

Myostatin rapid sequence evolution in ruminants predates domestication

Åsa Tellgren^{a,b}, Ann-Charlotte Berglund^{a,b}, Peter Savolainen^c,
Christine M. Janis^d, David A. Liberles^{a,b,*}

^a Computational Biology Unit, Bergen Centre for Computational Science, University of Bergen, 5020 Bergen, Norway

^b Stockholm Bioinformatics Center and Department of Biochemistry and Biophysics, Stockholm University, 10691 Stockholm, Sweden

^c Department of Biotechnology, Royal Institute of Technology, 10691 Stockholm, Sweden

^d Department of Ecology and Evolutionary Biology, Brown University, Providence, RI 02912, USA

Received 5 March 2004; revised 19 May 2004

Available online 11 September 2004

Abstract

Myostatin (GDF-8) is a negative regulator of skeletal muscle development. This gene has previously been implicated in the double muscling phenotype in mice and cattle. A systematic analysis of myostatin sequence evolution in ruminants was performed in a phylogenetic context. The myostatin coding sequence was determined from duiker (*Sylvicapra grimmia caffra*), eland (*Taurotragus derbianus*), gaur (*Bos gaurus*), ibex (*Capra ibex*), impala (*Aepyceros melampus rednilis*), pronghorn (*Antilocapra americana*), and tahr (*Hemitragus jemlahicus*). Analysis of nonsynonymous to synonymous nucleotide substitution rate ratios (K_a/K_s) indicates that positive selection may have been operating on this gene during the time of divergence of Bovinae and Antilopinae, starting from approximately 23 million years ago, a period that appears to account for most of the sequence difference between myostatin in these groups. These periods of positive selective pressure on myostatin may correlate with changes in skeletal muscle mass during the same period. © 2004 Elsevier Inc. All rights reserved.

Keywords: Myostatin; Positive selection; Phylogeny; Artiodactyl; Skeletal muscle mass

1. Introduction

Myostatin (GDF-8) is a member of the transforming growth factor- β (TGF- β) superfamily that acts as a negative regulator of muscle cell development (Lee and McPherron, 1999). The TGF- β superfamily contains a large number of homologous cytokine proteins with similar folds that are involved in regulating various cellular processes (see Shah et al., 2001 for a more detailed description). The myostatin gene is expressed in skeletal muscle, producing a precursor protein with an N-terminal secretory signal, a propeptide domain, and a C-terminal domain. Myostatin is proteolytically processed

at a conserved four amino acid RSRR site (cleaved by bone morphogenetic protein 1 or BMP-1) to produce a 13 kDa mature peptide (Wolfman et al., 2003). Small glutamine-rich tetratricopeptide repeat-containing protein (SGT) interacts with uncleaved myostatin and may regulate its secretion (Wang et al., 2003). Myostatin is active as a homodimer, but forms a noncovalent biologically inactive complex with its propeptide (Hill et al., 2002; Lee and McPherron, 2001). Other proteins in serum also interact with myostatin to negatively regulate it, including follistatin, follistatin related gene (FLRG), growth arrest specific gene 1 (GAS-1), and telethonin (Hill et al., 2002; Lee and McPherron, 2001; Nicholas et al., 2002).

Like other members of the TGF- β superfamily, myostatin signals through heteromeric receptor complexes.

* Corresponding author. Fax: +47 55584295.

E-mail address: liberles@cbu.uib.no (D.A. Liberles).

Myostatin binds activin type II receptor B (ActRIIB), and to a lesser extent ActIIA (Lee and McPherron, 2001). When bound to type II receptors, they recruit type I receptors like activin receptor-like kinase 4 (ALK4) or ALK5, inducing phosphorylation of Smad2 and Smad3. Phosphorylated Smad2 and Smad3 move to the nucleus, and bind to DNA binding proteins, ultimately downregulating myogenesis (Rebbapragada et al., 2003). Relatedly, myostatin antagonizes the effects of BMP-7 in adipose cells, also downregulating adipogenesis (Rebbapragada et al., 2003).

Mutations affecting myostatin function in cattle breeds like Belgian Blue and Piedmontese (McPherron and Lee, 1997), as well as in mice (McPherron et al., 1997) are known to result in a double muscling phenotype, where the number of specific types of muscle fibers is approximately doubled. Examination of the allele distribution and linkage analysis of cattle is hypothesized to show the signatures of selection (Wiener et al., 2003). Similarly, a systematic analysis of the ratio of nonsynonymous to synonymous nucleotide substitution rates (K_a/K_s) has shown significant K_a/K_s ratios (K_a/K_s greater than one is an indicator of genes potentially under positive selective pressures to alter functions along lineages of phylogenetic trees) in the lineages leading from the last common ancestor of sheep and cow to both sheep and cow. This was not observed in the lineage leading from the last common ancestor of sheep, cow, and pig to either pig or the cow–sheep last common ancestor (Bovidae) (in the database described in Liberles et al., 2001). Interestingly, this study also showed a burst of positive selection after a fish-specific gene duplication event that may also be related to musculature.

Based upon the Artiodactyl data, we tested whether the rapid sequence evolution of myostatin between sheep (*Ovis aries*) and cow (*Bos taurus*) was due to recent selection during domestication or if this was due to a more ancient event closer to the separation of sheep and cow around 23 million years ago (this date is based upon molecular inferences in Hassanin and Douzery, 1999). This was tested initially by sequencing myostatin from ibex and gaur, two intermediate relatives of sheep and cow that have never been domesticated. Ibex (*Capra ibex*) is a wild relative of goats, that is thought to have diverged from sheep 2.8–7.0 million years ago (Hassanin and Douzery, 1999). Gaur (*Bos gaurus*) is a close Indian relative of the cow that is thought to have diverged within the past 3.3 million years (Hassanin and Douzery, 1999).

Additionally, myostatin was also sequenced from four other species internal to Bovidae [eland (*Taurotragus derbianus*), duiker (*Sylvicapra grimmia caffra*), impala (*Aepyceros melampus rednilis*), and tahr (*Hemitragus jemlahicus*)] plus an immediate outgroup, pronghorn (*Antilocapra americana*). Goat (*Capra hircus*), a close relative of ibex, has been sequenced independently by

another group (Tay et al., 2004). Eland is a bovine that is more distant to cow than gaur (Hassanin and Douzery, 1999; Matthee and Davis, 2001). Duiker, and impala are Antilopinae that are more distant to sheep than goat and ibex, while tahr falls within the Caprini group including sheep, goat, and ibex (Groves and Shields, 1996; Hassanin and Douzery, 1999). Together these sequences allowed us to more precisely pinpoint the period where myostatin was under positive selective pressure.

2. Materials and methods

2.1. DNA samples

DNA samples from Gaur (*Bos gaurus*), Giant Eland (*Taurotragus derbianus*), Ibex (*Capra ibex*), Himalayan Tahr (*Hemitragus jemlahicus*), Kenyan Impala (*Aepyceros melampus rendilis*), Kaffir Crowned Duiker (*Sylvicapra grimmia caffra*), and Pronghorn (*Antilocapra americana*) were obtained from the Center for the Reproduction of Endangered Species (CRES) at the Zoological Society of San Diego. The sequences used from GenBank were from pig (*Sus scrofa*) (AF019623), cow (*Bos taurus*) (AF320998), sheep (*Ovis aries*) (AF019622), and goat (*Capra hircus*) (AY436347).

2.2. Polymerase chain reaction

Exon sequences were amplified from genomic DNA. GenBank sequences from *Bos taurus*, *Ovis aries*, and *Sus scrofa* were used to design primers. Forward primers for exon 1 were M02-1f, M02-2f, M03-14f, M03-15f, M04-34f, and M04-35f. Reverse primers for exon 1 were M03-16r and M03-25r. Primers in and around exon 2 were forward primers M03-20f, M03-26f, and M03-27f and reverse primers M02-4r, M03-17r, M03-28r, M03-29r, M04-36r, and M04-37r. Exon 3 was amplified with forward primers M03-22f, M03-30f, and M03-31f together with reverse primers M03-23r, M03-24r, M03-32r, and M03-33r. Primer locations and sequences can be visualized in Fig. 1 (Table 1). In a typical reaction, 10 ng template DNA was mixed with 20 mM Tris–HCl, 50 mM KCl, 0.2 mM of each dNTP, 1.5 or 2.5 mM MgCl₂, 2 μM of each primer and 1 U of platinum *Taq* DNA polymerase (Invitrogen, Applied Biosystems, Foster City, California), in a total volume of 50 μl. The amplification program consisted of predenaturation (94 °C for 2 min) followed by 35 cycles of denaturation (94 °C for 0.5 min), primer annealing (55 °C for 0.5 min) and extension (72 °C for 2.5 min), and a final extension (72 °C for 7 min).

In some cases nested PCRs were used, where 20 cycles were used in the first PCR followed by a second PCR with 20–35 cycles. Typically for exon 1, primers M03-14f and M03-17r (M02-4r for Eland) were used in a first

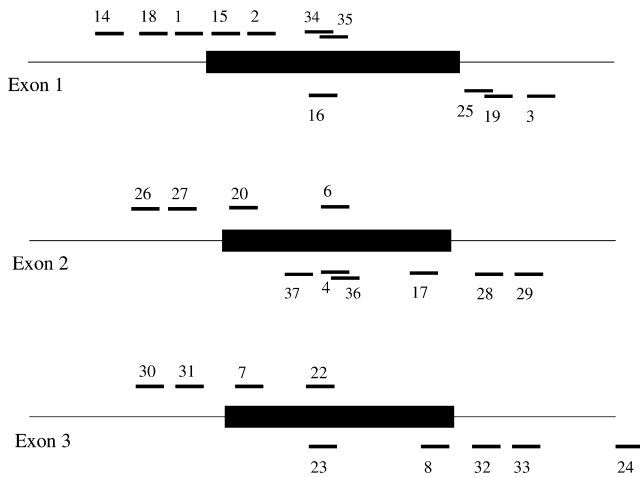


Fig. 1. The location on the gene of the primers described in Table 1 is shown. The use of each primer is described in Materials and methods.

Table 1

A list of primers used in PCR and sequencing reaction is shown

M02-1f	5'-AGAAGTAAGAACAAGGGAAAAG-3'
M02-2f	5'-ATCTGAATGAGAACAGCGAGC-3'
M02-3r	5'-AAACATAAACATGTTATAGTTGTC-3'
M02-4r	5'-TGCCTGGGTTTCATGTCAAGT-3'
M02-6f	5'-AACTTGACATGAACCCAGGCA-3'
M02-7f	5'-TAGAAGTCAAGGTAACAGACAC-3'
M02-8r	5'-CACAGCGATCTACTACCATG-3'
M03-14f	5'-GCTCACCCCTTGACTGTAACA-3'
M03-15f	5'-CATGCAAAAAGTCAAAATCTTTG-3'
M03-16r	5'-TAGGAGCTGTTTCCAGGCG-3'
M03-17r	5'-GAAGGTTACAGCAAGATCATG-3'
M03-18f	5'-AGATTCAGTGGTGTGGCAAGT-3'
M03-19r	5'-TACAAGCCAGCAGCTTGTTAA-3'
M03-20f	5'-AAGACGGTACAAGGTATACTG-3'
M03-22f	5'-TCCTCATACCCATCTTGTGC-3'
M03-23r	5'-GCACAAGATGGGTATGAGGA-3'
M03-24r	5'-CACATTCACATTATACAGCCAT-3'
M03-25r	5'-GCAGCTTGTTAAAAGAGCCTG-3'
M03-26f	5'-GCTGTTATGAATGAAATGCTAC-3'
M03-27f	5'-AGACTAAATATGCACACATTATTC-3'
M03-28r	5'-GTCTGGCTTATAAGCACAGG-3'
M03-29r	5'-TTTTATTGGGTACAGGGCTAC-3'
M03-30f	5'-ATGTGACATAAGCAAAATGATTAG-3'
M03-31f	5'-ATAATGAGTCTTGAGGTAGG-3'
M03-32r	5'-GTGGTACATAATTTACAGTTTC-3'
M03-33r	5'-CTTGTGCTTAAGTGAAGTGTAG-3'
M04-34f	5'-GCCTGGAAACAGCTCCTAAC-3'
M04-35	5'-CTCCTAACATTAGCAAAGATGC-3'
M04-36	5'-GCCTGGGTTTCATGTCAAGC-3'
M04-37	5'-GCCTGGGTTTCATGTCAAGC-3'

The location of these primers on the gene is shown in Fig. 1.

PCR followed by an additional PCR with primers M02-1f and M02-4r, M02-2f, and M04-37r (M04-36r for Pronghorn), as well as M03-14f and M03-25r. First primers used for exon 2 were M03-26f and M03-29r, followed by primers M03-27f and M03-28r. Exon 3 was amplified with M03-30f and M03-33r, followed by M03-31f and M03-32r.

2.3. Sequencing

Complete exon sequences were obtained from two repeated PCRs sequenced in forward and reverse. Primers used for PCR were also used for sequencing. In addition, primers M02-3r, M03-18f, and M03-19r were used for sequencing exon 1, M02-6f for exon 2, and M02-7f and M02-8r for exon 3.

In the reactions, 2–4 μ l of PCR was mixed with 3.2 pmol of sequencing primer and BigDye2 or BigDye3 cycle sequencing were performed according to the manufacturer's recommendations (PE Biosystems). The cycle sequencing steps used were 95°C for 30s, followed by 30 cycles of 96°C for 10s, 55°C for 5s and 64°C for 4min. The samples were analyzed on a ABI PRISM 377 (PE Biosystems). Sequences have been deposited in GenBank, with accession numbers as follows: duiker (bankit623239), eland (bankit629604), gaur (bankit629588), ibex (bankit629634), impala (bankit629648), pronghorn (bankit623243), and tahr (bankit629638).

2.4. Phylogeny and sequence analysis

The protein and DNA sequences aligned without gaps. This was confirmed using Darwin, which was also used for writing all sequence analysis and manipulation programs (Gonnet et al., 2000). The seqboot, DNAML, and consense programs from Phylip with 100 bootstrapped replicates (Felsenstein, 1989) together with MRBAYES (Huelsenbeck and Ronquist, 2001) were used for phylogeny construction from two datasets, third codon positions only (to minimize the effects of selection) and the entire coding sequences. In MRBAYES, a GTR model with gamma distributed rates across sites was used with 750,000 generations and a burnin of 100 trees.

Ratios of nonsynonymous to synonymous nucleotide substitution rates were calculated using a parsimony based method (Liberles, 2001) and maximum likelihood based methods (Yang, 1998; Yang and Nielsen, 2002) encoded in PAML (Yang, 1997). The parsimony-based method was run using a standard approach (method 2 from Liberles, 2001). The covarion-based partitioning method (Siltberg and Liberles, 2002) on the parsimony-based K_a/K_s ratios was run using the most conservative mode, with invariants defined from the top node of the tree including pig as the most distant sequence from the ruminants. All parsimony-based K_a/K_s approaches are encoded in Darwin (Gonnet et al., 2000) and available online at <http://www.bioinfo.no>.

Three sets of likelihood ratio tests were performed using PAML (Yang, 1997). The objective of the first set was to test the hypothesis that there is one K_a/K_s parameter for each branch versus the null hypothesis that all branches have the same K_a/K_s -parameter. These hypotheses were modeled using the models in Yang

(1998) (the one-ratio and the free-ratio model). The results of this comparison are shown in Table 2.

The second set of likelihood ratio tests specifically tests for comparison with relaxation of selective con-

Table 2

A likelihood ratio test based upon the method of Yang (1998) is described

Model	Log likelihood	<i>P</i>	Significant conclusions
One K_a/K_s ratio	-2379	—	$K_a/K_s = 0.20$
Free K_a/K_s over branches	-2358	0.001	K_a/K_s (Bovinae) = 1.32 K_a/K_s (Antilopinae) = 1.32 K_a/K_s (Cephalo-caprini) = 1.90

All sites were treated equivalently along a branch, but the range of available K_a/K_s ratios for each branch was varied.

Table 3

A likelihood ratio test is implemented to test for relaxation of selective constraint on the whole sequence

Model	Log likelihood	<i>P</i>	Significant conclusions
Two K_a/K_s ratios	-2367	—	K_a/K_s constrained to 1.00 over Bovinae, Antilopinae, Cephalo-caprini
Two K_a/K_s ratios unconstrained	-2367	0.35	K_a/K_s free but equal over Bovinae, Antilopinae, Cephalo-caprini K_a/K_s (Bovinae) = 1.39 K_a/K_s (Antilopinae) = 1.39 K_a/K_s (Cephalo-caprini) = 1.39
Four K_a/K_s ratios	-2366	0.35	K_a/K_s free over Bovinae, Antilopinae, Cephalo-caprini K_a/K_s (Bovinae) = 1.30 K_a/K_s (Antilopinae) = 1.27 K_a/K_s (Cephalo-caprini) = 1.96

A null model fixes K_a/K_s along the three positive selected branches from the model in Tables 2 to 1. Subsequent tests allow K_a/K_s to be positively selected together or independently along these lineages.

Table 4

A likelihood ratio test based upon the method of Yang and Nielsen (2002) was performed

Model	Log likelihood	<i>P</i>	Significant conclusions
Neutral site classes, $K_a/K_s = 0.1$	-2383	—	$p_0 = 0.72$
Positive selection, Bovinae branch, $K_a/K_s = 0.1$, est.	-2380	0.05	$p_0 = 0.51$, $p_1 = 0.17$ p_2 sites in propeptide and mature protein K_a/K_s (Bovinae) = 3.10
Positive selection, Antilopinae branch, $K_a/K_s = 0.1$, est.	-2377	0.005	$p_0 = 0.40$, $p_1 = 0.12$ p_2 sites in propeptide and mature protein K_a/K_s (Antilopinae) = 2.07
Positive selection, Cephalo-caprini branch	-2378	0.01	$p_0 = 0$, $p_1 = 0$ (not biologically realistic) K_a/K_s (Cephalo-caprini) = 1.49
Positive selection, Bovine, Antilopinae and Cephalo-caprini branches	-2367	0.0005	$p_0 = 0.31$, $p_1 = 0.06$ p_2 sites in propeptide and mature protein $K_a/K_s = 2.10$

Sites were divided into two categories, negatively selected and neutral, in the first model. The three subsequent modes allow a fraction of sites to be positively selected along an individual lineage and are compared to the first model for significance.

straint along specific branches versus the possibility of positive selection. The model under the null hypothesis constrains the K_a/K_s ratios on specific branches to 1, which is equivalent to assuming that the branches are evolving neutrally. Two different models are used for the alternative hypothesis that the three branches of interest are evolving under positive selection. The first model estimates a single parameter of equal ratios along the specific branches. The second model allows each of the three branches to have individual K_a/K_s parameters. The results of this test are shown in Table 3.

In the first two sets of tests, the models describing the hypotheses do not allow for different selective pressure along the gene sequences. The third set of likelihood ratio tests is as the second set a test of positive selection against neutral evolution. However, the models used for describing the hypotheses are based upon the site-branch models of Yang and Nielsen (2002). Here, the selective pressure is allowed to vary along the sequence. In this set, the null hypothesis allows sites to be negatively selected or neutral. Three additional hypotheses that allow sites to vary allow a particular branch with a third K_a/K_s category. A last hypothesis allows three branches together to have a separate K_a/K_s ratio and site category from the background. The results of these comparisons are shown in Table 4.

Given that the null hypothesis holds, the likelihood ratio test statistic can be compared to the χ^2 distribution with the appropriate degrees of freedom given from the null and alternative hypotheses. The properties of such likelihood ratio tests are described in more detail in Anisimova et al. (2001).

Finally, K_a was measured using the parsimony-based reconstruction approach in all 30 nucleotide blocks of contiguous sequence for myostatin excluding the signal peptide along the three lineages where K_a/K_s was significantly greater than one. K_a was calculated using the methodology and program of Liberles (2001).

3. Results

Myostatin gene sequences were obtained by PCR from duiker, eland, gaur, ibex, impala, pronghorn, and tahr in addition to available sequences from cow, goat, sheep, and pig. Two datasets, one based upon

the complete coding sequences and another based upon only third codon positions to minimize the effects of selection, were used to generate maximum likelihood phylogenetic trees for these sequences using two different methods (from DNAm1 from Phylip and MRBA-YES). The protein multiple sequence alignment, which

	1	15	16	▼	30	31	45	46	60	61	75	76	90
pig	MQKLQIYVYIYLFML	IVAGPVDLNENSEQK			ENVEKEGLCNACMWR		QNTKSSRLEAIKIQI		LSKLRLETAPNISKD		AIRQLLPKAPPLREL		
pronghornS.....G.....L.....
elandF.....K.....L.....E..T.....L..
gaurF.....L.....E..T.....L..
cowF.....L.....E..T.....L..
impalaF.....V.....Y.....L.....N.....
duikerF.....	T.....K.....L.....N.....
sheepF.....	L.....K.....L.....N.....
goatF.....	L.....K.....L.....N.....
tahrF.....	L.....K.....L.....N.....
ibexF.....	L.....K.....L.....SN.....
	91	105	106		120	121	135	136	150	151	165	166	180
pig	IDQYDVQRDDSSDGS	LEDDDYHATTETIIIT			MPTESDLLMQVEGKP		KCCFFKFSSKIQYNK		VVKAQLWIYLRPVKT		PTTVFVQILRLIKPM		
pronghornV.....T.....Q.....
elandF..G.A.....R..V.....T.....	L.....A.....
gaurF..A.....R..V.....T.....	L.....A.....
cowF..A.....R..V.....T.....	L.....A.....
impalaV.....AE..E.....
duikerV..V.....F.AE.QE.....H..I.....
sheepV..V.....AE.QE.....H.....
goatV.....AE.QE.....H..L.....I..A.....
tahrV..V.....AE.QE.....H.....
ibexV..V.....AE.QE.....H.....
	181	195	196		210	211	225	226	240	241	255	256	▼270
pig	KDGTRYTGIRSLKLD	MNPGTGIWQSIDVKT			VLQNWLKQPESNLGI		EIKALDENGHD LAVT		FPGPGEDGLNPFLEV		KVTDTPKRSRRDFGL		
pronghornA.....E.....A.....
elandR.....E.....T.....
gaurE.....T.....
cowE.....T.....
impalaN.....E.....E.....
duikerE.....E.....A.....
sheepE.....E.....
goatE.....E.....
tahrE.....E.....
ibexE.....E.....
	271	285	286		300	301	315	316	330	331	345	346	360
pig	DCDEHSTESRCCRYP	LTVDFEAFGWDWIIA			PKRYKANYCSGECEF		VFLQKYPHTHLVHQA		NPRGSAGPCCTPTKM		SPINMLYFNGKEQII		
pronghorn
elandEG..
gaurEG..
cowEG..
impala
duiker	L.....K.....
sheep	L.....K.....
goat	L.....K.....
tahr	L.....K.....
ibex	L.....K.....
	361	375											
pig	YGKIPAMVVDRCGCS												
pronghorn
eland
gaur
cow
impalaG.....
duikerG.....
sheepG.....
goatG.....
tahrG.....
ibexG.....

Fig. 2. A multiple sequence alignment of the translated myostatin protein sequences is shown. No gaps are evident and this was confirmed using the multiple sequence alignment algorithm encoded in Darwin (Gonnet et al., 2000). All positions that are different from the pig sequence are indicated. Arrows mark the divisions between the signal peptide (positions 1–23), propeptide (positions 24–266), and mature protein (positions 267–375). The corresponding DNA multiple sequence alignment can be found at http://www.ii.uib.no/~aasat/myo_dna_alignment.html.

lacked gaps, for these sequences is shown in Fig. 2 (the DNA sequence alignment can be viewed online at http://www.ii.uib.no/~aasat/myo_dna_alignment.html). All trees generated showed the same topology, as indicated in Fig. 3.

Two interesting groupings are indicated in this tree. Tahr and ibex are seen to be most closely related, with goat as an immediate outgroup, although with weak statistical support. Also, impala is observed as an outgroup to duiker together with Caprini. Few genes have been sequenced previously from these species and myostatin provides another datapoint in resolving such taxonomic issues in ruminants. Overall, our tree is consistent with existing knowledge on the relationships of these species (see for example, Hassanin and Douzery, 2003).

Nomenclature over the phylogenetic groupings where myostatin has been sequenced is defined for the purpose of clarity. The group of tahr, ibex, goat, and sheep is

referred to as Caprini. With the addition of duiker, it is referred to as Cephalo-caprini. With the addition of impala, it is known as Antilopinae. Cow, gaur, and eland are collectively known as Bovinae. Bovinae and Antilopinae are known as Bovidae. Adding pronghorn, the group is known as Ruminantia.

With a reliable tree (and nomenclature) in hand, sequence evolution was analyzed in the myostatin gene family. The K_a/K_s ratio is a valuable tool for detecting selective pressures between closely related species. A higher rate of K_a than K_s is indicative of a selective regime where individuals with mutations showed improved fitness over individuals without mutations. Using a parsimony reconstruction of ancestral states coupled to a standard method for counting nonsynonymous and synonymous substitutions and sites (Liberles, 2001), the lineage leading to Bovinae was observed to have a significant K_a/K_s ratio of 1.17 ± 0.01 , that leading to Cephalo-caprini a significant ratio of 1.42 ± 0.01 and that leading to Antilopinae an elevated ratio of 1.00 ± 0.01 (see Fig. 3). This last ratio becomes significantly >1 when excluding invariant sites in the covarion-based partitioning method (Siltberg and Liberles, 2002). The other branches are in the range 0–0.85, with only one other branch above 0.6 and only three other branches above 0.3. The branch with a value of 0.85 is that leading to goat.

Further, parsimony based methods may outperform likelihood based methods with very closely related sequences, but are known to severely underestimate the amount of substitution in more distantly related comparisons (Nielsen, 2002). A likelihood based analysis where K_a/K_s ratios were allowed to vary independently between branches (Yang, 1998) was also performed. This was compared with a model where K_a/K_s was estimated with one parameter over the entire tree. The results are described in Table 2. Overall, this result supports the use of multiple K_a/K_s parameters for each branch. This maximum likelihood analysis therefore also showed three positively selected branches for further testing, those leading to Bovinae ($K_a/K_s=1.32$), Antilopinae ($K_a/K_s=1.32$) and to Cephalo-caprini ($K_a/K_s=1.90$).

To test specifically against the possibility of relaxation of selective constraint, a null model was established with K_a/K_s along the three positively selected branches from Table 2 fixed to 1. Alternative hypotheses where K_a/K_s showed positive selection were tested and the results are shown in Table 3. These alternative models were preferred, but the results were not statistically significant to rule out relaxation of selective constraint.

From a biological perspective, it is unlikely that myostatin was removed from selective constraint, but that conservative selective pressures were subsequently retained without pseudogenization in all three independent lineages. A more reasonable explanation that accommodates the results from the parsimony based analysis is

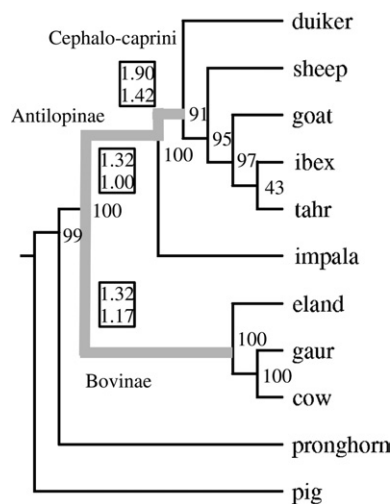


Fig. 3. Phylogenetic trees were constructed using DNAmI from Phylip (Felsenstein, 1989) and MRBAYES (Huelsenbeck and Ronquist, 2001), both using the whole DNA sequence and just the third codon positions. All four tree building calculations gave the same result, which is depicted with posterior probabilities from MRBAYES for the tree calculated from third codon positions (corresponding bootstrap values from DNAmI trees were generally slightly lower and ranged from 66 to 100). The tree and the DNA sequences from Fig. 2 were used to estimate the ratio of nonsynonymous to synonymous nucleotide substitution rate ratios (K_a/K_s) using both an approach involving parsimony ancestral sequence reconstruction and difference counting (Liberles, 2001) and a maximum likelihood approach (Yang, 1998) encoded in PAML (Yang, 1997). Three branches, depicted in thick gray, showed $K_a/K_s > 1$ using both the maximum likelihood and parsimony approaches for at least part of the sequences. In boxes next to all three branches, the K_a/K_s ratio from the first maximum likelihood approach is reported on top, with the K_a/K_s ratio from the parsimony approach on the bottom. The branch lengths measured in MRBAYES for the three positively selected branches closely approximate branch lengths measured in K_a using the parsimony based reconstruction. They are 0.015 for the Bovinae, 0.009 for Antilopinae, and 0.008 for Cephalo-caprini.

that a small subset of sites were under positive selective pressure with the vast majority under negative selective pressure. This hypothesis can also be tested in a maximum likelihood framework (Yang and Nielsen, 2002).

This maximum likelihood approach for detecting sites along specific branches under positive selective pressure was applied (Yang and Nielsen, 2002). A null hypothesis allowed sites and branches to be under negative selective pressure or evolving neutrally. In addition to negatively selected and neutrally evolving sites, models allowed for positively selected sites independently along each of the three branches previously indicated to be under positive selective pressure. Such positive selective pressure was again significantly supported, although with varying levels of statistical support (see Table 4). Combining all three branches into a single parameter increased the level of support for positive selection along the three lineages. The sites under positive selective pressure in the Bovinae and Antilopinae lineages corresponded with regions in both the propeptide and mature protein. Along the Cephalo-caprini lineage, all sites were indicated to be under positive selective pressure. This type of false positive from maximum likelihood methods has previously been reported (Suzuki and Nei, 2004).

The results from the first maximum likelihood K_a/K_s and the parsimony-based K_a/K_s methods are summarized in Fig. 3. Parsimony and maximum likelihood both point to positive selective pressure on the same three branches in myostatin in the early evolution of ruminant Artiodactyls.

With the false positive result from the maximum likelihood method to identify the regions of the myostatin gene under positive selective pressure along the Cephalo-caprini lineage, primary sequence windowing was done on the parsimony-based K_a results for the three positively selected lineages according to the analysis. K_a is used instead of K_a/K_s as a simple measure of variation along positively selected branches since positive selection acts on residues clustered together in a folded protein rather than simply along primary sequence (see Siltberg and Liberles, 2002 for a discussion of this).

As seen in Fig. 4A, substitution occurring during positive selective pressure in the lineage leading to Cephalo-caprini is detected localized in the propeptide region. In Figs. 4B and C, substitution during positive selective pressure is identified in both the propeptide and the C-terminal mature protein in the lineages leading to Bovinae and Antilopinae.

4. Discussion

Both parsimony and maximum likelihood analysis of the K_a/K_s ratios on the phylogenetic tree imply positive selective pressures in the evolution of myostatin in rumi-

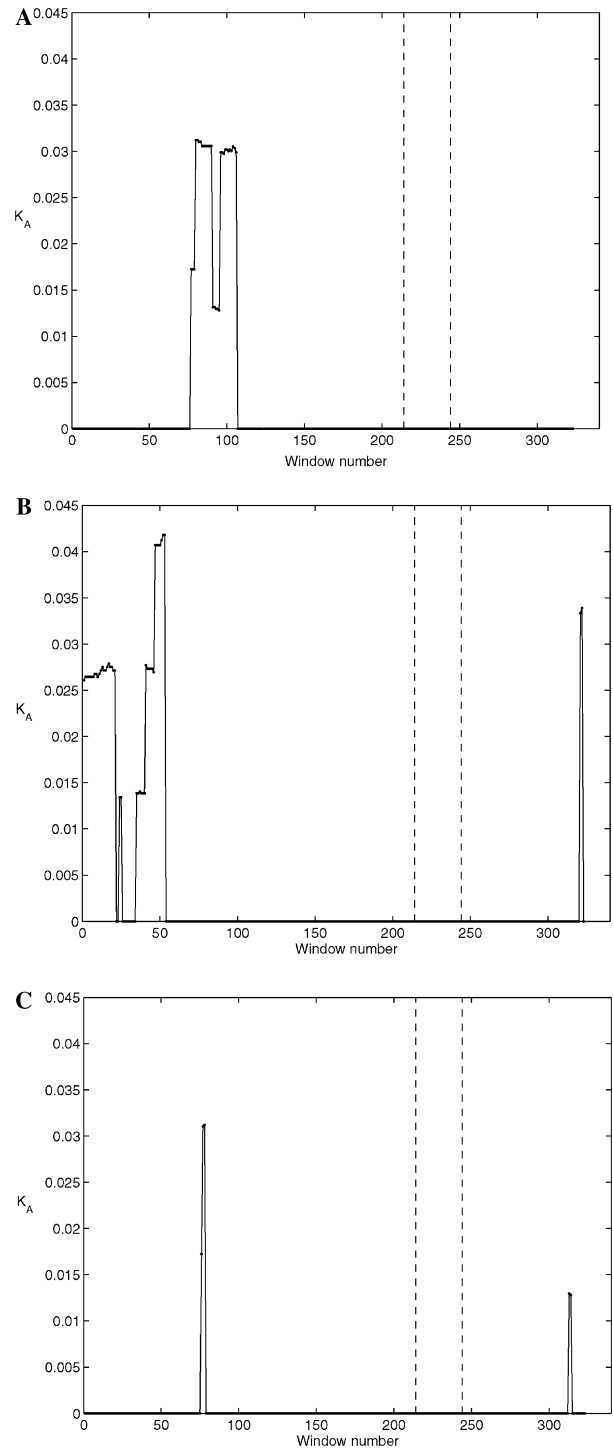


Fig. 4. (A) For primary sequence windows of 30 nucleotides, K_a was measured along the branch leading to Cephalo-caprini. Substitution is localized to the propeptide region. (B) Primary sequence K_a windowing of myostatin in 30 nucleotide blocks for the lineage leading to Bovinae shows substitution in both the propeptide and mature myostatin regions. (C) Primary sequence K_a windowing of myostatin in 30 nucleotide blocks for the lineage leading to Antilopinae also shows substitution in both the propeptide and mature myostatin regions. In all three figures, the signal peptide has been excluded and the boundary between the propeptide and mature protein is indicated with dotted lines to indicate the last window with only propeptide sequence and the first window with only mature protein.

nant Artiodactyls. That the substitutions occurring along these ancient branches have been subsequently conserved more recently in evolution implies that the proteins were functional rather than inactive both before and after the periods of positive selection. Given the ancestral nature of the selection events, subsequent conservation, and the large number of substitutions, it is unlikely that polymorphism in the gene today is related to ancestral polymorphism in the branches where positive selection was detected. From the conservation patterns across species, it appears that these changes have been fixed.

4.1. Myostatin and molecular structure

The positive selective pressures on myostatin have been localized to the regulatory propeptide along one branch and to both the propeptide and the mature C-terminal protein along the other two. While it is enticing to speculate about co-evolution of the interacting regulatory propeptide and mature protein, little is known about the residues responsible for this interaction, let alone their structures.

For the C-terminal mature protein, a close outparalog in the transforming growth factor beta superfamily (TGF- β), human bone morphogenetic protein 2 (BMP-2), has a crystal structure solved (Scheufler et al., 1999). This can be used as a rough template to analyze the substitutions that are occurring along each of the branches with high K_a/K_s ratios detected by the maximum likelihood approach. Of the substitutions, the most obvious candidates for a key functional role are two consecutive residues (356–357 in Fig. 2) near the expected receptor binding site that have changed from KE to EG along the lineage leading to Bovinae. An A to G substitution (position 366 in Fig. 2) has occurred along the lineage leading to Antilopinae at the expected dimer interface. Most of the remaining substitutions lie adjacent to the α -helix. As more structural information about myostatin, its propeptide, their interaction, and the interaction of myostatin with its target receptors becomes available, it will be possible to model the roles of such mutations on binding and regulatory events.

While the entire set of interacting partners for myostatin remains to be found, ActRIIB is the principal target in regulating myogenesis (Lee and McPherron, 2001; reviewed in Kocamis and Killefer, 2002). In addition to the propeptide, important co-evolution may also have occurred with other serum localized regulators. Myostatin signaling ultimately feeds into nuclear regulation of cell cycle progression.

4.2. Myostatin and phenotype

Ultimately, the positive selective pressures on myostatin may have been driven by phenotypic effects during

the time of divergence of Bovinae and Antilopinae. The divergence of Bovinae and Antilopinae was approximately 23 million years ago, as estimated through molecular dating (Hassanin and Douzery, 1999). The period 23–17 million years ago was a period of generally stable warm temperatures, possibly punctuated at the beginning by brief cooling at higher latitudes (Zachos et al., 1997). At around this time, the Bovinae moved south from Eurasian woodland environments into Africa, following the formation of a permanent land bridge around 24 million years ago (Jaeger, 2003; Kingdon, 1997). The first uncontested bovid (*Eotragus*) appeared in the fossil record around 17 million years ago, but had a body mass estimate of only around 20 kg (Solounias et al., 1995) (the correlation between body mass and skeletal muscle mass is not clear, but *Eotragus* is assumed to have a correspondingly ancestral skeletal muscle mass). However, by shortly after this time, increases in body mass were evident in some groups in both the Bovinae and Antilopinae lineages. One hypothesis is that selective pressures on myostatin drove this increase in body mass coupled to an increase in skeletal muscle mass, in turn driven by ecological changes in the environments of the various species. Myostatin may also only be part of a more complex story (especially plausible along the Antilopinae lineage), where substitution drove initial more modest changes in skeletal muscle and body mass, but set the stage for subsequent evolution driven both by myostatin and by other genes. Today, most Antilopinae are smaller, although roan antelope (*Hippotragus equinus*) can weigh up to 300 kg and muskox (*Ovibos moschatus*) up to 400 kg. Many bovine species can weigh more than 500 kg, although smaller (compared to other bovine species) more ancestral (in body mass) Bovinae like bushback (*Tragelaphus scriptus*) and 4-horned antelope (*Tetracerus quadricornis*) are also known. Myostatin may ultimately be one of several genes underlying changes in skeletal muscle and body mass.

With the sequencing and analysis of myostatin from several ruminant species, myostatin seems to be a crucial (but not the only) gene in regulating body musculature. It appears to have been under recent selective pressure in cattle as observed from population studies, probably occurring during domestication (Wiener et al., 2003), as well as having been under selective pressure during ruminant evolution and divergence during speciation. Further studies and insights will piece together the remaining links from ecology to genetics to biochemistry and structural biology in this important protein family.

Acknowledgments

We are thankful to Leona Chemnick and Oliver Ryder at the San Diego Zoo for providing us with

DNA samples. Rein Aasland provided laboratory space while Anna Söderholm, Matthew Betts and Christian Roth provided some technical assistance for which we are grateful. We thank Tim Hughes, Nathalie Reuter, Inge Jonassen and Eivind Coward for careful reading of this paper. Funding for this work was provided by a grant from the Carl Trygger Foundation and from the Swedish Foundation for Strategic Research and FUGE, the Norwegian Functional Genomics Platform.

References

- Anisimova, M., Bielawski, J.P., Yang, Z., 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol. Biol. Evol.* 18, 1585–1592.
- Felsenstein, J., 1989. PHYLIP—Phylogeny Inference Package. *Cladistics* 5, 164–166.
- Gonnet, G.H., Hallett, M.T., Korostensky, C., Bernardin, L., 2000. Darwin v. 2.0: an interpreted computer language for the bio-sciences. *Bioinformatics* 16, 101–103.
- Groves, P., Shields, G.F., 1996. Phylogenetics of the Caprinae-based on cytochrome b sequence. *Mol. Phylogenet. Evol.* 5, 467–476.
- Hassanin, A., Douzery, E.J.P., 1999. The tribal radiation of the family Bovidae (Artiodactyla) and the evolution of the mitochondrial cytochrome b gene. *Mol. Phylogenet. Evol.* 13, 227–243.
- Hassanin, A., Douzery, E.J.P., 2003. Molecular and morphological phylogenies of ruminantia and the alternative position of the moschidae. *Syst. Biol.* 52, 206–228.
- Hill, J.J., Davies, M.V., Pearson, A.A., Wang, J.H., Hewick, R.M., Wolfman, N.M., Qiu, Y., 2002. The myostatin propeptide and the follistatin-related gene are inhibitory binding proteins of myostatin in normal serum. *J. Biol. Chem.* 277, 40735–40741.
- Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17, 754–755.
- Jaeger, J.J., 2003. Isolationist tendencies. *Nature* 426, 509–510.
- Kingdon, J., 1997. *The Kingdon Field Guide to African Mammals*. Academic Press, London.
- Kocamis, H., Killefer, J., 2002. Myostatin expression and possible functions in animal muscle growth. *Dom. Anim. Endocrin.* 23, 447–454.
- Lee, S.J., McPherron, A.C., 1999. Myostatin and the control of skeletal muscle mass. *Curr. Opin. Gen. Dev.* 9, 604–607.
- Lee, S.J., McPherron, A.C., 2001. Regulation of myostatin activity and muscle growth. *Proc. Natl. Acad. Sci. USA* 98, 9306–9311.
- Liberles, D.A., 2001. Evaluation of methods for determination of a reconstructed history of gene sequence evolution. *Mol. Biol. Evol.* 18, 2040–2047.
- Liberles, D.A., Schreiber, D.R., Govindarajan, S., Chamberlin, S.G., Benner, S.A., 2001. The Adaptive Evolution Database (TAED). *Gen. Biol.* 2 (8), research0028.1-0028.6.
- Mathee, C.A., Davis, S.K., 2001. Molecular insights into the evolution of the family Bovidae: A nuclear DNA perspective. *Mol. Biol. Evol.* 18, 1220–1230.
- McPherron, A.C., Lawler, A.M., Lee, S.J., 1997. Regulation of skeletal muscle mass in mice by a new TGF-beta superfamily member. *Nature* 387, 83–90.
- McPherron, A.C., Lee, S.J., 1997. Double muscling in cattle due to mutations in the myostatin gene. *Proc. Natl. Acad. Sci. USA* 94, 12457–12461.
- Nicholas, G., Thomas, M., Langley, B., Somers, W., Patel, K., Kemp, C.F., Sharma, M., Kambadur, R., 2002. Titan-cap associates with, and regulates secretion of myostatin. *J. Cell. Physiol.* 193, 120–131.
- Nielsen, R., 2002. Mapping mutations on phylogenies. *Syst. Biol.* 51, 729–739.
- Rebbapragada, A., Benchabane, H., Wrana, J.L., Celeste, A.J., Attisano, L., 2003. Myostatin signals through a transforming growth factor B-like signaling pathway to block adipogenesis. *Mol. Cell. Biol.* 23, 7230–7242.
- Scheufler, C., Sebald, W., Hülsmeier, M., 1999. Crystal structure of human bone morphogenetic protein-2 at 2.7 Å resolution. *J. Mol. Biol.* 287, 103–115.
- Shah, P.K., Buslje, C.M., Sowdhamini, R., 2001. Structural determinants of binding and specificity in transforming growth factor-receptor interactions. *Proteins* 45, 408–420.
- Siltberg, J., Liberles, D.A., 2002. A simple covarion-based approach to analyse nucleotide substitution rates. *J. Evol. Biol.* 15, 588–594.
- Solounias, N., Barry, J.C., Bernor, R.L., Lindsay, E.H., Raza, S.M., 1995. The oldest bovid from the Siwaliks, Pakistan. *J. Vert. Paleont.* 15, 806–814.
- Suzuki, Y., Nei, M., 2004. False positive selection identified by ML-based methods: Examples from the *Sig1* gene of the diatom *Thalassiosira weissflogii* and the *tax* gene of a human T-cell lymphotropic virus. *Mol. Biol. Evol.* 21, 914–921.
- Tay, G.K., Iaschi, S.P.A., Bellinge, R.H.S., Chong, F.N., Hui, J., 2004. The development of sequence-based typing of myostatin (GDF-8) to identify the double muscling phenotype in the goat. *Small Ruminant Res.* 52, 1–12.
- Wang, H., Zhang, Q., Zhu, D., 2003. hSGT interacts with the N-terminal region of myostatin. *Biochem. Biophys. Res. Commun.* 311, 877–883.
- Wiener, P., Burton, D., Ajmone-Marsan, P., Dunner, S., Mommens, G., Nijman, I.J., Rodellar, C., Valentini, A., Williams, J.L., 2003. Signatures of selection? Patterns of microsatellite diversity on a chromosome containing a selected locus. *Heredity* 90, 350–358.
- Wolfman, N.M., McPherron, A.C., Pappano, W.N., Davies, M.V., Song, K., Tomkinson, K.N., Wright, J.F., Zhao, L., Sebald, S.M., Greenspan, D.S., Lee, S.J., 2003. Activation of latent myostatin by the BMP-1/tolloid family of metalloproteinases. *Proc. Natl. Acad. Sci. USA* 100, 15842–15846.
- Yang, Z., 1997. PAML: a program for phylogenetic analysis by maximum likelihood. *CABIOS* 13, 555–556.
- Yang, Z., 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* 15, 568–573.
- Yang, Z., Nielsen, R., 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* 19, 908–917.
- Zachos, J.C., Flower, B.P., Paul, H., 1997. Orbitally paced climate oscillations across the Oligocene/miocene boundary. *Nature* 388, 567–570.