Title: Adaptation of mammalian myosin II sequences to body mass

Mark N Wass*[1], Sarah T Jeanfavre*[1,2], Michael P Coghlan*[1], Martin Ridout[#], Anthony J Baines*[3] and Michael A Geeves*[3]

School of Biosciences*, and School of Mathematics[#], Statistics and Actuarial Science, University of Kent, Canterbury, UK

1. Equal contribution
2. Current address: Broad Institute, 415 Main Street, 7029-K, Cambridge MA 02142
3. Joint corresponding authors

Key words: selection, muscle contraction, mammalian physiology, heart rate

Address for correspondence:

Prof M.A.Geeves
School of Biosciences,
University of Kent,
Canterbury
CT1 7NJ UK

m.a.geeves@kent.ac.uk
tel 44 1227827597

Dr A J Baines
School of Biosciences,
University of Kent,
Canterbury
CT1 7NJ UK

a.j.baines@kent.ac.uk
tel 44 1227 823462

Abstract

The speed of muscle contraction is related to body size; muscles in larger species contract at a slower rate. We investigated the evolution of twelve myosin II isoforms across mammals to identify any adapted to increasing body mass. β-myosin head domain had the greatest rate of sequence divergence (0.05% per Myr) and it was the only domain where sequence divergence correlated with body mass (0.091% divergence per log mass unit). β-myosin is abundant in cardiac ventricle and slow skeletal muscle. Our analysis suggests that during evolution, β-myosin sequences have adapted to enable slower heart beating and contraction of slow skeletal muscle as body mass increased. Additionally, for eight of the twelve myosins, the ratio of divergence in the heads:tails was significantly different. For β-myosin the ratio was 3:1 while for the extraocluar, non-muscle A and embryonic myosin the ratio was <1:2. Our data provide new insights into the evolution of myosin function and indicate distinct evolutionary pressures on head and tail domains in individual isoforms.

Myosins are a large family of molecular motors responsible for the generation of forces and movements within eukaryotic cells. Different myosins from the 35 distinct subgroups [1] organise the actin cytoskeleton, drive cell motility, and participate in organelle & vesicle transport, cell division and signal transduction systems [2]. In mammals, the myosin IIs have two branches: one branch includes all of the striated muscle myosins IIs (at least 11 common isoforms) and the second includes both the non-muscle (three isoforms) and smooth-muscle (one gene alternately spliced to give two isoforms) lineage [3 4].

Striated muscles produce rapid movements of the whole organism (skeletal muscles) or pump blood to transport nutrients and waste products between specialist organs (cardiac muscle). A key feature of muscle cells is that the speed of contraction (specifically the maximum shortening speed) is a property of the specific myosin isoform(s) expressed in the cell [5]. The contraction parameters of a muscle adjust to the size of the organism; small animals have fast contracting muscle fibers. As size increases muscle contraction becomes slower to compensate for the greater momentum and inertia associated with a larger mass [4 6]. This has been widely observed in measured heart rates that are known to be correlated with basal metabolic rate and anti-correlated with body size across a wide range of species [7].

The rate of muscle contraction in different species can be controlled in two ways. Firstly, muscles can express different combinations of fast and slow contracting muscle fibres (containing different myosin isoforms) to adjust contraction parameters to the physiological requirements [4]. Secondly, during evolution individual myosin isoforms may adapt to different demand including the change in species size.

Here we consider the second of these mechanisms. If the maximum velocity of contraction, driven by a myosin isoform, varies with animal size then there must be specific alterations in the sequence of the myosin to generate the altered velocity. We have investigated evolution of the myosin-II family of proteins in mammals to

identify if myosin isoforms expressed in muscle have been adapted to body mass as animals have evolved from small to larger sizes. To do this we compare a set of 12 common myosin II isoforms (Table 1) from 12 mammals of varying size. Nine isoforms are from striated muscle, the smooth muscle isoform and two non-muscle isoforms. The non-muscle isoforms provide a negative control, as they act at the cellular level and are therefore unlikely to be influenced by body mass.

Results

All myosin II isoforms contain a head region, or motor-domain, and a tail region. The N-terminal globular head domain (approximately 760 amino acids) contains all of the requirements for motor activity, while the C-terminal tail region (~1200 amino acids) drives dimerization and oligomerization into myosin filaments. We have considered evolution of the head and tail domains separately as the shortening speed of muscle contraction is a property of the head domain and not the tail. 12 species were selected (Table 1), for which we had complete sequences for most of the 12 myosin II isoforms giving 130 sequences for the preliminary analysis. We excluded any sequences where sequence information was missing (see methods) as these could have affected the results obtained. We first considered evolutionary divergence by calculating the sequence identity of each pair of sequences within each isoform. The data is illustrated for three myosin isoforms in Figure 1A and the analysis for the other nine isoforms is given in the supplementary information (Figure S1). Sequence identity rather than conservation was used because evolution was considered within a single myosin isoform. A low level of divergence was expected and as such even conservative changes of amino acids may be relevant to adaptation to body mass. For the head domains eight of the 12 isoforms have a similar rate of divergence at ~0.035% per Myr (Figure 1B), while the β-myosin has a much higher rate of divergence (0.05% per Myr) and the embryonic and the non-muscle A and B head domains have a lower rate of divergence than average (0.01-0.02% per Myr). This is consistent with our hypothesis since a low divergence for non-muscle myosins suggests a tightly constrained sequence with little adaptive pressure. In contrast the β-myosin, which when in muscle shows the greatest dependence of velocity of

contraction on species size, has the greatest rate of divergence 2.5-5 fold higher than the non-muscle myosins.

For the tail domains, the β-myosin has the lowest rate of divergence (0.017% per Myr) and the extraocular isoform has the highest rate of divergence (0.095% per Myr). The myosin 2B isoform has a similar rate of divergence in the head and tail domains and so we compared the difference in the rate of divergence to this isoform (Figure 1B). This shows that there is greater divergence in the tail domains, with eight of the twelve domains having rates of divergence that are significantly different to that of the 2B isoform, whereas for the head the rate of divergence is significantly different for only four of the isoforms (Figure 1B).

The myosin domains studied here exhibit an average divergence rate of 0.035%/Myr for the heads and 0.04%/Myr for the tails. These are at the lower end of the divergence rates reported for proteins which range from the very highly conserved histones (eg, H2 0.0025% /Myr) through cytochrome c (0.067% /Myr) to 0.27% /Myr for haemoglobin or the immunoglobulin v region (1.25%/Myr) [8].

Comparing the ratio of head/tail sequence divergence (Figure 1B) many isoforms fall within the range 0.75 – 1.25 indicating that the two domains have a similar rate of divergence. However, the ratio of divergence for the head and tail regions were all significantly different from one except for the smooth muscle, non-muscle 2B, slow tonic muscle and fast muscle 2B myosin isoforms. Most strikingly the β-isoform has a much faster rate of divergence in the head than the tail (ratio of 2.9) while the opposite is true of extraocular myosin (ratio 0.36, due to high rate of tail divergence), non-muscle 2A and embryonic myosin (ratio 0.21 and 0.43 respectively, due to low rates of head divergence). It has been widely thought that the myosin head and tail domains have coevolved [1,9,10], however our analysis suggests that the rate of sequence divergence is not the same for the head and tail and varies for myosin II isoforms (see discussion).

Adaptation of myosin II to increasing body mass

To consider how myosin II isoforms may have evolved in adaptation to body mass, we investigated if there was a correlation between sequence divergence and body mass. Mass values were collected from the literature, and we quote the middle of the range of values reported. The range of masses cover more than six decades from 6 g to 10,000 kg (Supplementary Table 2). Being one of the smallest species, mouse was selected as a reference and the sequence identify of the other 11 species with the mouse sequence plotted against the species body mass (Figure 2A & B and Supplementary Figure 1). The results indicated a clear negative correlation of sequence identity with body mass for the β-myosin head domain (R= -0.945) with a gradient of 0.94% sequence change for each 10-fold increase in mass (Figure 2A). No such correlation exists for the β-myosin tail domain (R = 0.47). This result is quite striking and is consistent with organism mass being a driver of sequence divergence of the β-myosin, specifically the head domain.

None of the other myosin isoforms show a strong correlation between mass and sequence divergence in either the head or the tail, many of the lines are nearly horizontal (gradient <0.1% per log mass unit) and with low correlation coefficients and large error bars (Figure 2B). Other than the β-myosin head domain, the highest correlation coefficients were -0.61 for the myosin 2A head and -0.69 for the myosin 2x tail. Looking at the data in detail suggests some correlation if outliers are excluded but there is no strong reason to exclude individual data points from the small set of 12 sequences (Supplementary Figure 1).

This initial analysis used a small number of species, to ensure that the same set of species were represented in the analyses for all twelve myosin isoforms. This eliminates any internal bias from the species selected. To further establish if β-myosin has adapted to species mass we performed the same analysis for all of the available β-myosin sequences. To enable comparison with other isoforms we also expanded the analysis for the embryonic and non-muscle 2A myosin isoforms to provide a control. This added a further 27 species with a complete β-myosin sequence giving a total of 39 species. For the embryonic and the non-muscle

myosin 2A 50 and 60 sequences were used respectively. Many more partial sequences were available but excluded from our analysis because the calculation of sequence divergence would not be accurate.

For this larger data set of 39 β-myosin sequences, the relationship between mass and sequence divergence of the head domain is retained (gradient -0.73 ± 0.084% per log mass unit) with a slightly wider spread (R-value of -0.81, Figure 3). The analysis for the larger sets of the embryonic and non-muscle isoforms shows no mass dependence (gradient -0.091±0.053 and -0.106 ±0.060 % per log mass unit for embryonic and non-muscle myosin respectively with R = -0.22 in both cases; Figure 3). The data for the tail domains of all three isoforms show a small R-value of between -0.33 and -0.45 indicating no dependence on mass. The larger set of sequences enabled comparison between different evolutionary groups as the data contains species from Laurasiatheria (mostly ungulates and cetacean), Euarchontoglires (rodents and primates) and a couple of Afrotheria or Metatheria. The correlation between sequence divergence and body mass is maintained for these groups (Figure 3).

Location of sequence changes in the β-myosin

Since the sequence changes observed in each domain are relatively small (not exceeding 8%), it is of interest to examine where these sequence changes are positioned in the head and tail domains to evaluate if particular structural features of the domains are especially variable. The location of each of the sequence changes observed for the β, non-muscle 2A and embryonic myosins were considered (Figure 4). For each residue we considered the frequency of the consensus amino acids (black lines) and the number of different amino acids present at each position (red lines). In the β-myosin head domain, of the 778 amino acids 651 are totally conserved and a further 84 sites are highly conserved (i.e. fewer than four species have a different amino acid, for 54 of these sites only one species has a different amino acid). These changes occur in so few species that no conclusions can be drawn about the driver for these changes. Specific locations that show sequence changes in more than 4 species are of greater interest. There are 43 of these

(highlighted in Figure 4 by crossing the dotted line). The majority of these sites have only a single alternate amino acid and in most cases, the substitution is conservative.

The major functional regions of the motor domain are labelled by cyan bands in Figure 4. The sequence variations are scattered throughout the head both within and outside these functional areas with no identifiable pattern to the location of the changes. Variation in human β-myosin is known to cause inherited cardiomyopathies [11] with more than 300 mutations identified. Comparison of the location of human myopathy variants and those here show no simple correlation. For example, there are 23 cardiomyopathy linked mutations in the 68 amino acid converter region whereas there are very few in our data set. There are only 4 locations in the converter of the set of 39 β-myosins that have an alternate (conservative) amino acid substitution and this occurs in only one or two species. The one exception is converter residue Gly-747 which is replaced in 7 species by Ser.

Complete analysis of the sequence changes and their possible functional significance is beyond the scope of the current work but the type of issues raised by this analysis can be illustrated by consideration of Loop1. This is a surface loop near the entrance to the nucleotide-binding site and is a region that is hypervariable in both length and sequence between myosin classes and has been associated with controlling ADP release from myosin [12,13]. The rate constant for ADP release has ben implicated in settingthe maximum shortening velocity in some myosin IIs. In some species and muscle types the myosin gene is alternatively spliced in this region to generate myosin with different functions e.g. vertebrate smooth and scallop muscle myosins [14,15]. In the β-myosins analysed here, Loop 1 is identical in length throughout with the consensus sequence ^{202}GDRSKKDQTPGKG^{214} and is highly conserved. Only the three underlined sites have alternate residues (Figure 5). While patterns are observed, there is no simple relationship between residues changes and size. For example residue 208 changes between Asp and Glu with Glu predominant in larger species and Asp in smaller. Multiple amino acids are observed

at residue 210 but Asn and Thr predominate, with Asn occurring only once in the 15 largest species. Pro-211 is replaced in 15 species by Thr or by Ser, yet in the 19 smallest species Pro occurs 16 times with Thr only three times. A switch of amino acid between small and large species is not common to all of the 43 sites with alternative amino acids but such a pattern can be discerned in several locations for both conservative and non-conservative substitutions. Unlike Loop 1 the sites of these changes do not cluster to well defined functional sites.

In the β-myosin tail, there are very few changes, only four sites have more than two alternate residues, and ~30 sites have two alternate residues. The sites in which five or more species have a change in the residue are numbered in Fig 4. These are mostly in the N-terminal S2 region or in the C-terminal third of the tail. There is a hint of periodicity in the changes in the C-terminus with peaks at approximately 100 amino acid intervals.

In the other two myosins there are very few changes in sequence in the heads and those that stand out against the background are numbered in Figure 4. Very few of these have more than a single alternate amino acid. The sequence changes in the tail are much more common and are too frequent to pick out any specific details.

The large variation in the rates of myosin tail sequence divergence was not expected. Most myosin tails studied here have divergence rates of between 0.03 to 0.05% per Myr but β-myosin has an unusually low rate (0.017% per Myr) while the extraocular myosin has a rate of at least twice that of any other myosin (0.094% per Myr). Myosin II tails might be expected to be well-conserved since they require a strict seven amino acid repeat pattern to form the coiled-coil dimer, they must pack closely together in groups (both parallel and antiparallel packing) via multiple interactions to form the thick filament which may be composed of a single myosin isoform, or different isoforms. In addition, the myosin in the thick filament of the sarcomere has a large set of additional binding partners including titin, myosin binding protein C and M-line proteins. This set of interactions might be expected to limit the divergence of the tail region but they appear to cover a similar range to that

of the motor domain. The observation that the tail sequences may not all diverge at the same rate was perhaps unsurprising but such a large variation was not anticipated. At present we have no simple explanation for the distinct rates of divergence of the β and extraocular myosin tails.

Discussion

Our analysis of myosin evolution has identified two central findings: first that head and tail domains of certain myosin isoforms have not co-evolved at the same rates and secondly that sequence divergence of the β-myosin head domain is correlated with body mass indicating that the domain has evolved in adaptation to increasing body mass.

Previous studies have observed that across the wider myosin family the head and tail sequences are tightly coupled throughout evolution [9,10]. Particularly, Korn concluded that the myosin heads and tails have co-evolved since sequence analysis of the heads or tails resulted in the same pattern of evolution and the same assignment of myosins into subgroups. More recent analyses with increasing numbers of sequences have reached the same conclusion[1,3,16]. A study of the human myosin II sequences revealed almost identical degrees of divergence between the heads and tails of different isoforms[17]. Contrary to this study in a single species, we observed that myosin II isoforms can show distinct rates of divergence for their head and tail domains across a range of mammals. This is most clearly seen in the non-muscle 2a and extraocular myosins where the tail domains diverge at 3-4 times the rate that of the head domain, and in contrast in the β-myosin where the head diverges three times faster than the tail.

The high rate of β-myosin head domain divergence can be explained in the context of the adaptation to increasing body mass. The head domain directly controls muscle contraction velocity and the adaptation to changes in body mass occurred via variation in the head domain. Such variation did not occur in the tail domain because

it has a different functional role that is not directly relevant to speed of muscle contraction. In current views of muscle contraction the maximum shortening velocity (Vo) is related to the working stroke (d) of the myosin (head and neck) and the time (τ) during each ATP turnover for which the myosin remains attached to its site on the actin filament [18].

$$Vo = d/\tau$$

The working stroke has been estimated for a number of myosin II isoforms and is thought to be relatively invariant whilst τ does alter in parallel with changes to Vo [19] [20]. It is well established for the slow β-myosin isoform that the lifetime of the attached cross bridge is limited by the time required for ADP release after the working stroke ($1/k_{-ADP}$). k_{-ADP} has been measured for a number of purified β-myosin isoforms in vitro and the measured value predicts the measured Vo for a muscle contraction with high precision (see supplementary Figure 2).

Our investigation of adaptation to increasing body mass during evolution only identified a strong correlation for the β-myosin head domain. We propose that the variation identified in this domain is required to optimize the myosin for the required contraction velocity/heart rate. In the head domain we observed a 0.73% change in sequence per 10 fold change in mass. For a motor domain of 780 amino acids this translates to ~6 amino acids for each 10 fold change in mammal mass. Rat and human differ in mass by 230 fold (0.3 – 70 kg) and have 32 differences in the sequence of the β-myosin head. This pair of species is a useful comparator since the physiology of the heart and of slow skeletal muscle expressing this isoform and the protein biochemistry of the β-myosin has been studied in detail.

We suggest that three muscle contraction parameters are linked to changes in body mass and are therefore driven by the sequence changes in β-myosin; heart rate, muscle fibre maximum shortening velocity (Vo), and the rate constant limiting ADP release from the actin.myosin complex (k_{-ADP}). The heart rate, Vo and k_{-ADP} each change 3-5 fold between rat and man when measured under the same conditions. Heart-rate decreases from 330-450 b/min for rat to 60 – 180 b/min for man. The maximum shortening velocity of a skeletal muscle fiber expressing only β-myosin

decreases ~4 fold (1.42 to 0.33 μm/s per half sarcomere at 12 C[5]) and the rate constant of ADP release from the actin.β-myosin complex decreases almost 4 fold (100 s^{-1} to 27 s^{-1} [21,22]). These changes are driven by a subset of the 43 differences in the amino acid sequence.

Several areas of the myosin head have been shown to influence the rate constant of ADP release, including Loop1 illustrated in Fig 5 [12-15] and the region known as exon 7, which is alternately spliced in Drosophila myosin II [23,24]. Therefore the presence of variable residues in loop 1 (208, 210 and 211) and close to exon 7 (region D in Fig 4) that show variation with species mass, could alter ADP release.

While we have identified the positions that vary within the β-myosin head domain, to understand the role of individual residues in the function of myosin needs high-resolution structures of the different conformers of myosin found during the ATP driven cross-bridge cycle. At least four conformers are predicted to exist for all myosins [25,26] but there are only two crystal structures of the human β-myosin head domain in the protein databank (PDB: 4P7H and 4DB1) both in a similar conformation (sulphate or ATP analog in the active site) and there are no structures of any other mammalian sarcomeric myosins to use as templates for the other conformations. Thus detailed mapping of the sequence changes onto the structures to assess the potential role of each residue to alter myosin activity would be speculative at present. The alternative approach would be a comparison of a larger set of β-myosin sequences combined with a collection of defined contraction parameters associated with each myosin isoform. This may allow a correlation to be defined between specific sequence changes, body mass and contraction parameters.

The median size of a species is linked to evolutionary success in quite complex ways. Copes law[27,28] for example state that in general species tend to evolve from the small to large suggesting, at the simplest level, a survival advantage of large over small individuals/species in competition for resources or mate selection. There

are different factors that drive the need for larger animals to have slower contracting muscles. These include the mechanical constraints that require slower contracting muscles because of the larger mass to be moved and inertia. There may also be efficiency considerations as muscles are a significant source of heat generation and the balance between the need to keep warm and the need to lose heat alters between small and large animals. These two factors come together in the mammalian heart, where the basal metabolic rate dictates the rate at which the circulatory system delivers oxygen (plus other nutrients/waste products) and deals with heat dissipation between central and peripheral tissues and at the same time larger hearts beat more slowly as the volume of blood moved per contraction increases.

The head domain of most adult muscle myosin isoforms show a similar rate of sequence divergence of ~0.035%/Myr and faster than the non-muscle isoforms and with the exception of β-myosin there was no clear correlation between sequence divergence and mass. It is possible that further analysis of larger data sets may reveal different populations of mammal myosins that do or do not fit a-size dependence. One of the complexities of studying a wide range of mammals is that they do not all use their muscles and myosin isoforms in the same way or for the same purpose. Gait and posture, for example, vary considerably and may contribute different selective pressures to muscle and myosin divergence. Mammals are remarkably adaptable and muscle tissue is particularly plastic. The contraction parameters (including velocity) of a skeletal muscle can be adapted by adjusting the mix of fiber types (and hence myosin isoforms) used in a specific muscle. Thus, the selective pressure on a specific myosin isoform in skeletal muscle may be different in each mammal. The β-cardiac/slow muscle myosin is unusual in that there is only a single major slow isoform and therefore there is little scope to adjust the contraction parameters by mixing fibre types. Fast muscle expresses a combination of two or three fast isoforms in addition to some slow isoform and specialist isoforms. Data on detailed biomechanics on contraction of muscles fibres expressing a single myosin isoform are rarely available, making any detailed assessment of selective pressure difficult. The heart and slow muscle fibres, in contrast, are relatively similar in

function across mammals and the simple relationship between heart rate and size is well established. Only two myosin isoforms are expressed to any significant extent in the heart α and β and only in the smallest mammals does α-cardiac myosin replace β in the heart ventricle (the β myosin is however used in slow twitch muscle even in small mammals such as the mouse). This may therefore allow the size-sequence divergence relationship to be more easily defined.

In summary, we present evidence that the head and tail domains of some myosin isoforms have evolved at different rates and that these differences in divergence are likely to relate to specialised function. This is highlighted by the correlation of β-myosin sequence divergence with body mass, which we propose is required to fine tune the heart rate of the species.

Materials and Methods

Protein sequences were extracted from RefSeq [29] and Uniprot [30] as listed in Supplementary Tables 1 & 2. Sequences were selected either as canonical, well annotated isoforms (e.g. human MyHC7 Uniprot:P12883), or by using BLAST [31] to search for homologues of the human myosin proteins within the genomes of the other mammals. To ensure that each sequence corresponded to the requisite specific isoform, each sequence identified initially by database annotation or by BLAST, was further analysed using BLAST to compare it to UniProt to ensure that the sequence was complete and that it was most closely similar to the canonical standard for that isoform. Incomplete sequences were excluded from our analysis because sequence gaps could have a major effect on the sequence comparison for such closely related isoforms.

For each isoform, the protein sequences were divided into the Head (1-778, β-myosin numbering) and Tail (779 – 1938) regions. For each region of each myosin, the sequences were aligned using Clustal Omega [32]; the resulting multiple sequence alignment was used to construct a percentage identity matrix between the species. Sequence identity was used rather than sequence similarity as we are considering small changes (>93% identity, >98% similarity) within the isoform and substitutions

that would normally be classed as similar (e.g. aspartate to glutamate) may be relevant. A 2-fold change in a rate constant that controls, for example shortening velocity, requires only a change of the order of 1 kcal/mol in the activation energy according to transition state theory. This is comparable to a single H-bond or van der Waals interaction. Neighbour joining trees were also generated using Clustal Omega, with the opossum sequences set as the outgroup.

An evolutionary distance matrix was generated using evolutionary distances extracted from TimeTree [33]).

The masses of each species were extracted from a wide range of information sources and are listed in the Supplementary Information Tables 1 & 2. To compare sequence divergence against either evolutionary time or animal body mass, the relevant matrices were plotted against each other.

The rates of evolutionary divergence for each myosin domain were calculated from graphs of the sequence identity versus evolutionary distance using a linear model of regression fitted using unweighted least squares with the regression line passing through the origin (i.e. 100% sequence identity at an evolutionary distance of zero). To test for a difference in slope between the Head and Tail domains for each myosin isoform, the test statistic used was

$$z = \frac{\widehat{\gamma_H} - \widehat{\gamma_T}}{\sqrt{SE_H^2 + SE_T^2}}, \qquad \text{eqn 1}$$

where $\widehat{\gamma_H}$ and $\widehat{\gamma_T}$ are the slope estimates for the Head and Tail regions, and SE_H^2 and $SE_T^2$are the standard errors of these estimates. The value of z was compared to the standard normal distribution to determine significance. A similar approach (z-calculation) was used to compare the rates of divergence of pairs of Head domains and pairs of Tail domains.

Acknowledgements

MNW was supported by a Royal Society research grant. The authors would like to thank Martin Michaelis, Chris Toseland, and Carlo Reggiani for useful discussions and comments on the manuscript.

References

1. Odronitz, F. & Kollmar, M. Drawing the tree of eukaryotic life based on the analysis of 2,269 manually annotated myosins from 328 species. *Genome Biol.* **8,** R196 (2007).
2. Krendel, M. & Mooseker, M. S. Myosins: tails (and heads) of functional diversity. *Physiology (Bethesda)* **20,** 239–251 (2005).
3. Golomb, E. *et al.* Identification and Characterization of Nonmuscle Myosin II-C, a New Member of the Myosin II Family. *J. Biol. Chem.* **279,** 2800–2808 (2004).
4. Schiaffino, S. & Reggiani, C. Fiber types in mammalian skeletal muscles. *Physiol. Rev.* **91,** 1447–1531 (2011).
5. Pellegrino, M. A. *et al.* Orthologous myosin isoforms and scaling of shortening velocity with body size in mouse, rat, rabbit and human muscles. *The Journal of Physiology* **546,** 677–689 (2003).
6. Hill, A. V. The dimensions of animals and their muscular dynamics. *Sci Prog* **38,** 209–230 (1950).
7. Savage, V. M. *et al.* Scaling of number, size, and metabolic rate of cells with body size in mammals. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 4718–4723 (2007).
8. Whitford, D. *Protein Structure and Function*. (Wiley).
9. Goodson, H. V. & Spudich, J. A. Molecular evolution of the myosin family: relationships derived from comparisons of amino acid sequences. *Proc. Natl. Acad. Sci. U.S.A.* **90,** 659–663 (1993).
10. Korn, E. D. Coevolution of head, neck, and tail domains of myosin heavy chains. *Proc. Natl. Acad. Sci. U.S.A.* **97,** 12559–12564 (2000).
11. Buvoli, M., Hamady, M., Leinwand, L. A. & Knight, R. Bioinformatics assessment of beta-myosin mutations reveals myosin's high sensitivity to mutations. *Trends Cardiovasc. Med.* **18,** 141–149 (2008).
12. Spudich, J. A. How molecular motors work. *Nature* **372,** 515–518 (1994).
13. Goodson, H. V., Warrick, H. M. & Spudich, J. A. Specialized conservation of surface loops of myosin: evidence that loops are involved in determining functional characteristics. *J. Mol. Biol.* **287,** 173–185 (1999).
14. Kurzawa-Goertz, S. E., Perreault-Micale, C. L., Trybus, K. M., Szent-Györgyi, A. G. & Geeves, M. A. Loop I can modulate ADP affinity, ATPase activity, and motility of different scallop myosins. Transient kinetic analysis of S1 isoforms. *Biochemistry* **37,** 7517–7525 (1998).
15. Sweeney, H. L. *et al.* Kinetic tuning of myosin via a flexible loop adjacent to the nucleotide binding pocket. *J. Biol. Chem.* **273,** 6262–6270 (1998).
16. Foth, B. J., Goedecke, M. C. & Soldati, D. New insights into myosin evolution and classification. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 3681–3686 (2006).
17. Weiss, A. & Leinwand, L. A. The mammalian myosin heavy chain gene family. *Annu. Rev. Cell Dev. Biol.* **12,** 417–439 (1996).
18. Uyeda, T. Q. P., Kron, S. J. & Spudich, J. A. Myosin step size: Estimation from slow sliding movement of actin over low densities of heavy meromyosin. *J. Mol. Biol.* **214,** 699–710 (1990).
19. Palmiter, K. A., Tyska, M. J., Dupuis, D. E., Alpert, N. R. & Warshaw, D. M. Kinetic differences at the single molecule level account for the functional

diversity of rabbit cardiac myosin isoforms. *The Journal of Physiology* **519,** 669–678 (1999).
20. Bottinelli, R. & Reggiani, C. Human skeletal muscle fibres: molecular and functional diversity. *Progress in Biophysics and Molecular Biology* **73,** 195–262 (2000).
21. Bloemink, M. J., Adamek, N., Reggiani, C. & Geeves, M. A. Kinetic analysis of the slow skeletal myosin MHC-1 isoform from bovine masseter muscle. *J. Mol. Biol.* **373,** 1184–1197 (2007).
22. Nyitrai, M. *et al.* What limits the velocity of fast-skeletal muscle contraction in mammals? *J. Mol. Biol.* **355,** 432–442 (2006).
23. Swank, D. M. *et al.* An alternative domain near the ATP binding pocket of Drosophila myosin affects muscle fiber kinetics. *Biophys. J.* **90,** 2427–2435 (2006).
24. Miller, B. M., Bloemink, M. J., Nyitrai, M., Bernstein, S. I. & Geeves, M. A. A variable domain near the ATP-binding site in Drosophila muscle myosin is part of the communication pathway between the nucleotide and actin-binding sites. *J. Mol. Biol.* **368,** 1051–1066 (2007).
25. Geeves, M. A. & Holmes, K. C. The molecular mechanism of muscle contraction. *Adv. Protein Chem.* **71,** 161–193 (2005).
26. Sweeney, H. L. & Houdusse, A. Structural and functional insights into the Myosin motor mechanism. *Annu Rev Biophys* **39,** 539–557 (2010).
27. Stanley, S. M. An Explanation for Cope's Rule. *Evolution* **27,** 1 (1973).
28. Hone, D. W. E. & Benton, M. J. The evolution of large size: how does Cope's Rule work? *Trends Ecol. Evol. (Amst.)* **20,** 4–6 (2005).
29. Pruitt, K. D. *et al.* RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res.* **42,** D756–63 (2014).
30. UniProt Consortium. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **42,** D191–8 (2014).
31. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
32. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 539 (2011).
33. Hedges, S. B., Dudley, J. & Kumar, S. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22,** 2971–2972 (2006).

Tables and Figures

Table 1 myosin II Isoforms considered

Gene name	Heavy Chain	Isoform	Short name
MYH1	MyHC-1	Skeletal 2B	2B
MYH2	MyHC-2	Skeletal 2A	2A
MYH3	MyHC-3	Embryonic	EMB
MYH4	MyHC-4	Skeletal 2D/X	2X
MYH6	MyHC-6	α cardiac	α
MYH7	MyHC-7	β-myosin	β
MYH8	MyHC-8	Perinatal	PERI
MYH9	MyHC-9	Non-muscle A	NMA
MYH10	MyHC-10	Non-muscle B	NMB
MYH11	MyHC-11	Smooth Muscle	SM
MYH13	MyHC-13	Extraocular	EXOC
MYH7b	MyHC-7b	Slow Tonic	SlowT

Figure 1

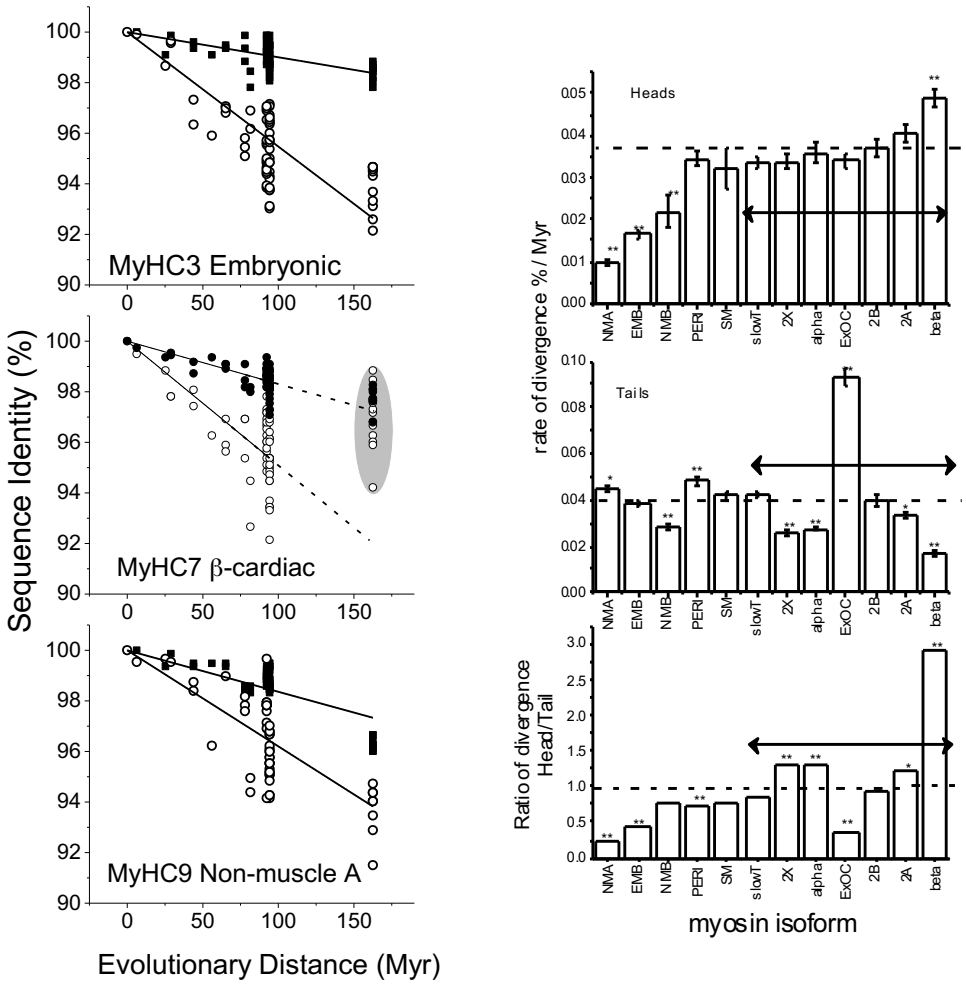

Figure 1. <u>Rate of sequence divergence of myosin II isoforms (Heads and Tails) for 12 species</u>

A) Pair-wise comparison of sequence identity for 12 species compared to evolutionary distance calculated by time tree. The best fit regression line through zero is superimposed. Points highlighted with a grey background represent data that have been excluded from the regression line (Opossum data in the myosin 7 plot).

Head domains of each isoform are shown in black squares with tail domains as empty circles.

B) The slope (± SE) of the Regression line of equivalent plots to Fig 1A for all available myosin 2 isoforms (original plots are in Supplementary Data Fig S1) shown separately for the Head and Tail segments and the ratio (head/tail) of the two. The dashed line represents the value for myosin 2B which is close to the means of all values in both Head and Tail plots. Values that differ from the myosin 2b value at the >95% (*) and >99% (**) significance level, based on z-scores (see Methods), are indicated. The dotted line in the ratio plot represents the value of 1 when both slopes are the same. Stars indicate the value is significantly different to one (* >95%; ** >99%). The double-headed arrows indicate the set of adult sarcomeric myosins.

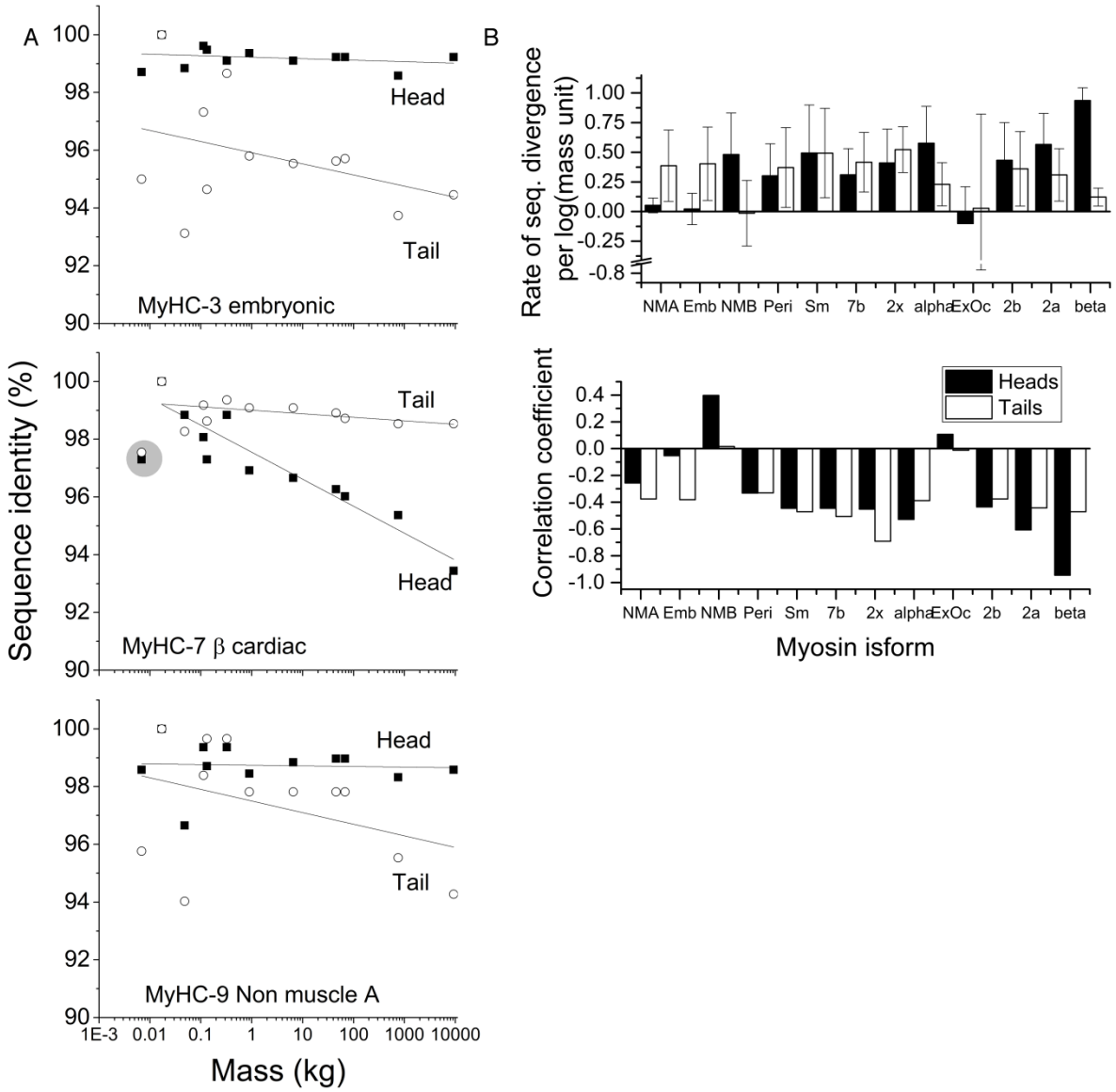

Figure 2. Dependence of myosin sequence divergence on the mass of the 12 mammalian species shown in Fig 1.

A) Sequence divergence data of the Head and Tails from Fig 1 is plotted against the log of the mass of the species and the correlation between the two is shown by the regression line. Brandt's Bat data has been excluded from the β-cardiac mass plot. Data for the myosin Head of each isoform are shown in filled squares and Tail domains shown as empty circles.

B) Upper panel: Rates of divergence per log mass unit (+SE) Lower panel: Correlation coefficients plotted for 12 myosin isoforms (individual data plots are shown in Supplementary Data, Fig S1). Filled bars are the data for the myosin heads empty bars are for the myosin Tails.

Figure 3

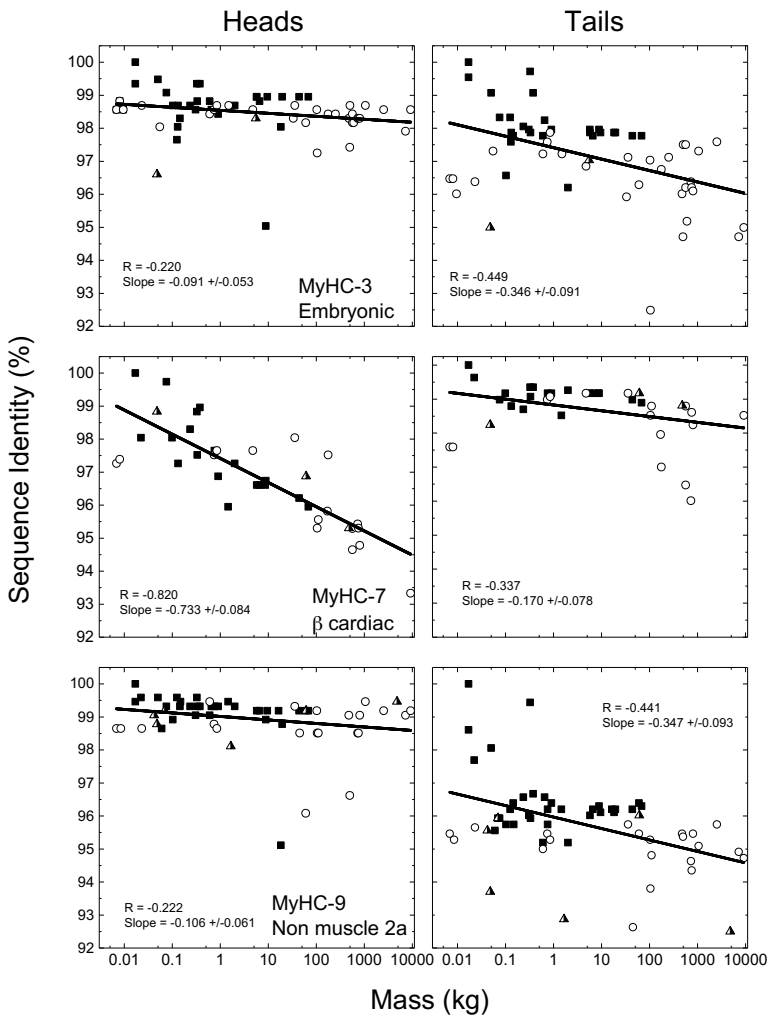

Figure 3 <u>The dependence of sequence identity on species mass for three myosin heavy chains.</u>

Percent sequence identity to the mouse sequence for three myosins as a function of species mass Details of the species mass and sequences used are given in Table S2. The data uses all complete sequences present in the data base in September 2015 and includes 59 MyHC3, 39 MyHC7and 60 MyHC9 sequences. Open circle are Laurasiatheria, filled squares are Euarchontoglires and half-filled triangles are Afrotheria or Metatheria. The regression lines are shown with the correlation coefficient R and the slopes in units of % identity/log mass unit.

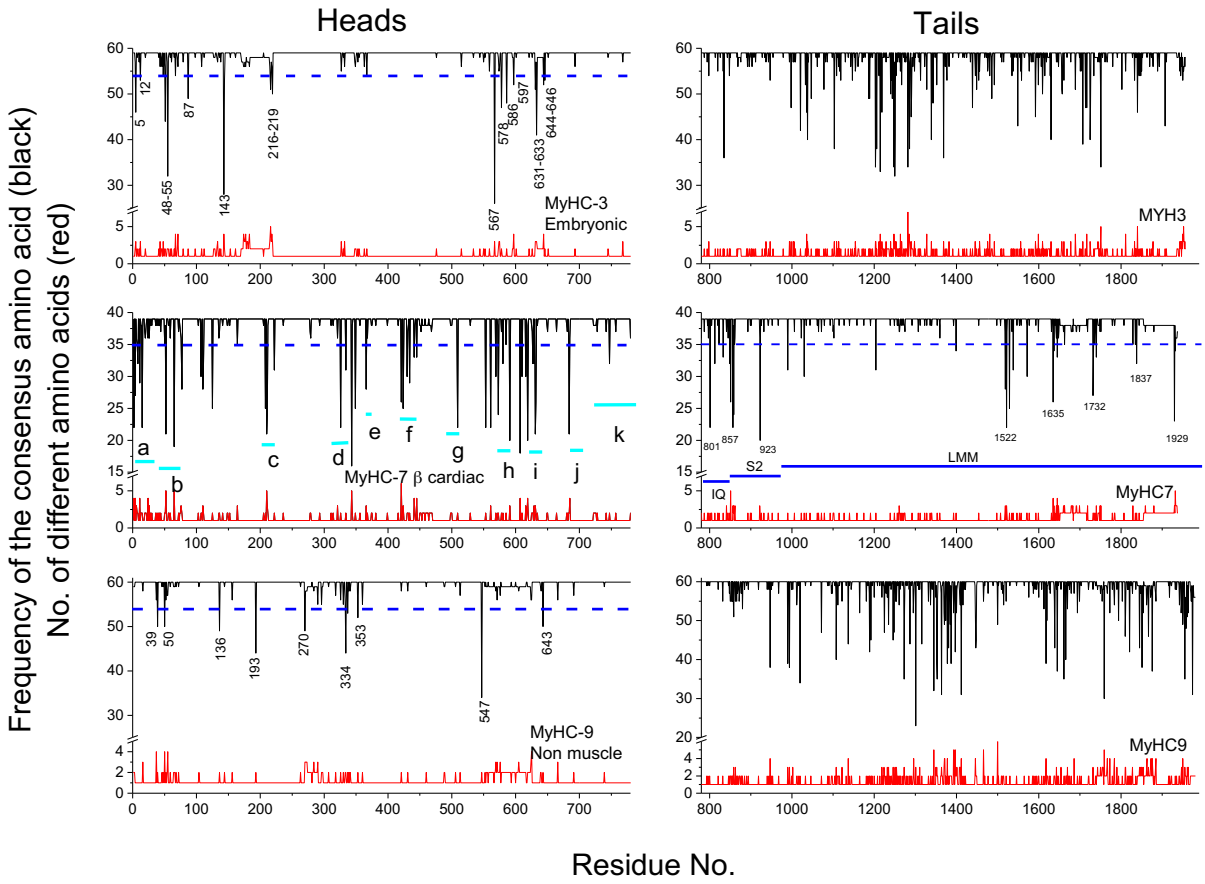

Fig 4 Location of non-conserved residues in the sequence of MYHC 3, 7 and 9. Black lines show the number of times the consensus amino acid occurs at each residue position in the sequence (maximum equals the no of sequences used MyHC-3, 59; MyHC-7, 39; MyHC-9, 60). Red shows the number of different amino acids occurring at that position in the sequence (minimum 1). The blue dashed line indicates a threshold for site with greater than 4 or 5 differences. The heads and tails are shown in different plots to allow an expanded scale for the heads. For the tail the residues in MyHC7 and the head region of MyHC3 & 9 numbers for regions with high variability are noted together with the 3 major domains of the tail; the lever arm or IQ domain, the S2 or subfragment-2 and LMM or Light Meromyosin. There are small offsets in sequence numbering between the 3 different myosins due small difference in the number of residues in head domain loops 1 & 2. For the β myosin (MyHC7) the plot for the head domain highlights (in cycan) key functional areas of the sequence as listed: a. N-terminus 1-36, b. SH3 domain 37-75, c. Loop 1 201-215, d. Drosophila exon 7 region 300-335, e. Loop 4 368-372, f. Helix-O 426-440, g. Relay helix/loop, h. Loop 3 558-573, i. Loop 2 615-637, j. SH helices 680-707, k. Converter 710-778

Species	Mass	202											214
Brandts bat	0.007	G	D	R	S	K	K	D	Q	N	T	G	K	G
Little brown bat	0.008	G	D	R	S	K	K	D	Q	N	T	G	K	G
Mouse	0.017	G	D	R	S	K	K	D	Q	T	P	G	K	G
Chinese hamster	0.022	G	D	R	S	K	K	D	Q	T	P	G	K	G
Opossum	0.048	G	D	R	S	K	K	E	Q	T	P	G	K	G
Kangaroo rat	0.076	G	D	R	S	K	K	D	Q	T	P	G	K	G
Golden hamster	0.099	G	D	R	S	K	K	E	Q	T	P	G	K	G
Tarsier	0.132	G	D	R	S	K	K	D	Q	N	P	G	K	G
Degu	0.235	G	D	R	S	K	K	D	Q	T	P	G	K	G
Rat	0.325	G	D	R	S	K	K	E	Q	A	T	G	K	G
Marmoset	0.33	G	D	R	S	K	K	D	Q	N	P	G	K	G
Mole-rat	0.375	G	D	R	S	K	K	D	Q	T	P	G	K	G
Black flying fox	0.74	G	D	R	S	K	K	D	Q	N	P	G	K	G
Night monkey	0.75	G	D	R	S	K	K	D	Q	N	P	G	K	G
Large fruit bat	0.85	G	D	R	S	K	K	D	Q	N	P	G	K	G
Guinea pig	0.9	G	D	R	S	K	K	E	Q	T	P	G	K	G
Flying Lemur	1.45	G	D	R	G	K	K	D	Q	G	P	G	K	G
Rabbit	2	G	D	R	S	K	K	D	Q	T	P	G	K	G
Cat	4.75	G	D	R	S	K	K	E	Q	T	P	G	K	G
Green monkey	5.75	G	D	R	S	K	K	D	Q	N	T	G	K	G
Crab eating macaque	6.55	G	D	R	S	K	K	D	Q	N	T	G	K	G
Pig-tailed macaque	6.55	G	D	R	S	K	K	D	Q	T	P	G	K	G
Mangabey	8.6	G	D	R	S	K	K	D	Q	N	T	G	K	G
Colobus	8.9	G	D	R	S	K	K	D	Q	N	T	G	K	G
Dog	35.5	G	D	R	S	K	K	D	Q	T	P	G	K	G
Bonobo	44	G	D	R	S	K	K	D	Q	G	P	G	K	G
Aardvark	61	G	D	R	S	K	K	E	Q	T	P	G	K	G
Human	68	G	D	R	S	K	K	D	Q	S	P	G	K	G
Dolphin	104.5	G	D	R	S	K	K	E	Q	T	S	G	K	G
Sheep	110	G	D	R	S	K	K	E	Q	T	P	G	K	G
Pig	169	G	D	R	S	K	K	D	Q	N	T	G	K	G
Siberian tiger	176.4	G	D	R	S	K	K	D	Q	T	P	G	K	G
Manatee	480	G	D	R	S	K	K	E	Q	T	T	G	K	G
Yak	562.5	G	D	R	S	K	K	E	Q	A	T	G	K	G
Horse	563.5	G	D	R	S	K	K	D	Q	T	S	G	K	G
Water buffalo	725	G	D	R	S	K	K	E	Q	T	T	G	K	G
Cattle	755	G	D	R	S	K	K	E	Q	A	T	G	K	G
Camel	800	G	D	R	S	K	K	E	Q	T	P	S	K	G
Minke whale	9200	G	D	R	S	K	K	E	Q	T	S	G	K	G

Figure 5. Variation in the β-myosin loop 1 region.

Loop one is shown for the 39 β-myosin sequences from smallest to greatest body mass (kg). The loop contains three positions (208, 210 and 211) where there is considerable variation and there is a trend for one amino acid in smaller species and a different one in larger species.