

# Myosin gene mutation correlates with anatomical changes in the human lineage

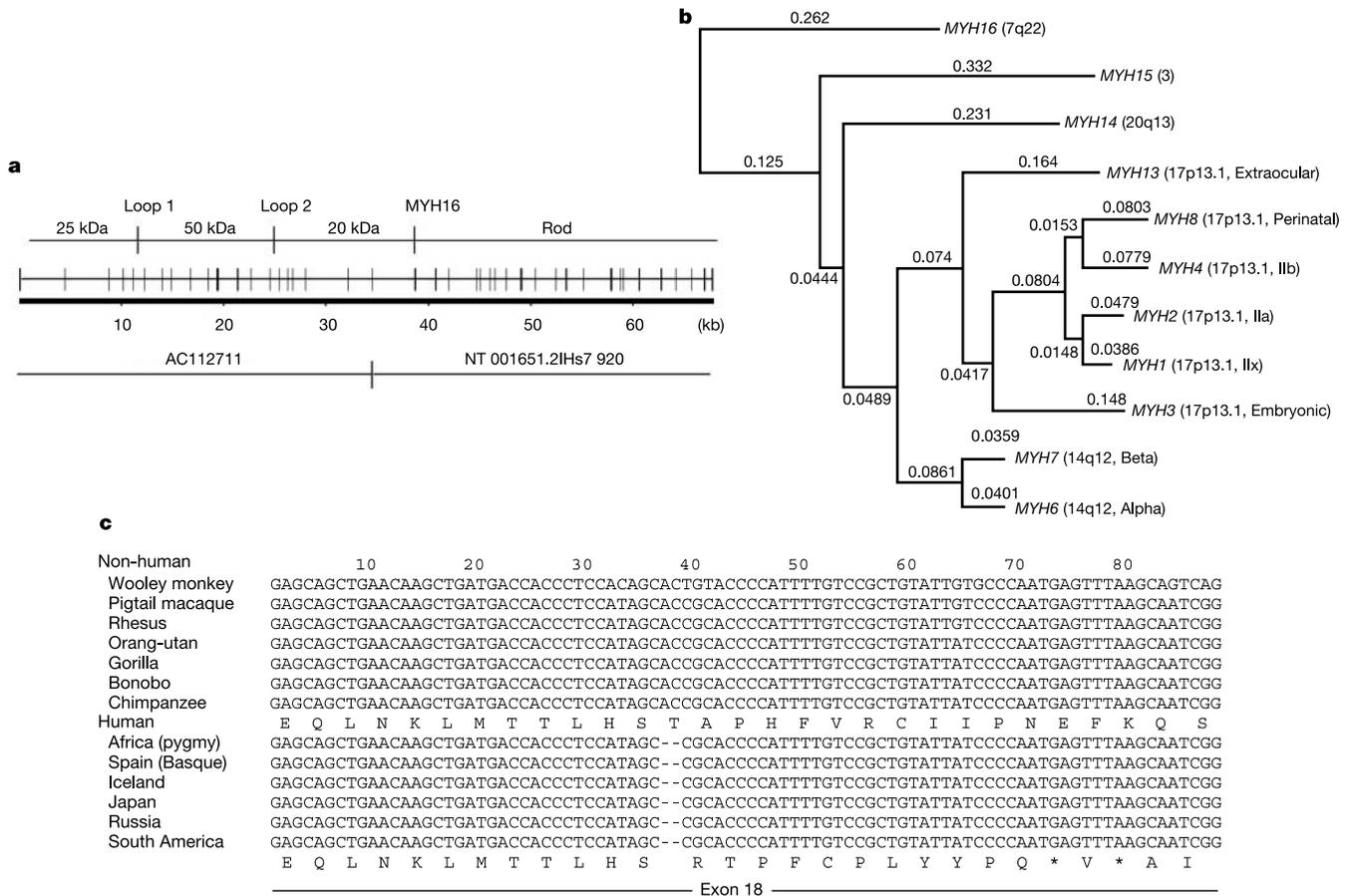
Hansell H. Stedman<sup>1,3</sup>, Benjamin W. Kozyak<sup>1</sup>, Anthony Nelson<sup>1</sup>, Danielle M. Thesier<sup>2</sup>, Leonard T. Su<sup>1</sup>, David W. Low<sup>1,5</sup>, Charles R. Bridges<sup>1</sup>, Joseph B. Shrager<sup>1,3</sup>, Nancy Minugh-Purvis<sup>2,4,5</sup> & Marilyn A. Mitchell<sup>1</sup>

<sup>1</sup>Department of Surgery and <sup>2</sup>Cell and Developmental Biology, <sup>3</sup>the Pennsylvania Muscle Institute, School of Medicine, and <sup>4</sup>Department of Anatomy and Cell Biology, School of Dental Medicine, University of Pennsylvania, and <sup>5</sup>Division of Plastic Surgery, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA

Powerful masticatory muscles are found in most primates, including chimpanzees and gorillas, and were part of a prominent adaptation of *Australopithecus* and *Paranthropus*, extinct genera of the family Hominidae<sup>1,2</sup>. In contrast, masticatory muscles are considerably smaller in both modern and fossil members of *Homo*. The evolving hominid masticatory apparatus—traceable to a Late Miocene, chimpanzee-like morphology<sup>3</sup>—shifted towards a pattern of gracilization nearly simultaneously with accelerated encephalization in early

*Homo*<sup>4</sup>. Here, we show that the gene encoding the predominant myosin heavy chain (*MYH*) expressed in these muscles was inactivated by a frameshifting mutation after the lineages leading to humans and chimpanzees diverged. Loss of this protein isoform is associated with marked size reductions in individual muscle fibres and entire masticatory muscles. Using the coding sequence for the myosin rod domains as a molecular clock, we estimate that this mutation appeared approximately 2.4 million years ago, predating the appearance of modern human body size<sup>5</sup> and emigration of *Homo* from Africa<sup>6</sup>. This represents the first proteomic distinction between humans and chimpanzees that can be correlated with a traceable anatomic imprint in the fossil record.

We obtained a DNA sequence by degenerate polymerase chain reaction (PCR) that suggested the existence of a hitherto unrecognized human sarcomeric myosin gene (*MYH16*, see Supplementary Information). An unannotated accession file posted during the finalization of the human genome sequence (AC112711) contained the only exact match to the query sequence on a BLAST search. Here, we use homology to other sarcomeric myosins to annotate this locus and generate the complete sequence of a hypothetical human sarcomeric myosin distantly related to the other known members of this subfamily (Fig. 1; see also Supplementary Information). This reconstruction suggested that a frameshift deletion at codon 660 of the messenger RNA of the deduced polypeptide was



**Figure 1** Molecular evolution of *MYH16*. **a**, Distribution of 42 predicted coding exons spanning 67,983 base pairs (bp) in the region of human chromosome 7q22 flanked 5' by *SMURF1* and 3' by *ARPC1A*. **b**, Phylogenetic reconstruction for all human sarcomeric myosin genes (heavy chain), showing early divergence of *MYH16* from others. Branch lengths shown are derived from a maximum likelihood analysis of the aligned cDNAs, beginning with the conserved proline at the head-rod junction. Non-sarcomeric class II

myosins (designated *MYH9*, -10 and -11; data not shown) are used to root the tree. **c**, Aligned DNA sequences for *MYH16* exon 18 representing seven non-human primate species and six geographically dispersed human populations, revealing the effect of frameshift on reading frame and deduced amino acid sequence. Note stop codon at position 72–74.

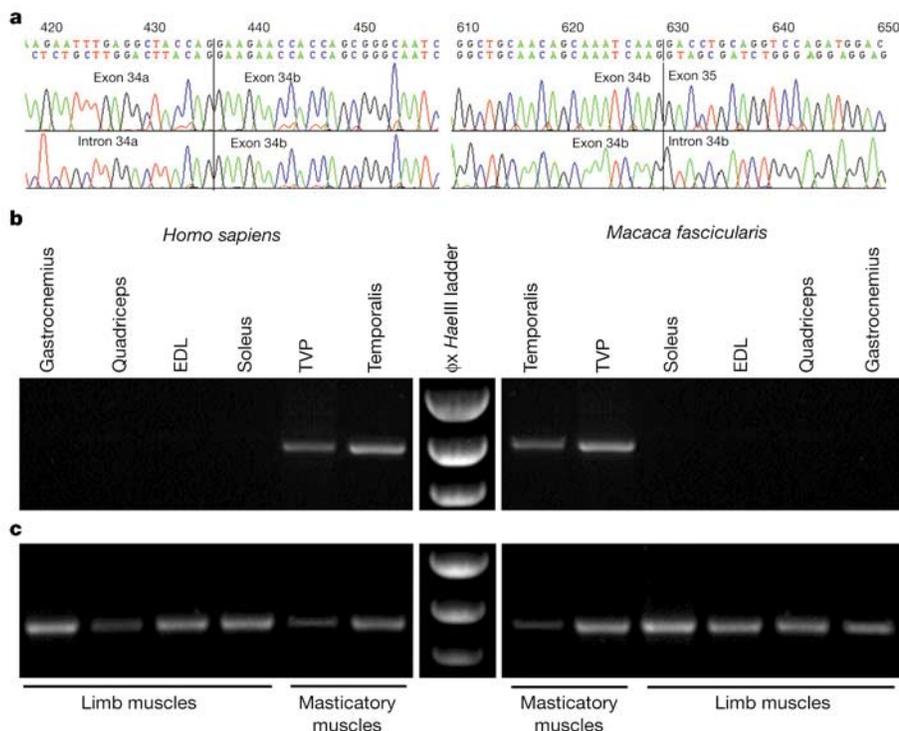
either a rare allele or a sequencing artefact. To resolve this uncertainty, we sequenced this region in DNA samples from members of geographically disparate human populations. As shown in Fig. 1c, the mutation is found in all modern humans sampled, including natives of Africa, South America, Western Europe, Iceland, Japan and Russia; thus, the inactivating mutation seems to be fixed in *Homo sapiens*. In contrast, all of the non-human primates for which sequence was obtained have an ACC codon that encodes a highly conserved threonine. The frameshift in the human coding sequence truncates the predicted 224-kDa myosin heavy chain to a 76-kDa fragment containing an unstable portion of the myosin head domain (Fig. 1c).

Animals homozygous for null mutations in sarcomeric myosin heavy chain genes sometimes possess severe functional defects in individual muscles that ordinarily accumulate the highest levels of specific isomyosins<sup>7,8</sup>. In reconstructing the pattern of expression of the *MYH16* gene in a recent common ancestor, scarcity of chimpanzee tissue led us to examine transcription in a wide range of muscles from macaque (*Macaca fascicularis*) and modern humans. Cross-species sequence similarity facilitated the design and use of several isoform-specific RT-PCR primers, with the finding that the amplification products uniquely contain the sequences predicted by our human *MYH16* gene annotation (Fig. 2a). Transcription is detectable only in the muscles of the head, specifically those derived from the embryonic first pharyngeal arch, including temporalis and tensor veli palatini (Fig. 2b).

In gross anatomical comparisons between humans versus great apes and monkeys, the relative size of individual masticatory muscle homologues contrasts remarkably, as is evident in Fig. 3a–k. The relative sizes of the temporalis and masseter muscles are reflected in the morphology of such craniofacial features as the temporal fossa and zygomatic arch (highlighted). Figure 3a–i illustrates the difference in these sites of bony attachment in modern macaque, gorilla and human skulls. At the histological level, the difference between

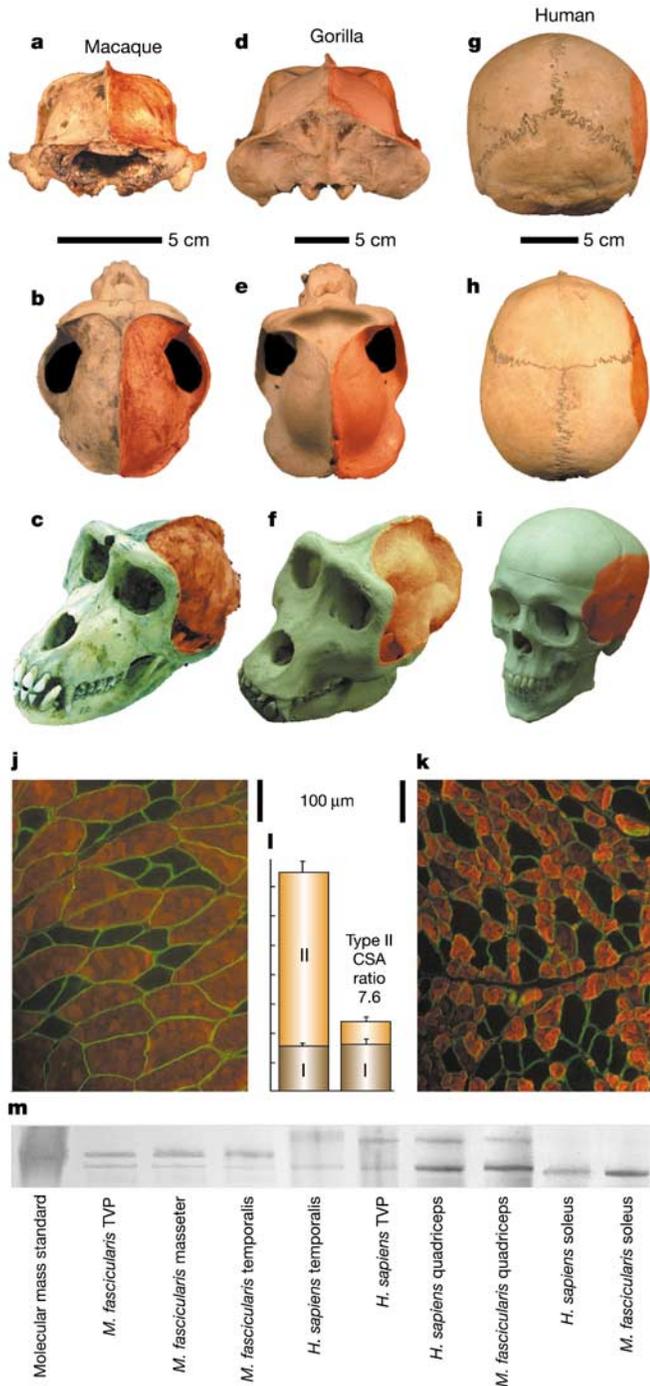
human and non-human primate temporalis muscle is highlighted by staining for type II (all fast twitch) sarcomeric myosin and interstitial laminin (Fig. 3j–l). Although the unstained type I/slow fibres are nearly identical in cross-sectional area, as confirmed in a reciprocal stain (data not shown), the type II fibres of *H. sapiens* are about one-eighth the size of those of *M. fascicularis*. The relative hypotrophy of the type II fibres resembles that seen in limb muscle in inclusion body myopathy-3 (IBM3)—a rare disease for which the only proven molecular aetiology is a myosin gene mutation (*MYH2*)<sup>9</sup>. Protein gel electrophoresis shows that the dominant myosin heavy chain isoform in the temporalis of *M. fascicularis* is specific to the masticatory muscles and is undetectable in the human temporalis (Fig. 3m). Peptide sequencing of this protein using mass spectrometry unambiguously establishes it as the product of the *M. fascicularis* *MYH16* orthologue (Supplementary Information), whereas the major isomyosins of the human temporalis have recently been identified as products of the *MYH1*, -2 and -7 genes<sup>10</sup>.

The fact that the human *MYH16* locus is still appropriately transcribed indicates that the coding sequence deletion identified here was not preceded by a silencing mutation in a transcriptional control domain. We postulate that the volume of the skeletal muscle fibres expressing the *MYH16* gene transcript is proportional to the total amount of myosin heavy chain accumulating in the cell, and that reliance on translation of the less abundant *MYH1* and -2 transcripts in the face of a frameshift mutation in *MYH16* has resulted in the eightfold reduction in the size of the type II fibres in the human masticatory muscles as compared with macaque. Genetic manipulation of muscle size has marked secondary effects on the anatomy of bony attachment sites, as illustrated by recent studies of myostatin signalling<sup>11,12</sup>. Moreover, experimental animal models of masticatory muscle resection or transposition have demonstrated the correlation between craniofacial morphology and the force of masticatory muscle contraction<sup>13,14</sup>. We reason



**Figure 2** Transcription of *MYH16*. **a**, Alignment of representative genomic and cDNA sequences demonstrate that *MYH16* is a transcribed and properly spliced pseudogene in *H. sapiens*. **b**, RT-PCR detects transcription of *MYH16* and its *M. fascicularis* orthologue only in muscles derived from the first pharyngeal arch (denoted as masticatory muscles).

**c**, Control amplifications were performed using identical first-strand cDNA template and a pair of 'universal' sarcomeric myosin heavy chain primers based on an *MYH* consensus sequence. EDL, extensor digitorum longus; TVP, tensor veli palatini;  $\phi$ X HaeIII ladder,  $\phi$ X genomic DNA digested with the restriction enzyme HaeIII.

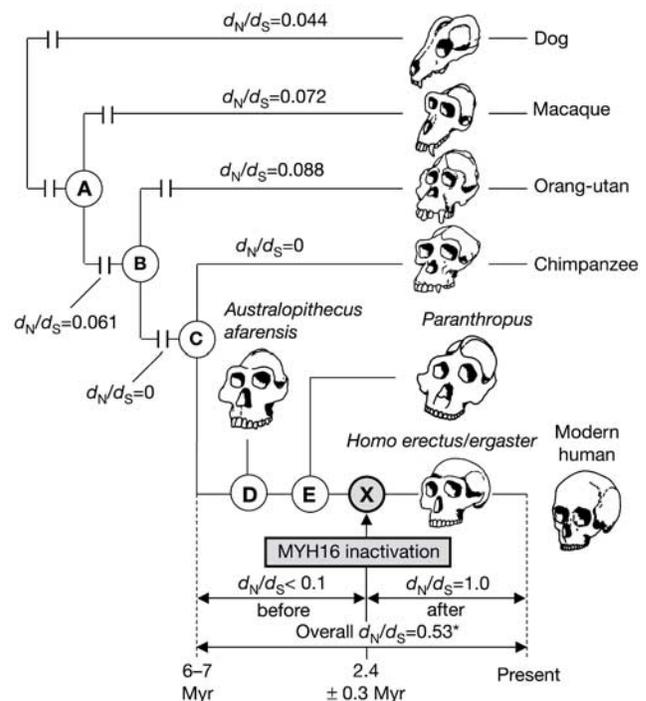


**Figure 3** Selective loss of type II myofibre size and muscle bulk in the human temporalis. **a–i**, Occipital, superior and oblique views showing temporal fossae (highlighted in red) and temporalis foramina (circumscribed by the zygomatic arch) in skulls from *M. fascicularis* (**a–c**), *Gorilla gorilla* (**d–f**) and *H. sapiens* (**g–i**). **j, k**, Immunofluorescently stained cryosections of temporalis from *M. fascicularis* (**j**) and *H. sapiens* (**k**). Magnification,  $\times 100$ . Rhodamine (red) fluorescence indicates type II myofibres, fluorescein (green) reveals interstitial laminin at fibre perimeter. Note that unstained type I fibres are the same size in both species. **l**, Quantification of relative cross-sectional area (CSA) by fibre type. **m**, Coomassie-blue-stained gel electrophoretograms of myosins from limb and first pharyngeal arch muscles of *M. fascicularis* and *H. sapiens*, showing an isomyosin of unique mobility in the first pharyngeal arch derivatives of *M. fascicularis*, midway between MYH1 and -2 bands co-migrating at the top and MYH7 at the bottom<sup>10</sup>.

that an abrupt evolutionary alteration in the size and contractile force generated by these muscles would have had pleiotropic effects on craniofacial morphology in the first homozygous *MYH16*-null human ancestor. Such a reduction would probably be inconsistent with the temporalis hypertrophy implicated in sagittal cresting and zygomatic flaring in non-human primates as well as in *Paranthropus* and *Australopithecus*. It is also likely that diminished contractile force would translate into a reduction in stress across patent sutures, sites of dura-mater-patterned growth in the immature neurocranium<sup>15</sup>.

To estimate the age of the gene inactivation, we aligned orthologous *MYH16* sequences for human, chimpanzee, orang-utan, macaque and dog, and deduced ancestral sequences (Fig. 4). The high ratio of silent (synonymous) to amino-acid-altering (non-synonymous) mutations indicates that the *MYH16* gene evolved under negative (purifying) darwinian selection in all ancestral lineages except that leading directly to *H. sapiens*<sup>16</sup>. The loss of selective constraint after gene inactivation allows non-synonymous mutations to accumulate at the neutral mutation rate<sup>17</sup>. Under the assumption that the neutral mutation rate has remained constant since the human–chimpanzee divergence at 6–7 million years (Myr) ago<sup>17,18</sup>, application of the formula of ref. 17 yields a time of  $2.4 \pm 0.3$  Myr for the appearance of the inactivating mutation in a hominid ancestor (see Supplementary Information).

The earliest hominid fossil record, restricted to the Late Miocene epoch of Africa, is comprised of a variety of taxa featuring small, chimpanzee-sized braincases and large masticatory complexes<sup>18–21</sup>. Emphasis on a robust masticatory adaptation continues through the Pliocene epoch in various species of *Australopithecus* and *Paranthropus*<sup>1,22</sup>. In contrast, a relatively gracile masticatory



**Figure 4** Divergence of the *MYH16* orthologue. Data are based on nucleotide sequences for the six largest exons in the rod-coding domain. Individual mutations are indicated by N for non-synonymous or S for synonymous.  $d_N$  and  $d_S$  refer to the mutational rates normalized to the number of relevant sites (840 N, 225 S) in pairwise comparisons between *MYH16* orthologues of extant species and inferred ancestors at nodes A, B and C. The atypically high  $d_N/d_S$  in the human ancestral lineage (asterisk,  $P = 0.0527$ ) indicates a loss of negative darwinian selection approximately 2.4 Myr ago (see Supplementary Information).

apparatus appears in *Homo erectus/ergaster* by 1.8–2.0 Myr<sup>5</sup>. Coinciding with this loss of masticatory strength is the marked increase in cranial capacity<sup>4,5</sup>, which characterizes the Pleistocene evolution of *Homo*. Our findings on the age of the inactivating mutation in the *MYH16* gene raise the intriguing possibility that the decrement in masticatory muscle size removed an evolutionary constraint on encephalization, as suggested by the anatomy of the muscle attachments relative to the sutures (compare Fig. 3a and d with Fig. 3g). □

## Methods

### Bioinformatics

The *MYH16* locus was initially recognized by PCR of human genomic DNA using degenerate primers targeting the coding sequences for two highly conserved  $\alpha$ -helices flanking 'loop 1' at the myosin 25/50-kDa junction<sup>23,24</sup>. As soon as a draft DNA sequence spanning this sequence was posted online (AC112711) a complete representation of the human *MYH16* locus was constructed by integrating this information for the head domain of the gene into the suite of computational algorithms described in ref. 24. The deduced complementary DNA and peptide sequences obtained by restoring an open reading frame across codon 660 (conforming to the primate consensus shown in Fig. 1) were then compared to sequences for all other class II myosin heavy chains encoded by the human genome using the ClustalW alignment algorithm in MacVector 7.1.1 (Accelrys). Phylogenetic reconstruction used the maximum likelihood execution within the program PAUP\* 4.0 beta<sup>25</sup>. All other methods, as detailed in Supplementary Information, gave a topologically similar tree with this input data set. Ancestral sequence reconstruction was achieved independently by using the maximum parsimony (PAMP) and maximum likelihood (BASEML) programs within the most recent release of the program PAML<sup>26</sup>; the results were identical, reflecting the degree of conservation among the aligned sequences. The program suite MEGA version 2.1 (ref. 27) was used to calculate  $d_N/d_S$  (see Fig. 4 legend for definition) ratios and relevant statistics on a per lineage basis.

### Messenger RNA isolation, PCR and sequence analysis

Messenger RNA was isolated using the MicroFastTrack 2.0 mRNA isolation kit (Invitrogen). Complementary DNA was synthesized using the Retroscript first strand synthesis kit (Ambion). *MYH16* RT-PCR amplifications were performed using HotStarTaq DNA Polymerase (Qiagen). Genomic DNA samples obtained from the Coriell Genetic Cell Repositories and the Southwest Foundation National Primate Research Center were amplified by PCR using the primer pairs and annealing temperatures listed in Supplementary Information. After purification of PCR products using Qiagen QIAquick gel extraction kit, according to the manufacturer's specifications, bidirectional DNA sequences were generated by the University of Pennsylvania School of Medicine, Department of Genetics Sequencing Core, after fluorimetric analysis of dideoxynucleotide chain termination products.

### Protein gel electrophoresis and peptide analysis

Myofibrils from human and cynomolgus macaque tissues were isolated and washed as described<sup>28</sup>. Polyacrylamide gel electrophoresis was performed as described<sup>29</sup>, using 4  $\mu$ g of total protein per lane. Myosin isoforms were visualized by staining gels with Bio-safe Coomassie blue stain (BioRad). A gel slice containing the dominant myosin heavy chain isoform from the *M. fascicularis* temporalis muscle was subjected to trypsin proteolysis and tandem mass spectrometry peptide sequence analysis at the Wistar Institute Proteomics Facility. The analysis was performed using microcapillary reverse-phase high-performance liquid chromatography nano-spray tandem mass spectrometry on a ThermoFinnigan LCQ quadrupole ion-trap mass spectrometer. The tandem mass spectrometry spectra were run against the non-redundant human protein sequence database and the predicted products of the newly recognized *MYH14*, -15 and -16 genes using the SEQUEST software package<sup>30</sup>.

### Tissue procurement

All human muscle samples were obtained under protocol number 203200 as approved by the Hospital of the University of Pennsylvania Institutional Review Board. All primate tissues were obtained under protocol number 706976 as approved by the University of Pennsylvania Animal Care and Use Committee.

### Antibody staining protocol

Frozen tissues were sectioned at a thickness of 8  $\mu$ m on a Leica cryotome. Sections were blocked with 2% filter-sterilized bovine serum albumin (BSA) for 3.5 h. BSA solution was removed without washing and replaced with a solution containing mouse-derived primary antibody against 'Fast' (Sigma, M-4276; dilution 1:800), 'Slow' (Sigma, M-8421; dilution 1:5,000), or IIA (SC-71; dilution 1:100) myosin heavy chain protein, as well as rabbit-derived primary antibody against laminin (Sigma, L-9393; dilution 1:2,500), and incubated overnight at 4 °C. Sections were then washed three times with 1  $\times$  PBS and incubated for 1 h in a secondary antibody solution containing Cy3-conjugated goat anti-mouse IgG (Jackson ImmunoResearch, 113-165-100; dilution 1:200) and fluorescein-isothiocyanate-conjugated sheep anti-rabbit IgG (Sigma, F-7512; dilution 1:160). After secondary antibody incubation, sections were washed three times in 1  $\times$  PBS and mounted with Vectashield containing 4,6-diamidino-2-phenylindole. Photographs were taken at  $\times$ 100 total magnification using a Zeiss Axiophot2 microscope equipped with a Nikon DXM1200 digital camera.

## Image analysis

Fibre size was determined using Image Pro Plus (Phase 3 Imaging Systems), running a macro optimized to calculate fibre area based on a laminin stain.

Received 5 April 2003; accepted 20 January 2004; doi:10.1038/nature02358.

1. Rak, Y. *The Australopithecine Face* (Academic, New York, 1983).
2. Aiello, L. & Dean, C. *An Introduction to Human Evolutionary Anatomy* (Academic, New York, 1998).
3. White, T. D., Suwa, G., Simpson, S. & Asfaw, B. Jaws and teeth of *Australopithecus afarensis* from Maka, Middle Awash, Ethiopia. *Am. J. Phys. Anthropol.* **111**, 45–68 (2000).
4. Tobias, P. V. *The Skulls, Endocasts, and Teeth of Homo habilis* (University Press, Cambridge, 1991).
5. Walker, A. & Leakey, R. (eds) *The Nariokotome Homo erectus skeleton* (Harvard Univ. Press, Cambridge, Massachusetts, 1993).
6. Vekua, A. *et al.* A new skull of early *Homo* from Dmanisi, Georgia. *Science* **297**, 85–89 (2002).
7. Acakpo-Satchivi, L. *et al.* Growth and muscle defects in mice lacking adult myosin heavy chain genes. *J. Cell Biol.* **139**, 1219–1229 (1997).
8. Allen, D. L., Harrison, B. C., Sartorius, C., Byrnes, W. C. & Leinwand, L. A. Mutation of the IIB myosin heavy chain gene results in muscle fiber loss and compensatory hypertrophy. *Am. J. Physiol. Cell Physiol.* **280**, C637–C645 (2001).
9. Martinsson, T. *et al.* Autosomal dominant myopathy: missense mutation (Glu-706  $\rightarrow$  Lys) in the myosin heavy chain IIa gene. *Proc. Natl Acad. Sci. USA* **97**, 14614–14619 (2000).
10. Korfage, J. A. & Van Eijden, T. M. Myosin heavy chain composition in human masticatory muscles by immunohistochemistry and gel electrophoresis. *J. Histochem. Cytochem.* **51**, 113–119 (2003).
11. McPherron, A. C., Lawler, A. M. & Lee, S. J. Regulation of skeletal muscle mass in mice by a new TGF- $\beta$  superfamily member. *Nature* **387**, 83–90 (1997).
12. Hamrick, M. W., McPherron, A. C., Lovejoy, C. O. & Hudson, J. Femoral morphology and cross-sectional geometry of adult myostatin-deficient mice. *Bone* **27**, 343–349 (2000).
13. Hohl, T. H. Masticatory muscle transposition in primates: effects on craniofacial growth. *J. Maxillofac. Surg.* **11**, 149–156 (1983).
14. Brennan, M. & Antonyshyn, O. The effects of temporalis muscle manipulation on skull growth: an experimental study. *Plast. Reconstr. Surg.* **97**, 13–24 (1996).
15. Warren, S. M., Brunet, L. J., Harland, R. M., Economides, A. N. & Longaker, M. T. The BMP antagonist noggin regulates cranial suture fusion. *Nature* **422**, 625–629 (2003).
16. Messier, W. & Stewart, C. B. Episodic adaptive evolution of primate lysozymes. *Nature* **385**, 151–154 (1997).
17. Chou, H. H. *et al.* Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. *Proc. Natl Acad. Sci. USA* **99**, 11736–11741 (2002).
18. Brunet, M. *et al.* A new hominid from the Upper Miocene of Chad, Central Africa. *Nature* **418**, 145–151 (2002).
19. Haile-Selassie, Y. Late Miocene hominids from the Middle Awash, Ethiopia. *Nature* **412**, 178–181 (2001).
20. Leakey, M. G., Feibel, C. S., McDougall, I., Ward, C. & Walker, A. New specimens and confirmation of an early age for *Australopithecus anamensis*. *Nature* **393**, 62–66 (1998).
21. Lockwood, C. A., Kimbel, W. H. & Johanson, D. C. Temporal trends and metric variation in the mandibles and dentition of *Australopithecus afarensis*. *J. Hum. Evol.* **39**, 23–55 (2000).
22. Asfaw, B. *et al.* *Australopithecus garhi*: a new species of early hominid from Ethiopia. *Science* **284**, 629–635 (1999).
23. Uyeda, T. Q. P., Ruppel, K. M. & Spudich, J. A. Enzymatic activities correlate with chimaeric substitutions at the actin-binding face of myosin. *Nature* **368**, 567–569 (1994).
24. Desjardins, P., Burkman, J., Shrager, J., Allmond, L. & Stedman, H. Evolutionary implications of three novel members of the human sarcomeric myosin heavy chain gene family. *Mol. Biol. Evol.* **19**, 375–393 (2002).
25. Swofford, D. *PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods)* (Sinauer Associates, Sunderland, Massachusetts, 1998).
26. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *CABIOS* **13**, 555–556 (1997).
27. Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244–1245 (2001).
28. Bamman, M. M., Clarke, M. S., Talmadge, R. J. & Feeback, D. L. Enhanced protein electrophoresis technique for separating human skeletal muscle myosin heavy chain isoforms. *Electrophoresis* **20**, 466–468 (1999).
29. Levine, S., Kaiser, L., Lefterovich, J. & Tikunov, B. Cellular adaptations in the diaphragm in chronic obstructive pulmonary disease. *N. Engl. J. Med.* **337**, 1799–1806 (1997).
30. MacCoss, M. J., Wu, C. C. & Yates, J. R. Probability-based validation of protein identifications using a modified SEQUEST algorithm. *Anal. Chem.* **74**, 5593–5599 (2002).

Supplementary Information accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank K. Brayman, N. Mirza, M. Ruckenstein and the University of Washington National Primate Research Center (Seattle) for providing access to the biopsy material used in this study; N. Gilmore of the Philadelphia Academy of Natural Sciences for access to specimens for photography; L. Joseph, D. Fonseca, R. McCourt and W. Ewens for assistance with the bioinformatic analysis; and P. Dodson, L. Whitaker, A. Kelly and S. Bartlett for advance reading of the manuscript. We also thank colleagues in the University of Pennsylvania Genomics Institute Bioinformatics Core, and Wistar Institute Proteomics Facility for their assistance. This work was supported in part by grants to H.H.S. from the NIH (NIAMS and NINDS), MDA, AFM, VA and the Genzyme Corporation.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to H.H.S. (hstedman@mail.med.upenn.edu).