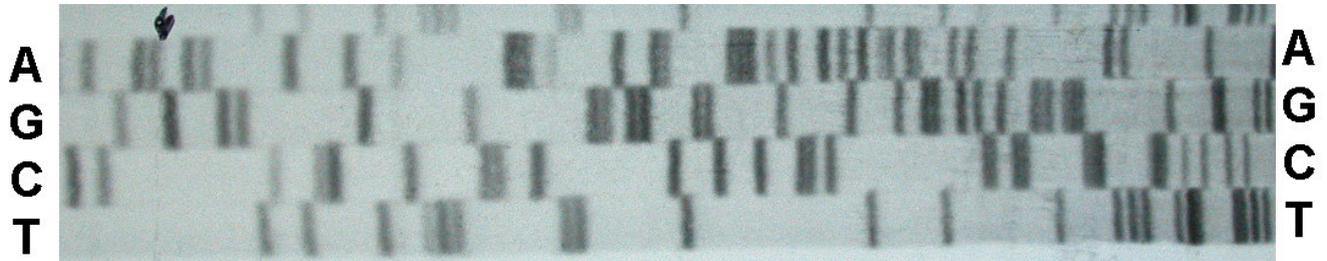


The DNA sequence for a unique 1.1 kb amplification product exactly matched bp 21209-21318 (autoradiograph shown) and 12211-12310 (autoradiograph not shown) of accession number AC112711 (arrow below denotes 3'-end of sense primer)..



Sense Primer -> G A G K T V N T K
CGCCGCGGGIAARACIGTIAAYACIAA... ->

10 20 30 40 50 60 70 80
K V I Q Y F A N I G G T G K Q T T D K K
GAAGGTCATCCAGTACTTTGCCAACATTGGAGGAACTGGCAAACAGACCACAGATAAGAAGgtagagccgaccgggtgggccc
GAAGGTCATCCAGTACTTTGCCAACATTGGAGGAACTGGCAAACAGACCACAGATAAGAAGgtagagccgaccgggtgggccc
K V I Q Y F A N I G G T G K Q T T D K K
10 21220 21230 21240 21250 21260 21270 21280 21290

90 100 110
atatttcccgtcttcaggcttctgtga... intron 7 approximately 1000 bp
atatttcccgtcttcaggcttctgtga...
21300 21310 21320

10 20 30 40 50 60 70 80
...ggctcccaggcagcagccaactggtcaactgactttgggtttctttctgcatcagGGGTCTCTGGAGGATCAAGTCATCC
...ggctcccaggcagcagccaactggtcaactgactttgggtttctttctgcatcagGGGTCTCTGGAGGATCAAGTCATCC
G S L E D Q V I
G S L E D Q V I
22240 22250 22260 22270 22280 22290 22300 22310

90 100
Q A N P V L E>
AGGCAAACCCTGTGCTGGAG...
AGGCAAACCCTGTGCTGGAG...
Q A N P V L E>
22320 22330

A F G N A K T V R **Antisense primer**
<- GCN TTY GGN AAY GCN AAR CAN GTN NGN
3'-CGI AAR CCI TTR CGI TTY TGC CAT GGCG-5'

(2) Other PCR Primers and annealing temperatures

MYH16 RTPCR (T_m=57°C)

forward CGGCTCAATCAAATCCTACGG,

reverse CTGGCATCCTCGTCCATCTG;

MYH RTPCR positive control (T_m=47°C)

forward GAGGCAAAAAGCGCCAGGA,

reverse TTGGTGAAGTTGATGCACAGCT.

Genomic DNA

Exon 18 (T_m=52°C) forward GTTGAGGTTTCTCTCAGAGCCTTG,

reverse CATGTGGGTGCTCTGCAACATC;

Exon 23 (T_m=54°C) forward TCTTGGTGTGGTCTTTGC, reverse TAGGGGCTTTAGGTATGG;

Exon 25 (T_m=48°C) forward GGTTCTTCTTGCCTTCTGAC,

reverse TGCCCACCAGGTAATGTGTG;

Exon 30 (T_m=48°C) forward CCGATGGTTTGGATTGTG, reverse TAAAGCAGCCTGTGAACG;

Exon 31 (T_m=48°C) forward GCTG-TGGCTGTCCTGTAAAC,

reverse TCTTCCTGATGACCCCAGAC;

Exon 34b (T_m=48°C) forward TGTGACGGTTTTTCCATTAC,

reverse TGAGTTAGCCCCCTTTAGG;

Exon 37 (T_m=53°C) forward TAGAGCACCCCTTTCCACCAAAC,

reverse AAGACAGGCATCTCACACACATACC.

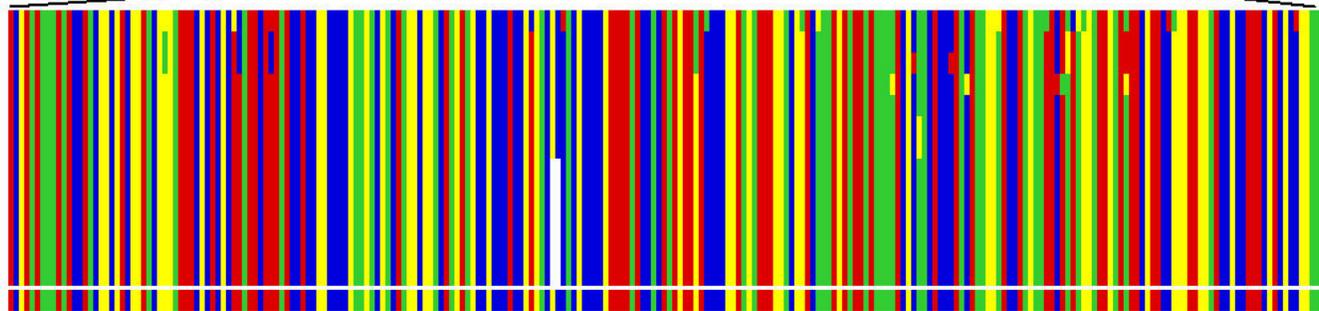
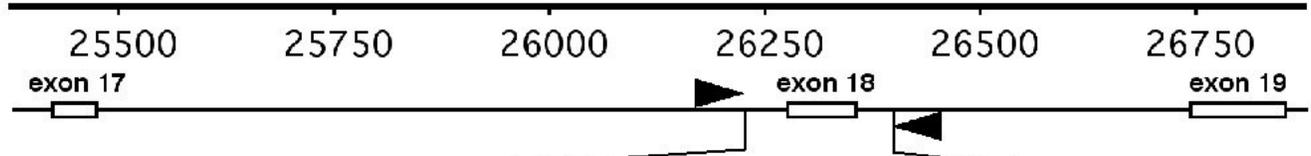
(3) Annotation of genomic DNA sequence for the MYH 16 Pseudogene

The *MYH16* locus in the human genome has been annotated as a Third Party Annotation (TPA) and is available under the GenBank TPA accession number: BK001410. Exon 18 sequences from the primate species shown below have been assigned the following GenBank accession numbers:

Chimpanzee (<i>Pan troglodytes</i>)	AY350722
Bonobo (<i>Pan paniscus</i>)	AY350721
Gorilla (<i>Gorilla gorilla</i>)	AY350720
Orangutan (<i>Pongo pygmaeus</i>)	AY350719
Pigtailed Macaque (<i>Macaca nemestrina</i>)	AY350717
Rhesus Macaque (<i>Macaca mulatta</i>)	AY350718
Woolley Monkey (<i>Lagothrix lagotricha</i>)	AY350716

(4) Alignment of Human and Non-human Primate DNA sequences for MYH 16 exon 18 and portions of the flanking introns

MYH 16, Position relative to start codon

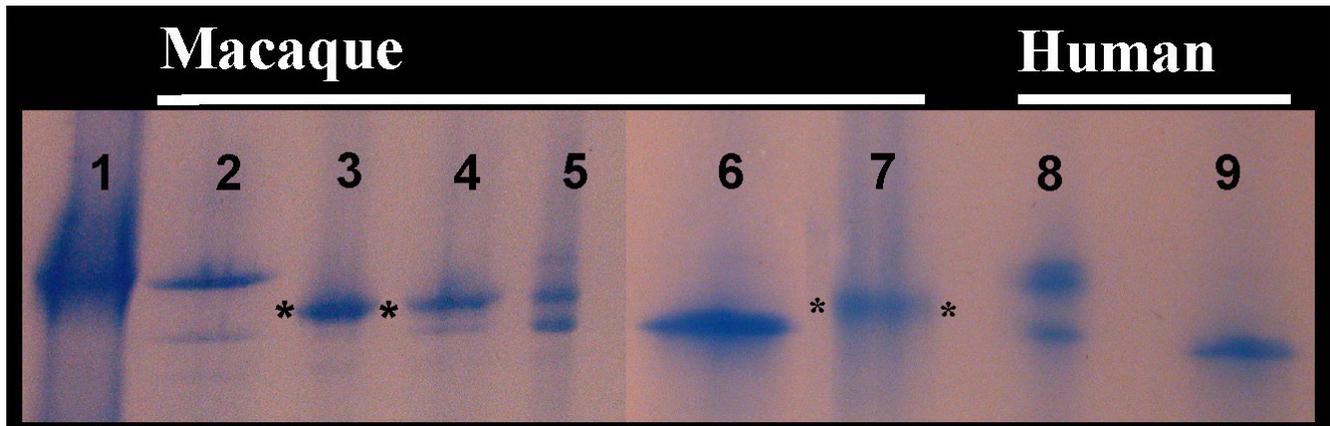


	28	38	48	58	68																																											
WoolleyMonkey	A	A	G	C	T	G	A	T	G	A	C	C	A	C	C	C	T	C	C	A	C	A	G	C	A	C	T	G	C	A	C	C	C	A	T	T	T	G	T	C	C	G	C	T	G	T	A	T
Pigtail Mac	A	A	G	C	T	G	A	T	G	A	C	C	A	C	C	C	T	C	C	A	T	A	G	C	A	C	C	G	C	A	C	C	C	A	T	T	T	G	T	C	C	G	C	T	G	T	A	T
Rhesus	A	A	G	C	T	G	A	T	G	A	C	C	A	C	C	C	T	C	C	A	T	A	G	C	A	C	C	G	C	A	C	C	C	A	T	T	T	G	T	C	C	G	C	T	G	T	A	T
Orangutan	A	A	G	C	T	G	A	T	G	A	C	C	A	C	C	C	T	C	C	A	T	A	G	C	A	C	C	G	C	A	C	C	C	A	T	T	T	G	T	C	C	G	C	T	G	T	A	T
Gorilla	A	A	G	C	T	G	A	T	G	A	C	C	A	C	C	C	T	C	C	A	T	A	G	C	A	C	C	G	C	A	C	C	C	A	T	T	T	G	T	C	C	G	C	T	G	T	A	T
Bonobo	A	A	G	C	T	G	A	T	G	A	C	C	A	C	C	C	T	C	C	A	T	A	G	C	A	C	C	G	C	A	C	C	C	A	T	T	T	G	T	C	C	G	C	T	G	T	A	T
Chimpanzee	A	A	G	C	T	G	A	T	G	A	C	C	A	C	C	C	T	C	C	A	T	A	G	C	A	C	C	G	C	A	C	C	C	A	T	T	T	G	T	C	C	G	C	T	G	T	A	T
African Pygmy	A	A	G	C	T	G	A	T	G	A	C	C	A	C	C	C	T	C	C	A	T	A	G	C	A	C	C	G	C	A	C	C	C	A	T	T	T	G	T	C	C	G	C	T	G	T	A	T
Basque	A	A	G	C	T	G	A	T	G	A	C	C	A	C	C	C	T	C	C	A	T	A	G	C	A	C	C	G	C	A	C	C	C	A	T	T	T	G	T	C	C	G	C	T	G	T	A	T
Icelandic	A	A	G	C	T	G	A	T	G	A	C	C	A	C	C	C	T	C	C	A	T	A	G	C	A	C	C	G	C	A	C	C	C	A	T	T	T	G	T	C	C	G	C	T	G	T	A	T
Japanese	A	A	G	C	T	G	A	T	G	A	C	C	A	C	C	C	T	C	C	A	T	A	G	C	A	C	C	G	C	A	C	C	C	A	T	T	T	G	T	C	C	G	C	T	G	T	A	T
Russian	A	A	G	C	T	G	A	T	G	A	C	C	A	C	C	C	T	C	C	A	T	A	G	C	A	C	C	G	C	A	C	C	C	A	T	T	T	G	T	C	C	G	C	T	G	T	A	T
Amerindian	A	A	G	C	T	G	A	T	G	A	C	C	A	C	C	C	T	C	C	A	T	A	G	C	A	C	C	G	C	A	C	C	C	A	T	T	T	G	T	C	C	G	C	T	G	T	A	T
consensus	A	A	G	C	T	G	A	T	G	A	C	C	A	C	C	C	T	C	C	A	T	A	G	C	A	C	C	G	C	A	C	C	C	A	T	T	T	G	T	C	C	G	C	T	G	T	A	T

	10	20	30	40	50	60	70	80
CHIMPANZEE	GAGCAGCTGAACRAAGCTGATGACCACCCCTCCATAGCACCCGCACCCCATTTTGTCCGCTGTATTATATCCCAATGAGTTTAAGCAATCGG							
	E Q L N K L M T T L H S T A P H F V R C I I P N E F K Q S							
HUMAN	GAGCAGCTGAACRAAGCTGATGACCACCCCTCCATAGC--CGCACCCCATTTTGTCCGCTGTATTATATCCCAATGAGTTTAAGCAATCGG							
	E Q L N K L M T T L H S R T P F C P L Y Y P Q * V * A I							

(5) Myosin gel electrophoresis for LC/MS-MS.

Coomasie 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 fast-IIa/IIx comigrating at top and slow-I at bottom in lane 8. Labeling: 1, m.w. standard; 2, tibialis anterior, 3, temporalis (asterisk denotes band cut for proteomic analysis below); 4, masseter; 5, tensor veli palatini; 6, soleus; 7, temporalis; 8, *H. sapiens* temporalis; 9, *H. sapiens* soleus. Note that the topmost band in lane 8 is a doublet, corresponding to the products of the *MYH 1* and 2 genes¹.



(6) Peptide analysis by LC/MS-MS.

The *M. fascicularis* temporalis myosin heavy chain band of unique electrophoretic mobility was extracted and processed for LC/MS-MS analysis in the Wistar Institute proteomics core as described:

http://www.wistar.upenn.edu/research_facilities/facilities/protein/service.html#a

Sequences highlighted in red correspond to positions of 32 peptides identified by LC-MS/MS analysis of a tryptic digest of the predominant myosin in the *M. fascicularis* temporalis muscle (major band in lane e-3). All peptides underlined correspond to confirmed high-probability spectral matches obtained in a SEQUEST search of a current non-redundant protein sequence database updated to include the predicted MYH 14, 15 and 16 gene products. Cross comparison of these to the MYH gene products aligned in the attached file "MYH 1-8 Peptide Alignment" demonstrates the uniqueness of the majority of these peptide sequence motifs.

MPGGYKGECG DDVDMPFFLA PPEKERIEAM NKPYDIKRSC WVKDEKEGFV AGEIQSEQGD QVTVKTITNQ TLTVKKDDIQ
QMNPPKFYQA SDMADMTFLN EASVLDNLRQ RYTNMRYTY SGLFCVTVNP YKWLPIYGAR VANMYKGKKR TEMPPHLFSI
SDNAYHDMLM DRENQSMLIT GESGAGKTEN TKKVIQYFAN IGGTGKQTTD KKGSLEDQVI QANPVLEAFG NAKTTRNNNS
SRFGKFIIRH FGTGKLAGA DIESYLLEKS RVISQQAER SYHIFYQILS NKKPELVESL LLVPNPKEYH WVSQGVTTVD
NMDDKEELQI TDEAFDVLGF SAEKMAVYK LTGGIMHFGN MKFKQKPRDE QAEVDTTEVA DKVAHLMGLN SGELQKGITR
PRVKVGNFV QKGQNMEOCQ NSIGALGKAV YDKMFKWLVA RINKTLDTKM QRQFFIGVLD IAGFEIFEFN SFEQLCINF
NEKLQOFFNH HMFVLEQEEY KREGIEWVFI DFGLDLQACI DLLEKPMGIF SILEEQCVFP KATDATFKAA LYDNHLGKSS
NFLKPKGGKS KGPEVHFELV HYAGTVGYNI TGWLEKNKDP LNETVVGLFQ KSSVAILALL FKEEEAPAGS KKQKRGSSFM
TVSNFYREQL NKLMTTLHST APHFVRCIIP NEFKQSGVID AHLIMHQLAC NGVLEGIRIC RKGFPNRLQY PEFKQRYQVL
NPNVIPQGFV DNKKASELLL AAILDLVNEY KIGHTKVFFR AGILARLEDM RDERLAKIMT MLQCRLRGFL MRVEFKKMLE
RRMGLKVIQQ NVHKFLQLRF WGWWKLYNKV KPLLNVARQE EEMKAKEEEL RKAMAQTQEL VNKVKELEEK TATLSQEKND
LTIQLQAEQE NLMDAERLT WMMKTKMDLE SQISDMRERL EEEEGMAASL SAAKRKLEGE LSDLKRDLG LETTLAKTEK
EKQALDHKVR TLTGDSLRE DSITKLQKEK RALEELHQKT LDDLQAEEDK VNHLTKNNSK LSTQIHELED NWEQEKKIRA
EVEKARRKAE SDLKMTIDNL NEMERSKLDL EEVVKRDLE INSVNSKYED EQSLNSTLQR KLKEHQDRIE ELEEELEAER
AMRAKVEKQR SDLSRDLEDL SDRLEEAGGA TSAQIEQNRK REAELLKLR ELEEAALQSE ATASTLRKKH VDSMAELTEH
VESLQRVSK LEKDKQVMKA EIDDLNASME TIQKSKMNAE AHVRKLEDSL SEANAKVAEL ERNQAEINAI RTRLQAENSE
LSREYEESSQ RLNQILRIKT SLTSQVDDYK RQLDEESKSR STAVVSLANT KHDLDLVKEQ LEEEQGGKSE LQRLVSKLNT
EVTTWTKYE TDAIQRTEEL EETKRKLAAR LQEAEEAAET AQARAASLEK NKQRLQAEVE DLTIDLEKAN AAAAALDKKQ
RLFDKMLAEW QQKCEELQVE VDSSQKECRM YMTESFKIKT AYEESLEHLE SVKKNKTLO EEIKDLIDQL GEGGRSVHEL
QKLKKLEME KEELQVALEE AESSLEVEES KVIRIQLELA QVKADIDRRI HEKEEFEAT RKNHQRAIES LQASLEAEAK
GRAEALRLKK KMETDLNEME IQLDHANKNN SELVKTLLKRL QQQIKDLQVQ MDEDARQHEE LRKQYNLQER RLSLLQTELE
EVRSALEGSE RSRKLEQEV VEITFWHNEI NIQNQSLLV KRKLESDVQR ISNEHEELIS EFRLTEERAK KAMMDAARMA
EELRQEQDHC MHLEKIKKNY EVTIKDLQAK MEEAEQLALK GKKRTIMKLE ARIKELETEL DGEQKQHVT VKTLCCKNERR
LKELVFQTEE DHKTNRMQA LVEKLQNKLK VYKRQIEEAE DQANQTLARY RKTVHELDDA EDRAGMAETA LNKLRTHRV
AGKGITSVEI IQVSKTGTSK TLSEE

(7) Reconstruction of Ancestral Sequences

The limited availability of fresh masticatory muscle samples from endangered primate species required us to sequence coding regions from genomic DNA. We selected a sampling of the largest exons deduced from our human MYH 16 pseudogene reconstruction and designed PCR primers from the flanking intron sequences. Sequence for *Canis familiaris* and *Macaca fascicularis* were obtained by RTPCR of mRNA prepared from biopsies of temporalis muscle. These sequences are available under the following Genbank Accession numbers (corresponding sequence from *Homo sapiens* can be found in our Third Party Annotation of the MYH16 locus, as described above – TPA# BK001410):

<u>Species/Exon</u>	<u>23</u>	<u>25</u>	<u>30</u>	<u>31</u>	<u>34b</u>	<u>37</u>
Chimpanzee (<i>Pan troglodytes</i>)	AY350724	AY350726	AY485943	AY485947	AY350730	AY350728
Orangutan (<i>Pongo pygmaeus</i>)	AY485954	AY485955	AY485956	AY485957	AY485948	AY485958
Cynomolgus Macaque (<i>Macaca fascicularis</i>)	AY350723	AY350725	AY485942	AY485945	AY350729	AY350727
Dog (<i>Canis familiaris</i>)	AY485951	AY485952	AY485944	AY485946	AY485950	AY485953
Gorilla (<i>Gorilla gorilla</i>)					AY485949	

These sequences were compiled into a contiguous open reading frame to facilitate codon-by-codon analysis. An unambiguous alignment of these sequences (no insertions or deletions were required) was obtained with the program ClustalW, as implemented in the MacVector software package (Accelrys, Symantec Corporation). This sequence alignment was exported into the appropriate file format for ancestral sequence reconstruction by two algorithms: maximum parsimony and maximum likelihood. We used the program PAMP for the former and BASEML for the latter, both as contained within the most recently posted release of the PAML software package, [paml3.13d.tar.gz](ftp://abacus.gene.ucl.ac.uk/pub/paml/), from the ftp download website: <ftp://abacus.gene.ucl.ac.uk/pub/paml/>

The reconstructed ancestral sequences were identical at all nucleotide positions with the sole exception of an ambiguity which was resolved by sequencing exon 34b of the MYH 16 ortholog from gorilla and rerunning the analysis for this exon (AY485949). Ancestral sequences are available from the author upon request as either merged files or as separate files for the individual exons.

The combined ancestral and extant species sequence file alignment was converted into a file format native to the MEGA2 software package as downloaded from <http://www.megasoftware.net/> and used to calculate synonymous and non-synonymous mutations for all pairwise sequence comparisons. In addition, statistics on the probability of negative (purifying) darwinian selection were calculated using the Z-test function as implemented in the MEGA2 package². Relevant data are summarized in Figure 4, a complete spreadsheet is available in the MSEXcel file: MYH 16 Distances.xls.

(8) Statistical Significance of Differences in dN/dS Ratios Between Lineages

To supplement the *ad hoc* calculations on statistical significance in the MYH 16 Distances.xls file, we used the method of Yang³ and the software described therein (specifically the codeml program of the PAML package) to examine the statistical significance of the lineage-specific differences in dN/dS ratios (ω) seen in our sampling of the *MYH 16* coding sequence. This statistical test allows one to assess whether a two-ratio model fits the data better than a one-ratio model by comparing the likelihood (l) given by maximum likelihood (ML) analysis for each branch when ω is constrained to be the same as all other branches (the null hypothesis), with l given when ω for that branch was free to settle upon its own value. Only in the case of the human lineage does the difference between these ω values achieve a statistical significance of approximately $p = 0.05$, with a lineage-specific ω of approximately 0.5. Other lineage-specific values of ω average approximately 0.07, with a maximum of 0.12, implying purifying darwinian selection. In all cases an overall tree topology similar to that in Figure 4 was assumed.

Table 1 dN/dS ratios and log likelihood values under different models

Model	dN/dS Ratios Allowed	Free Parameter	p	l	ω_0	ω_{free}
A	1	none	12	1897.49	0.0522	—
B	2	ω_D	13	1896.50	0.0746	0.0376
C	2	ω_M	13	1897.38	0.0499	0.0718
D	2	ω_P	13	1896.85	0.0471	0.1168
E	2	ω_C	13	1897.36	0.0526	0.0001
F	2	ω_H	13	1895.62	0.0478	0.5120

Table 1 dN/dS ratios and log likelihood differences under different models

Model A constrains the dN/dS ratio (ω) to be the same for all branches. Models B-F allow the dN/dS ratio for a single branch (species) to vary from the others. p , number of parameters; l , log likelihood value; ω_0 represents the background dN/dS ratio among all branches not free to vary; ω_D , dog; ω_M , macaque; ω_P , pongo (orang-utan); ω_C , chimpanzee; ω_H , human.

Table 2 Significance of log likelihood value difference between models

Null Hypothesis Tested	Assumption Made	Models Compared	2Δl	p-value
$\omega_D = \omega_0$	$\omega_M = \omega_P = \omega_C = \omega_H = \omega_0$	A and B	1.986796	0.1587
$\omega_M = \omega_0$	$\omega_D = \omega_P = \omega_C = \omega_H = \omega_0$	A and C	0.229252	0.6320
$\omega_P = \omega_0$	$\omega_D = \omega_M = \omega_C = \omega_H = \omega_0$	A and D	1.277616	0.2583
$\omega_C = \omega_0$	$\omega_D = \omega_M = \omega_P = \omega_H = \omega_0$	A and E	0.256468	0.6125
$\omega_H = \omega_0$	$\omega_D = \omega_M = \omega_P = \omega_C = \omega_0$	A and F	3.753482	0.0527*

Table 2 Significance of log likelihood value differences between models

For each model described in Table 1, we test the null hypothesis that the dN/dS ratio along each branch is the same whether it is free to vary (Models B-F) or is constrained to have the same dN/dS ratio as all of the other branches (Model A).

(9) Estimation of the age of the MYH 16 gene inactivation

See Chou et al., 2002, Proc Natl Acad Sci USA 11736-41, for a detailed description of the rationale for this analysis. Briefly, the assumption is made that non-synonymous mutations are selected against until the gene is inactivated, thereafter mutations at both synonymous and non-synonymous sites accumulate at the neutral mutation rate. Quantification of lineage-specific mutation rates at synonymous and non-synonymous sites remote from the inactivating deletion provides the information necessary for the calculation. We define the following terms as described in Chou, et al., 2002:

t = time since last common ancestor (LCA) human/chimpanzee

t_1 = time since MYH 16 deletional inactivation

k = neutral mutation rate = (number synonymous mutations/number of synonymous sites)/ t
= $dS/t = 1/2$ neutral site divergence human-chimpanzee/ t

f_N = dN/dS (average for all species) = .046 from data in MYH 16 Distances.xls and Figure 4.

$f_N k (t-t_1) + kt_1 = dN(\text{LCA to human lineage})$ LCA is represented by Node C in Figure 4.

$f_N k t + kt_1(1-f_N) = dN(\text{LCA to human lineage})$

$t_1(1-f_N) = (dN(\text{LCA to human lineage})/k) - f_N t$

$t_1 = ((dN(\text{LCA to human lineage})/k) - f_N t) / (1-f_N)$

= $((dN(\text{LCA to human})/(dS(\text{LCA to Human})) - f_N) / (1-f_N)) t$

= $((dN(\text{LCA to human})/(1/2 \text{ human-chimp neutral site divergence}) - f_N) / (1-f_N)) t$

The calculation requires normalization to the fossil record, for which we focus on the human-chimpanzee divergence to minimize the effects of lineage-specific slowing of the molecular clock⁴. Recent fossil data suggests a divergence as remote as 6 to 7 mya⁵. The level of codon bias in the *MYH* genes is high (data not shown), perhaps reflecting their genomic contexts or their high levels of expression⁶. This may diminish the rate of the silent molecular clock as determined from *MYH* synonymous positions. Thus we substitute the silent site mutation rate obtained from a genome-wide sampling of non-coding, non-repetitive human and chimpanzee DNA sequences for the minimal sampling of non-synonymous sites represented by our sequencing of *MYH 16* orthologs⁴. This global number has been defined with precision as the number of substitutions per 100 sites between the two species: 1.19 +/- 0.016 (Yi). Substituting the foregoing numbers into the last equation above we obtain:

$t_1 = ((2/840) / ((1/2) * .0119 +/- 0.00016) - .04643) / (1-.04643)) 6.5 +/- 1 \text{ mya} = (.37 +/- .01) (6.5 +/- 1)$

= 2.4 +/- .25 mya

We recognize that with the current data sampling this estimate is based on a small number of non-synonymous substitutions between human and chimpanzee, as is necessarily the case. Although this calculation yields a broad temporal range, we note that hominid fossils exhibiting decreased masticatory robusticity fall into the recent end of this window.

1. Korfage, J. A., Brugman, P. & Van Eijden, T. M. Intermuscular and intramuscular differences in myosin heavy chain composition of the human masticatory muscles. *J Neurol Sci* **178**, 95-106 (2000).
2. Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244-5 (2001).
3. Yang, Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol* **15**, 568-73 (1998).
4. Yi, S., Ellsworth, D. L. & Li, W. H. Slow molecular clocks in Old World monkeys, apes, and humans. *Mol Biol Evol* **19**, 2191-8 (2002).
5. Brunet, M. et al. A new hominid from the Upper Miocene of Chad, Central Africa. *Nature* **418**, 145-51 (2002).
6. Sharp, P. M., Averof, M., Lloyd, A. T., Matassi, G. & Peden, J. F. DNA sequence evolution: the sounds of silence. *Philos Trans R Soc Lond B Biol Sci* **349**, 241-7 (1995).