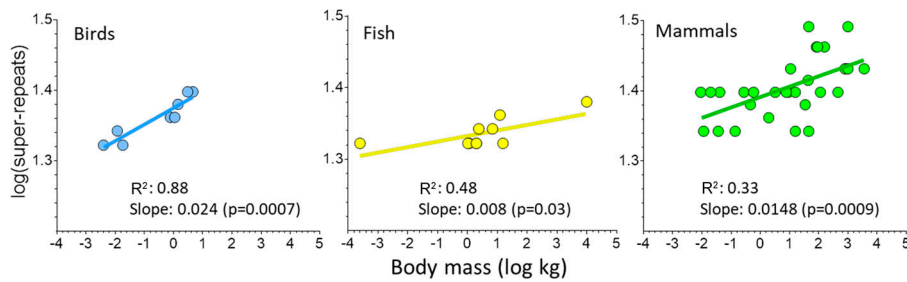


Figure 10. **Number of super-repeats versus body weight in birds, fish, and mammals.** Linear regression analysis reveals that there is a positive correlation between number of super-repeats and body weight, with slopes that are significantly different from 0. The model explains the variance well in birds ( $R^2 = 0.87$ ), followed by fish ( $R^2 = 0.48$ ) and mammals ( $R^2 = 0.33$ ).

log(number of super-repeats) versus log(body mass) (see Materials and methods for details). We evaluated animal classes for which more than three species had been analyzed (excluding amphibians and reptiles) and combined mammals and primates into a single group, ensuring a large number of analyzed species. Results are shown in Fig. 10. We performed a linear regression analysis, which revealed positive slopes that are significantly different from 0, with P values of 0.03, 0.0007, and 0.0009 for fish, birds, and mammals, respectively. The model explains the variance well in birds ( $R^2 = 0.88$ ), followed by fish ( $R^2 = 0.48$ ) and mammals ( $R^2 = 0.33$ ). The scaling factor ( $b$ ) was lowest for fish (0.008), highest for birds (0.024), and intermediate for mammals (0.015), and the difference between birds and fish reached statistical significance ( $P = 0.009$ ). We conclude that there is evolutionary pressure to increase the number of super-repeats with increasing body size.

## Discussion

Analyzing the nebulin gene in a wide range of species revealed a highly similar structure that consists of well-conserved N and C termini with a large number of simple repeats in between, most of which are organized in seven-domain super-repeats. The largest differences between species exists in the number of super-repeats, which ranges from 21 to 31, representing  $\sim 231$  kD at the protein level. We have shown previously that in fast muscle, nebulin regulates the length of the thin filament, with each super-repeat controlling a  $\sim 38.5$ -nm-long thin filament segment (Kiss et al., 2020). Focusing on fast muscles from species with a widely varying number of super-repeats and plotting the number of super-repeats against the previously measured TFL resulted in a linear relationship with a slope of  $38.3 \pm 1.7$  nm (Fig. 7). This is consistent with foundational work in the field that was based on the correlations between nebulin molecular weight and TFL (Kruger et al., 1991) and supports that in fast muscle, nebulin is a TFL ruler. The mechanistic basis of this is likely the stabilizing effect of nebulin on the thin filament, by making actin, tropomyosin, and tropomodulin (a thin filament's pointed end capping protein) less dynamic (Pappas et al., 2010).

Immuno-electron micrographs of fast muscles revealed that the pointed end of the thin filament (edge of the H-zone of the sarcomere) stops at the N terminus of nebulin (Fig. 8 B), as expected from a strict ruler. However, this is in contrast to results from immunofluorescence studies with an antibody against tropomodulin, to mark the thin filament's pointed end, and an antibody to nebulin's N terminus (the same antibody used in Fig. 8 B). These studies show a gap between the two labeled

epitopes, i.e., the thin filament protrudes beyond nebulin (Littlefield and Fowler, 2008; Gokhin and Fowler, 2013; Greaser and Pleitner, 2014; Kiss et al., 2020). This gap can be as large as  $\sim 100$ – $250$  nm in slow muscle, where it is now well accepted that a nebulin-free distal thin filament segment is present (Littlefield and Fowler, 2008; Gokhin and Fowler, 2013; Greaser and Pleitner, 2014; Kiss et al., 2020). However, the earlier studies revealed that even in fast twitch muscle (e.g., mouse EDL), a  $\sim 50$ -nm gap exists (Littlefield and Fowler, 2008; Gokhin and Fowler, 2013; Kiss et al., 2020). If thin filaments indeed extend beyond nebulin, this should be clearly detectable in electron micrographs of sarcomeres labeled with an antibody to nebulin's N terminus. However, our work reveals that this is not the case (Fig. 8 B). How to resolve this conundrum? One possibility is that the distance between nebulin's N terminus and tropomodulin detected in immunofluorescence studies is due to the physical arrangement of how tropomodulin is anchored to the thin filament's pointed end, where exactly the primary antibody binds to tropomodulin, and how the secondary antibody binds to the primary antibody. Future work is required to test this and other possible explanations.

It is noteworthy that the nebulin gene in different species always differed by an integral number of super-repeats, i.e., no gene sequences were found that contained a partial super-repeat, suggesting that the existence of partial super-repeats is not tolerated. This is consistent with studies on a mouse model in which a partial super-repeat has been deleted (Ottenheijm et al., 2013; Joureau et al., 2017), mimicking a mutation found in patients with NEM (Lehtokari et al., 2009; Yonath et al., 2012). Mice carrying the partial super-repeat deletion die soon after birth and express greatly reduced levels of nebulin protein (Ottenheijm et al., 2013), which has been explained by a mismatch between mutant nebulin and other thin filament proteins that are regularly spaced along the thin filament, increasing the sensitivity of nebulin to proteolysis (Ottenheijm et al., 2013). Thus, only the deletion/addition of an integer number of nebulin super-repeats at the level of the nebulin gene is tolerated and will be propagated during evolution, and not the partial deletion/addition of super-repeats. Preferential expansion in the central part of nebulin indicates that super-repeat duplication events at nebulin ends are less favorable; potentially, this is related to stronger actin-binding affinities at nebulin ends (Laitila et al., 2019). Addition of super-repeats through copy-number variation in the triplicate (TRI) region (located in human super-repeats 17–21) has been shown to be tolerated in human if one TRI copy (corresponding to two super-repeats) is added (Kiiski et al., 2016). In contrast, addition of two or more

copies is likely pathogenic (Kiiski et al., 2016). Interestingly, super-repeat expansion in the two species with the longest nebulin sequences found in this study (camel and chimpanzee) occurred through gain of an additional copy of this region.

Why does the number of super-repeats vary in different species? Is there a functional advantage to having more/less super-repeats? The correlation that was found between the number of super-repeats and body size (Fig. 10) provides insights. In the three classes of animals in which a sufficient number of species were analyzed (fish, birds, and mammals), the number of super-repeats increases with animal size, with scaling factors that are modest (0.0078–0.0234) but significant. Thus, as animals get bigger the number of super-repeats tends to increase and thin filaments become longer. The functional consequence of longer thin filaments is that the force–sarcomere length relation will be shifted to longer sarcomere length (by twice the TFL difference; Granzier et al., 1991). This allows optimal force levels to be produced at longer operating sarcomere lengths. Indeed, fish muscle with a low number of super-repeats functions at a short sarcomere length range (1.8–2.2  $\mu\text{m}$ ; Rome and Sosnicki, 1991), whereas in humans, as an example of a species with a large number of super-repeats, the sarcomere length working range is much longer, with typical values of  $\sim$ 2.2–3.0  $\mu\text{m}$  (Burkholder and Lieber, 2001). As visualized in Fig. 8 C, the shift in the force–sarcomere length relation ensures that the working range is at or near the optimal length for force generation in species as different as fish and human.

A longer operating sarcomere length range results in fewer sarcomeres in series per unit length of muscle fiber. Isometric force will therefore be produced more efficiently: per unit length of fiber, fewer myosin heads exist, and consequently ATP usage during isometric contraction will be less and the economy of ATP utilization will be increased. Additionally, it is expected that fewer sarcomeres in series will reduce the shortening velocity of muscle (the sum of the shortening of all sarcomeres in series). It was pointed out a long time ago (Hill, 1950) that animals of highly variable size (>1,000-fold variation) move at comparable speed (<4-fold variation) because as body size increases, the shortening velocity of muscle decreases. Some of this velocity reduction can be explained by a shift toward more slow fiber types in larger animals, but the reduction in shortening velocity with body size can also be seen in fibers that express the same myosin isoform (orthologous myosin), either fast or slow types (Pellegrino et al., 2003). This slowing of fibers expressing orthologous myosin isoforms in large animals is likely to have multiple explanations, and our work indicates that this includes a reduction in the number of sarcomeres in series, made possible through an increase in the number of nebulin super-repeats. The allometric coefficient for shortening speed of fibers that express orthologues type 2B myosin in vertebrate species has been determined at 0.041 (Pellegrino et al., 2003), and for nebulin, the allometric coefficient that we found for mammals is 0.0148 (Fig. 10), i.e.,  $\sim$ 37% of the shortening speed allometric factor. Thus, changes in nebulin can account for a considerable fraction of the increase in shortening speed with body size.

The N-terminal region of nebulin contains in all analyzed species eight simple repeats, whereas the C-terminal end contains

a variable number of simple repeats followed by a serine-rich domain and the terminal SH3 domain. In previous work, the SH3 domain was deleted in the mouse, which resulted in increased susceptibility to eccentric contraction-induced injury (Yamamoto et al., 2013), whereas another study deleted the final two unique C-terminal domains, the serine-rich region and the SH3 domain, which caused a moderate myopathy phenotype reminiscent of NEM (Li et al., 2019). The last three simple repeats (preceding the serine-rich domain) are extraordinarily conserved between species (92.56%, 86.59%, and 94.41%) suggesting that they perform an important function in the Z-disk that should be addressed in future research.

It is interesting that the C-terminal region of nebulin contains a varying number (5–18) of Z-repeats. These repeats are known from PCR studies in human and RNA sequencing studies in mouse to display a high degree of alternative splicing (Millevoi et al., 1998; Donner et al., 2004; Kiss et al., 2020). They are located within the Z-disk (Millevoi et al., 1998), a protein-dense region that functions as the attachment site for the thin filaments and that houses mechanosensing signaling molecules (Luther, 2009; Knöll et al., 2011). These Z-repeats are known to be alternatively spliced in human (Millevoi et al., 1998) and mouse (Kiss et al., 2020), with high expression in slow-twitch muscle (soleus) and low expression in fast-twitch muscle (EDL). Because the Z-disk width is large in slow muscle ( $\sim$ 150 nm) and small ( $\sim$ 60 nm) in fast muscle (Vigoreaux, 1994; Luther, 2009; Knöll et al., 2011), the alternatively spliced Z-repeats have been proposed to play a role in Z-disk width regulation (Millevoi et al., 1998; Kiss et al., 2020).

The width of the Z-disk is thought to reflect the mechanical strength of the Z-disk, but it is unclear why slow muscles require wide Z-disks and fast muscle can get away with narrow Z-disks, since it is the fast muscle that rapidly generates high force levels (Schiaffino and Reggiani, 2011). A possible explanation for wide Z-disks in slow muscle is that it is not only the level of force but also the duration that the force is applied that necessitates a wide Z-disk. Recent single-molecule studies provide evidence that individual interactions between Z-disk components (the interaction between  $\alpha$ -actinin and titin was studied) can be weak (to allow for protein turnover), but long-term stable anchoring can still be achieved by having an array of interactions (i.e., there can be up to four  $\alpha$ -actinin molecules interacting with titin in a wide Z-disk; Grison et al., 2017). We similarly propose that by having a large number of nebulin Z-repeats in wide Z-disks, multiple interactions between nebulin and other Z-disk components are made possible that, combined, result in stable anchoring during the long-lasting contractions that slow muscles produce.

How the Z-disk width is regulated has not been directly tested. It has been speculated that titin regulates the Z-disk width, because titin's N terminus spans the full Z-disk and contains differentially expressed Z-repeats that bind the structural Z-disk protein  $\alpha$ -actinin (Young et al., 1998). However, the number of titin Z-repeats and their spacing are incompatible with titin solely performing this task (Luther and Squire, 2002). Instead, it has been suggested that nebulin contributes to Z-disk width regulation with an important role played by the

differentially spliced Z-repeats (Millevoi et al., 1998). This notion is supported by the correlation between the number of nebulin Z-disk repeats and the Z-disk width (discussed above) and, furthermore, by the widened Z-disks that exist in nebulin-deficient muscle (Bang et al., 2006; Witt et al., 2006). Our present work suggests that the potential for wide Z-disks, by expressing all available Z-repeats, varies in different species. Few comparative studies have been performed of the Z-disk, but the limited insights stemming from studies on slow-twitch perch muscle (Z-repeats: 7), mouse slow fibers (Z-repeats: 11), and human slow fibers (Z-repeats: 13) with Z-disk widths of ~90 nm (Akster et al., 1985), ~105 nm (Li et al., 2019), and ~115 nm (Millevoi et al., 1998), respectively, reveals a positive correlation between the number of Z-repeats contained in the nebulin gene and the Z-disk width of slow muscle. Future comparative work on fibers from a wide range of species should examine this correlation more closely.

Comparing the number of Z-repeats with the number of super-repeats in the 53 analyzed species revealed a positive correlation (Fig. 9). Thus, when the number of super-repeats is high, the number of Z-repeats tends to be high as well. As argued above, a large number of super-repeats allows species that are large to generate force efficiently and reduce shortening speed, and a large number of Z-repeats allows slow muscle to have Z-disks that are wide so that they can withstand the high force levels produced during long-lasting contractions. Thus, the finding that the number of Z-repeats and super-repeats is positively correlated is consistent with the functional consequence of increasing the number of super-repeats. It is interesting that many of the fish species fall below the linear regression line in Fig. 9, suggesting that nebulin in fish has fewer Z-repeats than expected based on the correlation that considers all species. This might be explained by the low forces and high shortening speeds of fish muscle (Akster et al., 1985; Granzier et al., 1991), due to their rapid intrinsic myosin kinetics (Mead et al., 2020), and perhaps the low temperatures at which most fish species operate, which additionally could depress force levels.

In summary, the many genomic sequences that recently became available made it possible to analyze the complex and large nebulin gene in a wide range of species to gain novel insights that could not be obtained otherwise. Standout findings are the high variability in the number of nebulin super-repeats and that the number of super-repeats contained within the nebulin gene of a species scales with the size of the animal. We showed that in large mammals this will increase the length of the thin filament, and proposed that this will increase the efficiency of force production by shifting the sarcomere length working range to longer length, and thereby lowering the number of sarcomeres in series. The scaling factor was largest for birds, which is likely to reflect the high evolutionary pressure for energy efficiency owing to the high-energy demand of sustained flight. The reduced number of sarcomeres in series in large animals is expected to slow the shortening velocity of their muscle. It has been known since 1950 that as species increase in size, the shortening velocity of their muscle is reduced, and the present work contributes to understanding the mechanistic basis. The Z-repeats are located in a region of the molecule that is known to

undergo differential splicing, and we found that their number varies greatly in different species. The number of these Z-repeats is correlated with the number of super-repeats, supporting the notion that these Z-repeats play a role in increasing the width of the Z-disk in slow muscle. Finally, we propose that the increased Z-disk width of slow muscle is largely driven by the necessity to withstand forces that are long lasting.

## Acknowledgments

Olaf S. Andersen served as editor.

We are grateful to Ms. Zaynab Hourani for protein analysis.

This work was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (grant R01 AR053897 to H. Granzier).

H. Granzier is the Allan and Alfie Endowed Chair for Heart Disease in Women Research. The remaining authors declare no competing financial interests.

Author contributions: design of research: J. Gohlke, H. Granzier, and J.E. Smith, III; data collection and analysis: J. Gohlke, J. Lindqvist, and P. Tonino; writing: J. Gohlke, J.E. Smith, III, and H. Granzier.

Submitted: 30 September 2020

Accepted: 20 November 2020

## References

- Akster, H.A., H.L. Granzier, and H.E. ter Keurs. 1985. A comparison of quantitative ultrastructural and contractile characteristics of muscle fibre types of the perch, *Perca fluviatilis* L. *J. Comp. Physiol. B.* 155: 685–691. <https://doi.org/10.1007/BF00694582>
- Anderson, S.L., J. Ekstein, M.C. Donnelly, E.M. Keefe, N.R. Toto, L.A. LeVoci, and B.Y. Rubin. 2004. Nemaline myopathy in the Ashkenazi Jewish population is caused by a deletion in the nebulin gene. *Hum. Genet.* 115: 185–190. <https://doi.org/10.1007/s00439-004-1140-8>
- Bang, M.L., M. Caremani, E. Brunello, R.L. Lieber, J. Chen, V. Lombardi, and M. Linari. 2009. Nebulin plays a direct role in promoting strong actin-myosin interactions. *FASEB J.* 23:4117–4125. <https://doi.org/10.1096/fj.09-137729>
- Bang, M.L., X. Li, R. Littlefield, S. Bremner, A. Thor, K.U. Knowlton, R.L. Lieber, and J. Chen. 2006. Nebulin-deficient mice exhibit shorter thin filament lengths and reduced contractile function in skeletal muscle. *J. Cell Biol.* 173:905–916. <https://doi.org/10.1083/jcb.200603119>
- Björklund, A.K., S. Light, R. Sagit, and A. Lofsson. 2010. Nebulin: a study of protein repeat evolution. *J. Mol. Biol.* 402:38–51. <https://doi.org/10.1016/j.jmb.2010.07.011>
- Burkholder, T.J., and R.L. Lieber. 2001. Sarcomere length operating range of vertebrate muscles during movement. *J. Exp. Biol.* 204:1529–1536.
- Colombo, I., M. Scoto, A.Y. Manzur, S.A. Robb, L. Maggi, V. Gowda, T. Cullup, M. Yau, R. Phadke, C. Sewry, et al. 2015. Congenital myopathies: Natural history of a large pediatric cohort. *Neurology.* 84:28–35. <https://doi.org/10.1212/WNL.0000000000001110>
- Crooks, G.E., G. Hon, J.M. Chandonia, and S.E. Brenner. 2004. WebLogo: a sequence logo generator. *Genome Res.* 14:1188–1190. <https://doi.org/10.1101/gr.849004>
- Donner, K., M. Sandbacka, V.L. Lehtokari, C. Wallgren-Pettersson, and K. Pelin. 2004. Complete genomic structure of the human nebulin gene and identification of alternatively spliced transcripts. *Eur. J. Hum. Genet.* 12:744–751. <https://doi.org/10.1038/sj.ejhg.5201242>
- Edgar, R.C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* 5:113. <https://doi.org/10.1186/1471-2105-5-113>
- Gokhin, D.S., and V.M. Fowler. 2013. A two-segment model for thin filament architecture in skeletal muscle. *Nat. Rev. Mol. Cell Biol.* 14:113–119. <https://doi.org/10.1038/nrm3510>

- Granzier, H.L., H.A. Akster, and H.E. Ter Keurs. 1991. Effect of thin filament length on the force-sarcomere length relation of skeletal muscle. *Am. J. Physiol.* 260: C1060–C1070. <https://doi.org/10.1152/ajpcell.1991.260.5.C1060>
- Greaser, M.L., and J.M. Pleitner. 2014. Titin isoform size is not correlated with thin filament length in rat skeletal muscle. *Front. Physiol.* 5:35. <https://doi.org/10.3389/fphys.2014.00035>
- Gregorio, C.C., K. Trombitás, T. Centner, B. Kolmerer, G. Stier, K. Kunke, K. Suzuki, F. Obermayr, B. Herrmann, H. Granzier, et al. 1998. The NH2 terminus of titin spans the Z-disc: its interaction with a novel 19-kD ligand (T-cap) is required for sarcomeric integrity. *J. Cell Biol.* 143: 1013–1027. <https://doi.org/10.1083/jcb.143.4.1013>
- Grisson, M., U. Merkel, J. Kostan, K. Djinović-Carugo, and M. Rief. 2017.  $\alpha$ -Actinin/titin interaction: A dynamic and mechanically stable cluster of bonds in the muscle Z-disk. *Proc. Natl. Acad. Sci. USA.* 114:1015–1020. <https://doi.org/10.1073/pnas.1612681114>
- Hill, A.V. 1950. The dimensions of animals and their muscular dynamics. *Science Progress.* 38:208–229.
- Jin, J.P., and K. Wang. 1991a. Cloning, expression, and protein interaction of human nebulin fragments composed of varying numbers of sequence modules. *J. Biol. Chem.* 266:21215–21223.
- Jin, J.P., and K. Wang. 1991b. Nebulin as a giant actin-binding template protein in skeletal muscle sarcomere. Interaction of actin and cloned human nebulin fragments. *FEBS Lett.* 281:93–96. [https://doi.org/10.1016/0014-5793\(91\)80366-B](https://doi.org/10.1016/0014-5793(91)80366-B)
- Joureau, B., J.M. de Winter, K. Stam, H. Granzier, and C.A. Ottenheijm. 2017. Muscle weakness in respiratory and peripheral skeletal muscles in a mouse model for nebulin-based nemaline myopathy. *Neuromuscul. Disord.* 27:83–89. <https://doi.org/10.1016/j.nmd.2016.10.004>
- Kawai, M., T.S. Karam, J. Kolb, L. Wang, and H.L. Granzier. 2018. Nebulin increases thin filament stiffness and force per cross-bridge in slow-twitch soleus muscle fibers. *J. Gen. Physiol.* 150:1510–1522. <https://doi.org/10.1085/jgp.201812104>
- Kazmierski, S.T., P.B. Antin, C.C. Witt, N. Huebner, A.S. McElhinny, S. Labeit, and C.C. Gregorio. 2003. The complete mouse nebulin gene sequence and the identification of cardiac nebulin. *J. Mol. Biol.* 328:835–846. [https://doi.org/10.1016/S0022-2836\(03\)00348-6](https://doi.org/10.1016/S0022-2836(03)00348-6)
- Kiiski, K., V.L. Lehtokari, A. Löytynoja, L. Ahlsten, J. Laitila, C. Wallgren-Pettersson, and K. Pelin. 2016. A recurrent copy number variation of the NEB triplicate region: only revealed by the targeted nemaline myopathy CGH array. *Eur. J. Hum. Genet.* 24:574–580. <https://doi.org/10.1038/ejhg.2015.166>
- Kiss, B., J. Gohlke, P. Tonino, S. Hourani, J. Kolb, J. Strom, O. Alekhina, J.E. Smith III, C. Ottenheijm, C. Gregorio, and H. Granzier. 2020. Nebulin and Lmod2 are critical for specifying thin-filament length in skeletal muscle. *Sci. Adv.* 6:eabc1992. <https://doi.org/10.1126/sciadv.abc1992>
- Kiss, B., E.J. Lee, W. Ma, F.W. Li, P. Tonino, S.M. Mijailovich, T.C. Irving, and H.L. Granzier. 2018. Nebulin stiffens the thin filament and augments cross-bridge interaction in skeletal muscle. *Proc. Natl. Acad. Sci. USA.* 115:10369–10374. <https://doi.org/10.1073/pnas.1804726115>
- Knöll, R., B. Buyandelger, and M. Lab. 2011. The sarcomeric Z-disc and Z-discopathies. *J. Biomed. Biotechnol.* 2011:569628. <https://doi.org/10.1155/2011/569628>
- Kruger, M., J. Wright, and K. Wang. 1991. Nebulin as a length regulator of thin filaments of vertebrate skeletal muscles: correlation of thin filament length, nebulin size, and epitope profile. *J. Cell Biol.* 115:97–107. <https://doi.org/10.1083/jcb.115.1.97>
- Kumar, S., G. Stecher, and K. Tamura. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33: 1870–1874. <https://doi.org/10.1093/molbev/msw054>
- Labeit, S., and B. Kolmerer. 1995. The complete primary structure of human nebulin and its correlation to muscle structure. *J. Mol. Biol.* 248:308–315. [https://doi.org/10.1016/S0022-2836\(95\)80052-2](https://doi.org/10.1016/S0022-2836(95)80052-2)
- Labeit, S., T. Gibson, A. Lakey, K. Leonard, M. Zeviani, P. Knight, J. Wardale, and J. Trinick. 1991. Evidence that nebulin is a protein-ruler in muscle thin filaments. *FEBS Lett.* 282:313–316. [https://doi.org/10.1016/0014-5793\(91\)80503-U](https://doi.org/10.1016/0014-5793(91)80503-U)
- Lahmers, S., Y. Wu, D.R. Call, S. Labeit, and H. Granzier. 2004. Developmental control of titin isoform expression and passive stiffness in fetal and neonatal myocardium. *Circ. Res.* 94:505–513. <https://doi.org/10.1161/01.RES.0000115522.52554.86>
- Laing, N.G., and C. Wallgren-Pettersson. 2009. 161st ENMC International Workshop on nemaline myopathy and related disorders, Newcastle upon Tyne, 2008. *Neuromuscul. Disord.* 19:300–305. <https://doi.org/10.1016/j.nmd.2009.02.002>
- Laitila, J., J. Lehtonen, V.L. Lehtokari, L. Sagath, C. Wallgren-Pettersson, M. Grönholm, and K. Pelin. 2019. A nebulin super-repeat panel reveals stronger actin binding toward the ends of the super-repeat region. *Muscle Nerve.* 59:116–121. <https://doi.org/10.1002/mus.26350>
- Laitila, J.M., E.L. McNamara, C.D. Wingate, H. Goulee, J.A. Ross, R.L. Taylor, R. van der Pijl, L.M. Griffiths, R. Harries, G. Ravenscroft, et al. 2020. Nebulin nemaline myopathy recapitulated in a compound heterozygous mouse model with both a missense and a nonsense mutation in Neb. *Acta Neuropathol. Commun.* 8:18. <https://doi.org/10.1186/s40478-020-0893-1>
- Lawlor, M.W., C.A. Ottenheijm, V.L. Lehtokari, K. Cho, K. Pelin, C. Wallgren-Pettersson, H. Granzier, and A.H. Beggs. 2011. Novel mutations in NEB cause abnormal nebulin expression and markedly impaired muscle force generation in severe nemaline myopathy. *Skelet. Muscle.* 1:23. <https://doi.org/10.1186/2044-5040-1-23>
- Lehtokari, V.L., R.S. Greenleaf, E.T. DeChene, M. Kellinsalmi, K. Pelin, N.G. Laing, A.H. Beggs, and C. Wallgren-Pettersson. 2009. The exon 55 deletion in the nebulin gene—one single founder mutation with worldwide occurrence. *Neuromuscul. Disord.* 19:179–181. <https://doi.org/10.1016/j.nmd.2008.12.001>
- Lehtokari, V.L., K. Kiiski, S.A. Sandaradura, J. Laporte, P. Repo, J.A. Frey, K. Donner, M. Marttila, C. Saunders, P.G. Barth, et al. 2014. Mutation update: the spectra of nebulin variants and associated myopathies. *Hum. Mutat.* 35:1418–1426. <https://doi.org/10.1002/humu.22693>
- Lehtokari, V.L., K. Pelin, M. Sandbacka, S. Ranta, K. Donner, F. Muntoni, C. Sewry, C. Angelini, K. Bushby, P. Van den Bergh, et al. 2006. Identification of 45 novel mutations in the nebulin gene associated with autosomal recessive nemaline myopathy. *Hum. Mutat.* 27:946–956. <https://doi.org/10.1002/humu.20370>
- Li, F., E.R. Barton, and H. Granzier. 2019. Deleting nebulin's C-terminus reveals its importance to sarcomeric structure and function and is sufficient to invoke nemaline myopathy. *Hum. Mol. Genet.* 28:1709–1725. <https://doi.org/10.1093/hmg/ddz016>
- Li, F., D. Buck, J. De Winter, J. Kolb, H. Meng, C. Birch, R. Slater, Y.N. Escobar, J.E. Smith III, L. Yang, et al. 2015. Nebulin deficiency in adult muscle causes sarcomere defects and muscle-type-dependent changes in trophicity: novel insights in nemaline myopathy. *Hum. Mol. Genet.* 24: 5219–5233. <https://doi.org/10.1093/hmg/ddv243>
- Lindqvist, J., W. Ma, F. Li, Y. Hernandez, J. Kolb, B. Kiss, P. Tonino, R. van der Pijl, E. Karimi, H. Gong, et al. 2020. Triggering typical nemaline myopathy with compound heterozygous nebulin mutations reveals myofibrillar structural changes as pathomechanism. *Nat. Commun.* 11:2699. <https://doi.org/10.1038/s41467-020-16526-9>
- Littlefield, R.S., and V.M. Fowler. 2008. Thin filament length regulation in striated muscle sarcomeres: pointed-end dynamics go beyond a nebulin ruler. *Semin. Cell Dev. Biol.* 19:511–519. <https://doi.org/10.1016/j.semdb.2008.08.009>
- Lukoyanova, N., M.S. VanLoock, A. Orlova, V.E. Galkin, K. Wang, and E.H. Egelman. 2002. Each actin subunit has three nebulin binding sites: implications for steric blocking. *Curr. Biol.* 12:383–388. [https://doi.org/10.1016/S0960-9822\(02\)00678-4](https://doi.org/10.1016/S0960-9822(02)00678-4)
- Luther, P.K. 2009. The vertebrate muscle Z-disc: sarcomere anchor for structure and signalling. *J. Muscle Res. Cell Motil.* 30:171–185. <https://doi.org/10.1007/s10974-009-9189-6>
- Luther, P.K., and J.M. Squire. 2002. Muscle Z-band ultrastructure: titin Z-repeats and Z-band periodicities do not match. *J. Mol. Biol.* 319: 1157–1164. [https://doi.org/10.1016/S0022-2836\(02\)00372-8](https://doi.org/10.1016/S0022-2836(02)00372-8)
- Mead, A.F., G.G. Kennedy, B.M. Palmer, A.M. Ebert, and D.M. Warshaw. 2020. Mechanical Characteristics of Ultrafast Zebrafish Larval Swimming Muscles. *Biophys. J.* 119:806–820. <https://doi.org/10.1016/j.bpj.2020.06.036>
- Millevoi, S., K. Trombitas, B. Kolmerer, S. Kostin, J. Schaper, K. Pelin, H. Granzier, and S. Labeit. 1998. Characterization of nebulin and nebulin and emerging concepts of their roles for vertebrate Z-discs. *J. Mol. Biol.* 282:111–123. <https://doi.org/10.1006/jmbi.1998.1999>
- Nord, A., K. Carey, P. Hornbeck, and T. Wheeler. 2018. Splice-Aware Multiple Sequence Alignment of Protein Isoforms. *ACM BCB.* 2018:200–210. <https://doi.org/10.1145/3233547.3233592>
- Ogut, O., M.M. Hossain, and J.P. Jin. 2003. Interactions between nebulin-like motifs and thin filament regulatory proteins. *J. Biol. Chem.* 278: 3089–3097. <https://doi.org/10.1074/jbc.M205853200>
- Ottenheijm, C.A., D. Buck, J.M. de Winter, C. Ferrara, N. Piroddi, C. Tesi, J.R. Jasper, F.I. Malik, H. Meng, G.J. Stienen, et al. 2013. Deleting exon 55 from the nebulin gene induces severe muscle weakness in a mouse model for nemaline myopathy. *Brain.* 136:1718–1731. <https://doi.org/10.1093/brain/awt113>
- Ottenheijm, C.A., P. Hooijman, E.T. DeChene, G.J. Stienen, A.H. Beggs, and H. Granzier. 2010. Altered myofibrillar function depresses force

- generation in patients with nebulin-based nemaline myopathy (NEM2). *J. Struct. Biol.* 170:334–343. <https://doi.org/10.1016/j.jsb.2009.11.013>
- Ottenheijm, C.A., C.C. Witt, G.J. Stienen, S. Labeit, A.H. Beggs, and H. Granzier. 2009. Thin filament length dysregulation contributes to muscle weakness in nemaline myopathy patients with nebulin deficiency. *Hum. Mol. Genet.* 18:2359–2369. <https://doi.org/10.1093/hmg/ddp168>
- Pappas, C.T., K.T. Bliss, A. Zieseniss, and C.C. Gregorio. 2011. The Nebulin family: an actin support group. *Trends Cell Biol.* 21:29–37. <https://doi.org/10.1016/j.tcb.2010.09.005>
- Pappas, C.T., P.A. Krieg, and C.C. Gregorio. 2010. Nebulin regulates actin filament lengths by a stabilization mechanism. *J. Cell Biol.* 189:859–870. <https://doi.org/10.1083/jcb.201001043>
- Pellegrino, M.A., M. Canepari, R. Rossi, G. D'Antona, C. Reggiani, and R. Bottinelli. 2003. Orthologous myosin isoforms and scaling of shortening velocity with body size in mouse, rat, rabbit and human muscles. *J. Physiol.* 546:677–689. <https://doi.org/10.1113/jphysiol.2002.027375>
- Rome, L.C., and A.A. Sosnicki. 1991. Myofibrillar overlap in swimming carp. II. Sarcomere length changes during swimming. *Am. J. Physiol.* 260:C289–C296. <https://doi.org/10.1152/ajpcell.1991.260.2.C289>
- Schiaffino, S., and C. Reggiani. 2011. Fiber types in mammalian skeletal muscles. *Physiol. Rev.* 91:1447–1531. <https://doi.org/10.1152/physrev.00031.2010>
- Schneider, T.D., and R.M. Stephens. 1990. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* 18:6097–6100. <https://doi.org/10.1093/nar/18.20.6097>
- Stedman, H., K. Browning, N. Oliver, M. Oronzi-Scott, K. Fischbeck, S. Sarkar, J. Sylvester, R. Schmickel, and K. Wang. 1988. Nebulin cDNAs detect a 25-kilobase transcript in skeletal muscle and localize to human chromosome 2. *Genomics.* 2:1–7. [https://doi.org/10.1016/0888-7543\(88\)90102-4](https://doi.org/10.1016/0888-7543(88)90102-4)
- Tonino, P., B. Kiss, J. Strom, M. Methawasin, J.E. Smith III, J. Kolb, S. Labeit, and H. Granzier. 2017. The giant protein titin regulates the length of the striated muscle thick filament. *Nat. Commun.* 8:1041. <https://doi.org/10.1038/s41467-017-01144-9>
- Tonino, P., C.T. Pappas, B.D. Hudson, S. Labeit, C.C. Gregorio, and H. Granzier. 2010. Reduced myofibrillar connectivity and increased Z-disk width in nebulin-deficient skeletal muscle. *J. Cell Sci.* 123:384–391. <https://doi.org/10.1242/jcs.042234>
- Trinick, J. 1992. Understanding the functions of titin and nebulin. *FEBS Lett.* 307:44–48. [https://doi.org/10.1016/0014-5793\(92\)80899-R](https://doi.org/10.1016/0014-5793(92)80899-R)
- Vigoreaux, J.O. 1994. The muscle Z band: lessons in stress management. *J. Muscle Res. Cell Motil.* 15:237–255. <https://doi.org/10.1007/BF00123477>
- Wallgren-Petersson, C., C.A. Sewry, K.J. Nowak, and N.G. Laing. 2011. Nemaline myopathies. *Semin. Pediatr. Neurol.* 18:230–238. <https://doi.org/10.1016/j.spen.2011.10.004>
- Wang, K., M. Knipfer, Q.Q. Huang, A. van Heerden, L.C. Hsu, G. Gutierrez, X.L. Quian, and H. Stedman. 1996. Human skeletal muscle nebulin sequence encodes a blueprint for thin filament architecture. Sequence motifs and affinity profiles of tandem repeats and terminal SH3. *J. Biol. Chem.* 271:4304–4314. <https://doi.org/10.1074/jbc.271.8.4304>
- Warren, C.M., and M.L. Greaser. 2003. Method for cardiac myosin heavy chain separation by sodium dodecyl sulfate gel electrophoresis. *Anal. Biochem.* 320:149–151. [https://doi.org/10.1016/S0003-2697\(03\)00350-6](https://doi.org/10.1016/S0003-2697(03)00350-6)
- Witt, C.C., C. Burkart, D. Labeit, M. McNabb, Y. Wu, H. Granzier, and S. Labeit. 2006. Nebulin regulates thin filament length, contractility, and Z-disk structure in vivo. *EMBO J.* 25:3843–3855. <https://doi.org/10.1038/sj.emboj.7601242>
- Yamamoto, D.L., C. Vitiello, J. Zhang, D.S. Gokhin, A. Castaldi, G. Coulis, F. Piasek, M.C. Filomena, P.J. Eggenhuizen, P. Kunderfranco, et al. 2013. The nebulin SH3 domain is dispensable for normal skeletal muscle structure but is required for effective active load bearing in mouse. *J. Cell Sci.* 126:5477–5489. <https://doi.org/10.1242/jcs.137026>
- Yonath, H., H. Reznik-Wolf, M. Berkenstadt, S. Eisenberg-Barzilai, V.L. Lehtokari, C. Wallgren-Petersson, L. Mehta, R. Achiron, Y. Gilboa, S. Polak-Charcon, et al. 2012. Carrier state for the nebulin exon 55 deletion and abnormal prenatal ultrasound findings as potential signs of nemaline myopathy. *Prenat. Diagn.* 32:70–74. <https://doi.org/10.1002/pd.2905>
- Young, P., C. Ferguson, S. Bañuelos, and M. Gautel. 1998. Molecular structure of the sarcomeric Z-disk: two types of titin interactions lead to an asymmetrical sorting of alpha-actinin. *EMBO J.* 17:1614–1624. <https://doi.org/10.1093/emboj/17.6.1614>

## Supplemental material

Provided online are two tables in Excel format and three datasets in PDFs. Table S1 lists genomic coordinates of analyzed species. Table S2 lists super-repeat sequences. Data S1 lists the sequence alignment of N-terminal region of nebulin. Data S2 lists the sequence alignment of C-terminal domain of last super-repeat. Data S3 shows multiple species' alignment of exons 178–180.