ISSN 2095-8137 CN 53-1229/Q
Volume 36 Issue 1
18 January 2015

ZR
动物学研究

ZOOLOGICAL RESEARCH

CODEN: DOYADI
www.zoores.ac.cn

Science Press

ZOOLOGICAL RESEARCH

Volume 36, Issue 1 18 January 2015

CONTENTS

Cover image: Brown forest skink, *Sphenomorphus indicus* (Gray, 1853). Photo by Ying-Chao HU

ZOOLOGICAL RESEARCH

Genomic organization and evolution of ruminant lysozyme *c* genes

David M. IRWIN[1,2,*]

[1] *Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada*

[2] *Banting and Best Diabetes Centre, University of Toronto, Toronto, Canada*

ABSTRACT

Ruminant stomach lysozyme is a long established model of adaptive gene evolution. Evolution of stomach lysozyme function required changes in the site of expression of the lysozyme *c* gene and changes in the enzymatic properties of the enzyme. In ruminant mammals, these changes were associated with a change in the size of the lysozyme *c* gene family. The recent release of near complete genome sequences from several ruminant species allows a more complete examination of the evolution and diversification of the lysozyme *c* gene family. Here we characterize the size of the lysozyme *c* gene family in extant ruminants and demonstrate that their pecoran ruminant ancestor had a family of at least 10 lysozyme *c* genes, which included at least two pseudogenes. Evolutionary analysis of the ruminant lysozyme *c* gene sequences demonstrate that each of the four exons of the lysozyme *c* gene has a unique evolutionary history, indicating that they participated independently in concerted evolution. These analyses also show that episodic changes in the evolutionary constraints on the protein sequences occurred, with lysozyme *c* genes expressed in the abomasum of the stomach of extant ruminant species showing the greatest levels of selective constraints.

Keywords: Lysozyme *c*; Ruminants; Gene family; Gene duplication; Concerted evolution; Mosaic evolution

INTRODUCTION

Ruminant mammals such as cow, sheep, and deer, rely on foregut fermentation to extract nutrients from their diet of plant material (Clauss et al, 2010; Janis, 1976; Mackie, 2002; Stevens & Hume, 1998). Foregut fermentation, by bacteria and microbes, produces short chain fatty acids that are absorbed through the stomach wall and provide energy for the ruminant animals; however, the microbial population responsible for this fermentation incorporates many of the other nutrients, such as nitrogen based compounds, into their own growing populations (Mackie, 2002; Stevens & Hume, 1998). To extract these essential nutrients from the microbial population, ruminant animals must break open these bacterial and microbial cells, to release their contents, to allow the stomach digestive enzymes in the abomasum to extract nutrients from their contents (Stevens & Hume, 1998). Since bacterial cells are typically resistant to mammalian digestive enzymes, ruminant species have recruited the anti-bacterial enzyme, lysozyme *c*, to break open these cells (Callewaert & Michiels, 2010; Dobson et al, 1984; Irwin et al, 1992; Mackie, 2002; Prager & Jollès, 1996). Recruitment of lysozyme *c* as a digestive enzyme has occurred at least twice within mammals, on the lineages leading to the ruminant artiodactyls and the leaf-eating monkeys (Dobson et al, 1984; Stewart et al, 1987; Stewart & Wilson, 1987), with a similar recruitment of a calcium-binding lysozyme occurring in the hoatzin, a leaf-eating bird (Kornegay et al, 1994; Kornegay, 1996).

Recruitment of lysozyme *c* to become a digestive enzyme required changes both in the site of expression of the gene encoding this enzyme and in the amino acid sequence of the enzyme to allow function in the acidic stomach (Dobson et al, 1984; Irwin et al, 1992; Irwin, 1996; Prager, 1996). The major site of expression of lysozyme in mammals is macrophages, but it also secreted into some body fluids (such as tears), where it participates in host defense against bacterial infection (Callewaert & Michiels, 2010; Prager & Jollès, 1996; Short et al, 1996). The molecular basis for the recruitment of expression, at high levels, of lysozyme *c* in stomach cells is unknown. Typical mammalian lysozyme *c*

Received: 08 October 2014; Accepted: 02 December 2014

Foundation item: This study was supported by grants from the Natural Sciences and Engineering Research Council (RGPIN 183701)

*Corresponding author, E-mail: david.irwin@utoronto.ca

DOI:10.13918/j.issn.2095-8137.2015.1.1

enzymes function in an environment at a neutral pH, and one that is free of digestive enzymes (Callewaert & Michiels, 2010; Prager & Jollès, 1996; Prager, 1996). Lysozyme *c* function in the abomasum of the stomach of ruminant animals, to digest bacterial cell walls, required adapting the lysozyme *c* protein sequence to function at an acidic pH and becoming resistant to the actions of stomach digestive enzymes and acids found in the abomasum (Dobson et al, 1984; Jollès et al, 1989; Prager, 1996). A number of convergent amino acid changes were seen between the lysozyme *c* sequences that have adapted for function in the stomachs of the langur, a leaf-eating monkey, and ruminants, have been identified and presumed to account for much of the functional adaptation (Stewart & Wilson, 1987; Stewart et al, 1987; Swanson et al, 1991; Prager, 1996). Some of these adaptive changes include replacement of lysine residues with arginine, which removes potential cleavage sites for digestive enzymes found in the stomach, and the loss of an aspartate-proline dipeptide, which is an acid-labile peptide bond (Jollès et al, 1989; Prager, 1996; Stewart & Wilson, 1987; Stewart et al, 1987; Swanson et al, 1991). These putative adaptive amino acid replacements are inferred to occur early in ruminant evolution, and thus may parallel the origin and evolution of the ruminant lifestyle (Irwin et al, 1992; Irwin, 1996).

Recruitment of lysozyme *c* to a digestive role in ruminants is associated with an expansion of the size of the lysozyme *c* gene family (Jiang et al, 2014; Irwin & Wilson, 1989; Irwin et al, 1989, 1992). Most mammals have only one or a few lysozyme *c* genes, while ruminant species have 10 or more (Callewaert & Michiels, 2010; Irwin & Wilson, 1989; Irwin et al, 1989, 1996; Prager & Jollès, 1996; Irwin et al, 2011; Jiang et al, 2014). The lysozyme *c* gene family of the cow has been better characterized than those of most other ruminants, where it was found that only some of the genes are expressed in the abomasum, while others retain more ancestral type of roles (Irwin & Wilson, 1989; Irwin et al, 1993; Irwin, 2004). Similar observations have been made for the lysozyme *c* genes of sheep (Jiang et al, 2014). Several lysozyme *c* proteins, and their cDNAs, have been characterized from the abomasums of the cow, sheep and deer (Dobson et al, 1984; Jollès et al, 1989; Irwin & Wilson, 1989, 1990). Intriguingly, phylogenetic analysis of the coding and 3' untranslated portions of the lysozyme *c* cDNA sequences yielded different trees, with the coding sequences implying duplications of the genes on each species lineage and the 3 untranslated region indicating more ancient duplications before the divergence of these species (Irwin & Wilson, 1990). Selection at the protein level (e.g., lineage-specific adaptation of the protein sequences) does not explain the differences in the phylogenies for the two regions, as synonymous difference (those that do not change the coding potential) also yield the same conclusions. It was concluded that the differences in the phylogenies was due to concerted evolution, mediated by gene conversion, acting on the coding sequences, while the 3' untranslated regions only experienced divergent evolution (Irwin & Wilson, 1990; Irwin et al, 1992; Irwin, 1996; Wen & Irwin, 1999; Yu & Irwin, 1996). Characterization of the genomic sequences of lysozyme genes expressed in the abomasum from the cow and sheep suggested that the concerted evolution was limited to only the coding exons, and did not involve the intronic sequences separating these exons (Irwin et al, 1993; Wen & Irwin, 1999). An analysis of larger number of lysozyme mRNA sequences, including genes that are not expressed in the stomach, suggested that some of the genes expressed in non-stomach tissues might have also experienced concerted evolution (Irwin, 1995, 2004; Takeuchi et al, 1993).

The previous analyses were largely limited to lysozyme *c* genes expressed in ruminant species. With the recent completion of draft genomic sequences from several ruminant species (including, cow, yak, zebu, goat, Tibetan antelope, and sheep: Canavez et al, 2012; Dong et al, 2013; Ge et al, 2013; Jiang et al, 2014; Qiu et al, 2012; Zimin et al, 2009) it now possible to more completely characterize the complete complement of lysozyme *c* genes (including genenes that are not expressed) in the genomes of these species and examine the molecular evolution of these genes. Here we describe the lysozyme *c* gene complements of the cow and several other ruminant species. The lysozyme *c* gene cluster has largely been maintained within true ruminant (Infraorder Pecora) species. Analysis of these sequences shows that the ancestor of cow, sheep, and goats had 10 lysozyme *c* genes, several of which were pseudogenes that were retained by diverse species. The exons of the lysozyme *c* genes have differing evolutionary histories, suggesting that concerted evolution acted independently on each exon.

MATERIALS AND METHODS

Database searches

Previous searches of mammalian genomes indicated that the cow genome had about 12 lysozyme *c* genes located in a cluster on cow chromosome 5, many of which were incompletely annotated in the Ensembl assembly (Irwin et al, 2011). To better characterize the lysozyme *c* gene cluster in the cow (*Bos taurus*) genome we used the Blast algorithm (Altschul et al, 1990) to search the UMD 3.1 cow genome assembly (from Ensembl release 75 in June 2014; http://www.ensembl. org/index.html) with known and predicted cow lysozyme *c* cDNA and protein sequences. Lysozyme *c* genes from the sheep (*Ovis aries*; Oar_v3.1), pig (*Sus scrofa*; Sscrofa10.2), bottlenose dolphin (*Tursiops truncates*; Turtru1), dog (*Canis lupus familiaris*; CanFam3.1), panda (*Ailuropoda melanoleuca*; AilMel1), horse (*Equus caballus*; EquCab2), and rhinoceros (*Ceratotherium simum simum*; CerSimSim1 - preEnsembl) genomes from the Ensembl database were characterized by the approaches described above. A similar search strategy was used to identify lysozyme *c* genes in the yak (*Bos grunniens*), zebu (*Bos indicus*), water buffalo (*Babalus babalis*), Tibetan antelope (chiru; *Pantholopus hodgsonii*), goat (*Capra hircus*), aplaca

(*Vicugna pacos*), minke whale (*Balaenoptera acutorostrara scammoni*), killer whale (*Orcinus orca*), Yangtze River dolphin (*Lipotes vexillifer*), and sperm whale (*Physeter catodon*) genomes from the NCBI Genomes (chromosome), Whole-genome shotgun contigs (wgs), and Nucleotide collection (nr/nt) databases (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Genomic alignments and assignment of orthology

Genomic sequences encompassing lysozyme *c* genes were downloaded from the Ensembl and NCBI databases. Intron-exon boundaries of the 4 exons and the 5' and 3' flanking sequences of the new lysozyme *c* genes were annotated based on genomic alignments of genes using MultiPipMaker (Schwartz et al, 2000, 2003), using previously characterized artiodactyl lysozyme *c* genes (Irwin et al, 1993, 1996; Irwin, 1995; Yu & Irwin, 1996; Wen & Irwin, 1999) as guides. Gene neighborhood organization was assessed as previously described for lysozyme *c* genes (Irwin et al, 2011) with the flanking *Yeats*4 and *Cpsf*6 genes identified using Blast. Ruminant lysozyme *c* genes were named based on orthology (based on phylogeny, see below) and genomic location. Genes present in the common ancestor of sheep and cow were numbered (*Lyz*1-*Lyz*10), while lineage-specific duplicates have a letter (a-c) that follows the gene number. The alpaca lysozyme *c* genes were numbered arbitrarily, thus their numbers do not indicate orthology with the ruminant genes. All other species examined here have a single copy lysozyme *c* gene.

Phylogenetic analysis

Predicted protein coding sequences for lysozyme *c* cDNA sequences, extracted from the genomic alignments, were aligned with Muscle (Edgar, 2004) as implemented in Mega6.06 (Tamura et al, 2013). Alignments were edited manually to insert gaps to maintain open reading frames (due to the presence of frame shifting insertions in some pseudogenes). Phylogenetic trees were constructed by the maximum likelihood, neighbor-joining and parsimony methods using Mega6.06 (Tamara et al, 2013). Alternative phylogenetic hypothesis, derived from the phylogenies of the different exons, were tested using Tree-puzzle (Strimmer & von Haeseler, 1996) as implemented on the Mobyle @Pasteur web site (http://mobyle. pasteur.fr/cgi-bin/portal.py?#welcome; Néron et al, 2009).

RESULTS AND DISCUSSION

Number and organization of lysozyme *c* genes in the cow genome

Analyses of genomic Southern blots had concluded that there were about 10 lysozyme *c* genes in the cow genome (Irwin & Wilson, 1989; Irwin et al, 1989), with many of these genes clustered on chromosome 5 (Gallagher et al, 1993). A recent (2011) search of the Btau4.0 (2nd release, assembled 2007) of the cow genome sequence assembly identified 12 lysozyme *c* genes on chromosome 5 of the cow genome (Irwin et al, 2011). The genes identified in this search account for all of the previously characterized cow lysozyme *c* cDNA and protein sequences (Irwin et al, 2011). To better characterize the cow lysozyme *c* gene cluster, we searched the most current version (UMD3.1 – 3rd release) of the cow genome assembly (assembled 2009; Zimin et al, 2009) with Blast using the previously characterized lysozyme *c* cDNA and protein sequences. Our new searches identified a total of 14 lysosome *c* genes, 11 of which were annotated by Ensembl as genes (Table 1 and Figures 1, 2, and S1, supporting information at http://www.zoores.ac.cn/). The difference in the number of intact genes identified by the searches of the two different genome assemblies (12 in Btau4.0 and 11 in UMD3.1) is due to the earlier Btau4.0 assembly containing two copies of the tracheal lysozyme *c* gene (*CowC* and *CowD* in Irwin et al, 2011) while the most current assembly UMD 3.1 contains only single copy of this gene (here named *Lyz*2b).

In addition to the 11 annotated genes, each of which is composed of 4 exons consistent with the structure of a typical mammalian lysozyme *c* gene (Irwin et al, 1996; Callewaert & Michiels, 2010), Blast hits were found to map to additional locations that were distant from the annotated genes. Examination of these Blast hits suggested that they belong to three partial genes, which had not previously been annotated, with each being composed of only two, not four, exons (Table 1 and Figures 1, 2, and S1). The newly identified partial *Lyz*3a and *Lyz*3c genes contain exons 1 and 2 and exons 3 and 4, respectively, but are separated from each other by the *Lyz*3b gene that contains all 4 coding exons (Table 1 and Figures 1 and 2). No sequences similar to the missing exons were found near the *Lyz*3a and *Lyz*3c genes. The third partial gene, *Lyz*9, contains only exons 1 and 4, with the sequences between these exons showing no similarity to exons 2 or 3 of other lysozyme *c* genes. Most of the lysozyme *c* genes have the same orientation (annotated as the minus strand), but 4 of the 14 are on the opposite strand, indicating that the origin of this gene family is not just a simple series of tandem gene duplications. The cow genome, therefore, was found to contain 14 identifiable lysozyme *c* genes (Table 1 and Figures 1,2, and S1).

Pairwise sequence comparisons revealed that the DNA sequence identities of the coding sequences among the 14 genes ranged from 74.8% to 97.5%, with most pairs showing 80%-90% identity (Table 2). The similarity between *Lyz*3a and *Lyz*3c could not be measured, as there is no overlap between these two genes (*Lyz*3a has exons 3 and 4, while *Lyz*3c has exons 1 and 2, see Table 1). These two partial genes were most similar to the *Lyz*3b gene, showing greater than 96% identify in the coding sequence (Table 2), raising the possibility that they are recent gene duplicates. The *Lyz*3a, *Lyz*3b, and *Lyz*3c genes are also adjacent to each other in the genome, suggesting that the *Lyz*3a and *Lyz*3b gene were generated by partial tandem duplications of different parts of the *Lyz*3b gene (Figure 1). The partial gene *Lyz*9 did not show particularly strong similarity to any other specific cow lysozyme *c* gene (Table 2), suggesting

that it is not a product of a very recent segmental duplication event (Liu et al, 2009; Seo et al, 2013). Among the intact lysozyme genes, the coding sequence of the genes encoding the lysozymes expressed in the abomasum (Irwin & Wilson, 1989; Irwin et al, 1993) *Lyz*5/*Lyz*6/*Lyz*7 share about 97% identity and the *Lyz*2a/*Lyz*2b/*Lyz*2c genes, which includes the tracheal lysozyme gene (Takeuchi et al, 1993), share about 96% identity (Table 2). The high level of identity

Table 1 Locations of cow lysozyme *c* genes

Gene	Chromosome	Strand	Bases	Ensembl gene ID	Other name	Unigene	Sites of expression	Functional
Lyz1	5	Minus	44 421 190-44 426 117	ENSBTAG00000011941	Milk	Bt.67194: 448 ESTs	Rumen, omasum	Intact
Lyz2a	5	Plus	44 421 190-44 426 117	ENSBTAG00000022971		None[b]		Intact
Lyz2b	5	Plus	44 443 587-44 448 198	ENSBTAG00000000198	Trachael	Bt.64327: 121 ESTs	Rumen	Intact
Lyz3a	5	Minus	44 489 738-44 491 389	NA (exons 3 and 4)[a]		None[b]		Pseudogene
Lyz3b	5	Minus	44 502 011-44 507 108	ENSBTAG00000039170	ψNS4	None[b]		Pseudogene
Lyz3c	5	Minus	44 521 088-44 522 921	NA (exons 1 and 2)[a]		None[b]		Pseudogene
Lyz2c	5	Plus	44 533 912-44 538 847	ENSBTAG00000020564		Bt.105675: 31 ESTs	Rumen, intestine	Intact
Lyz4	5	Plus	44 554 351-44 559 113	ENSBTAG00000026323	Intestinal	Bt.49176: 162 ESTs	Intestine	Intact
Lyz5	5	Minus	44 573 344-44 578 495	ENSBTAG00000026088	Stomach 2	Bt.29367: 363 ESTs	Abomasum	Intact
Lyz6	5	Minus	44 599 815-44 607 109	ENSBTAG00000046511	Stomach 1	Bt.89770: 102 ESTs	Abomasum	Intact
Lyz7	5	Minus	44 631 778-44 638 267	ENSBTAG00000046628	Stomach 3	Bt.80498: 74 ESTs	Abomasum	Intact
Lyz8	5	Minus	44 652 535-44 656 613	ENSBTAG00000026322		None[b]		Pseudogene
Lyz9	5	Minus	44 673 817-44 676 391	NA (exons 1 and 4)[a]		None[b]		Pseudogene
Lyz10	5	Minus	44 713 289-44 720 946	ENSBTAG00000026779	Kidney	Bt.64645: 61 ESTs	Lymphoreticular, blood	Intact

[a] – Not annotated as a gene by Ensembl.

[b] – No ESTs with greater than 95% sequence identity identified in the NCBI EST database.

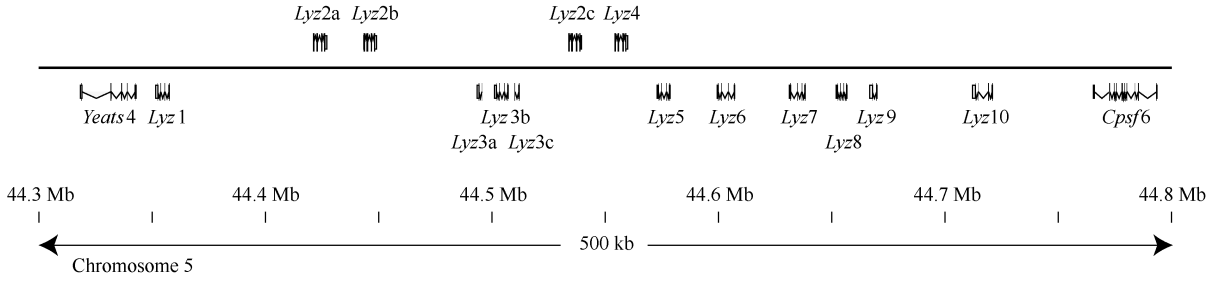

Figure 1 Organization of lysozyme *c* genes in the cow genome

Schematic of the arrangement of lysozyme *c* genes. And their neighbors, in the cow genome. Vertical lines represent exons, with splicing indicated by the lines joining the exons. Gene names are indicated above (plus strand) or below (minus strand) indicating strand with coding potential. Sizes of genes and distances are proportional. The genes are located between 44.3 and 44.8 Mb on chromosome 5.

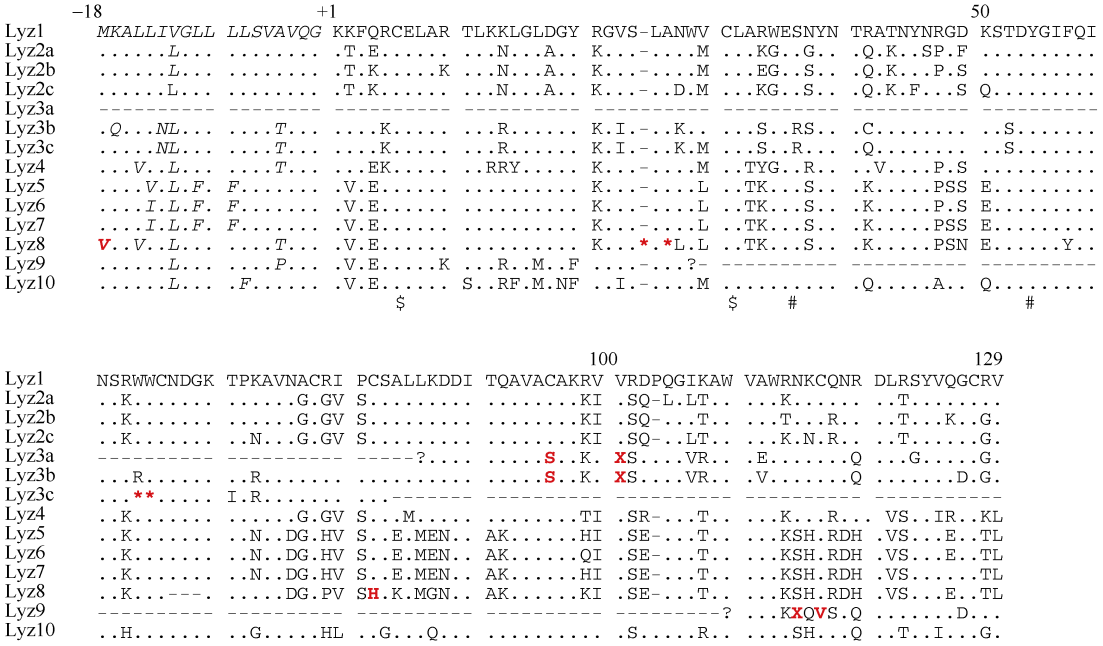

Figure 2 Amino acid sequences predicted by cow lysozyme *c* genes

Sequences of predicted lysozyme *c* proteins from the cow genome are shown in single letter code, with differences from the Cow Lyz1 sequence shown and identities indicated by dots (.). Sequence is numbered avove the sequences from the N-terminus of the Lyz1 sequence, with the signal peptide numbered backwards and in italics. Dashes (-) indicate gaps introduced to maximize alignment. Question marks (?) indicate incomplete codons due to missing sequence. Residues involved in disulfide bridging ($) and active site residues (#) are marked below the sequences. Residues marked in red are likely damaging pseudogenes and disrupt initiation, disrupt disulfide bridging or introduce stop codons. Asterisks (*) indicate inframe stop codons. Xs refer to codons that have less than 3 bases and thus cause frame shifts. The initiation codon of Lyz8 is not methionine (M).

Table 2 Pairwise percent DNA sequence identity between cow lysozyme *c* coding sequences

	Lyz2a	*Lyz2b*	*Lyz3a*	*Lyz3b*	*Lyz3c*	*Lyz2c*	*Lyz4*	*Lyz5*	*Lyz6*	*Lyz7*	*Lyz8*	*Lyz9*	*Lyz10*
Lyz1	89.4	88.7	91.3	91.3	90.7	89.0	86.5	82.2	83.1	82.9	80.9	85.6	89.0
Lyz2a		96.2	87.4	85.3	84.4	96.6	90.1	84.7	85.6	85.4	82.8	84.7	84.5
Lyz2b			88.1	85.3	84.1	96.8	90.5	84.9	85.8	85.4	82.8	83.2	84.5
Lyz3a				96.4	NA[a]	88.1	83.0	74.8	75.6	76.3	76.3	77.3	89.9
Lyz3b					96.3	85.3	85.3	79.0	79.5	79.7	77.6	84.2	86.5
Lyz3c						84.1	86.0	82.1	82.4	82.4	79.5	86.0	86.4
Lyz2c							89.6	84.9	85.8	85.6	82.8	83.7	84.7
Lyz4								84.7	84.7	85.4	83.0	83.7	83.3
Lyz5									96.4	97.5	92.2	83.2	82.9
Lyz6										96.8	92.2	84.2	82.0
Lyz7											92.4	84.7	82.9
Lyz8												82.7	80.0
Lyz9													85.6

[a] – These two genes do not overlap.

shared by these sets of genes suggests that these triplets are products of recent segmental duplication / gene duplication events (Liu et al, 2009; Seo et al, 2013). *Lyz5*/*Lyz6*/*Lyz7* are adjacent to each other and are in the same orientation (Table 1; Figure 1), thus could be generated by a simple series of tandem gene duplication

events. A more complicated duplication history is needed to explain the diversification of the *Lyz*2a/*Lyz*2b/*Lyz*2c genes. While *Lyz*2a and *Lyz*2b are in tandem, several other lysozyme *c* genes (*Lyz*3a, *Lyz*3b, and *Lyz*3c) are located between the *Lyz*2a/2b gene pair and the *Lyz*2c gene (Table 1 and Figure 1).

Since the three partial genes *Lyz*3a, *Lyz*3b, and *Lyz*9 do not contain all four coding exons; they cannot predict intact open reading frames. In addition to the missing exon sequence, all three of these genes also contain in frame stop codons or frameshifts that also would prevent translation (Figure 2). Among the lysozyme *c* genes possessing all four exons, two, *Lyz*3b and *Lyz*8, fail to predict intact open reading frames (Figure 2). *Lyz*3b, which was previously called lysozyme ψNS4 (Irwin, 1995) contains a frameshift, which is shared with *Lyz*3a, which prevents translation of the reading frame, while *Lyz*8 contains both in frame stop codons and a replacement at the initiating codon (Figure 2). Thus of the 14 lysozyme *c* genes found in the cow genome, only 9 potentially encode functional lysozyme *c* proteins. To further investigate the functional potential of these lysozyme *c* genes we searched for evidence of expression for all 14 cow lysozyme *c* genes in the NCBI expressed sequence tag (EST) database. ESTs were found for only 8 of the 9 intact genes, and for none of the 5 pseudogenes (Table 1). While *Lyz*2a has an intact open reading frame (Figure 2), no ESTs highly similar (>98% identity) to it were found in the NCBI database (Table 1), raising the possibility that this gene is not expressed.

Many of the cow lysozyme *c* gene annotations in the Ensembl database do not include 5' and/or 3' untranslated sequences. Since previous work had shown that the 5' and 3' untranslated sequences of known lysozyme *c* genes from diverse mammalian species have considerable sequence similarity (Irwin & Wilson, 1989; Irwin, 1995, 2004), we used this similarity to predict the extent of these regions for each gene (see Figure S1) from alignments generated by MultiPipMaker (Schwartz et al, 2000, 2003). Complete 5' untranslated regions could be predicted for all of the lysozyme *c* genes that had exon 1, however the full 3' untranslated regions could not be predicted for all exon 4 sequences, as the 3' end of the 3' untranslated region could not be found for the cow *Lyz*3a and *Lyz*3b genes (see Figure S1). This observation is consistent with an earlier failure to identify homologous sequences for the entire 3' untranslated region of the cow ψNS4 (*Lyz*3b) gene (Irwin, 1995, 2004).

Lysozyme *c* genes in other ruminant genomes

To better characterize the evolutionary history of the ruminant lysozyme *c* genes, we identified lysozyme *c* genes in the genomes of other ruminant species and their close relatives (Table 3 and Figures 3 and S1). As expected, from previous work (Callewaert & Michiels, 2010; Irwin et al, 1989, 1996, 2011; Prager & Jollès, 1996), only a single lysozyme *c* gene was found in the genomes of carnivores (dog, *Canis lupus familiaris*; and panda, *Ailuropoda melanoleuca*) and perrisodactyls (horse, *Equus caballus*; and rhinoceros, *Ceratotherium simum simum*) (Table 3 and Figures 3 and S1). The single lysozyme *c* gene in the outgroup species is located between the *Yeats*4 and *Cpsf*6 genes (Figure 3), as found in most other mammalian species (Irwin et al, 2011). This ancestral mammalian genomic arrangement has been retained in the cow, with the amplification of the lysozyme *c* genes occurring between the *Yeats*4 and *Cpsf*6 genes (Irwin et al, 2011) (Figures 1 and 3). The tylopod lineage (e.g., camels and alpacas) represents one branch of the earliest divergence within artiodactyls (Morgan et al, 2013; Romiguier et al, 2013), with these species being pseudoruminants with a simpler multi-chambered stomach than the true ruminants (Clauss et al, 2010; Janis, 1976; Mackie, 2002). Searches of the alpaca (*Vicugna pacos*) genome in the Ensembl database identified three genomic sequences encoding partial lysozyme *c* gene sequences, indicating that multiple lysozyme *c* genes exist in this genome (results not shown). Searches of the NCBI Genomes (chromosomes) database identified an updated larger genomic contig that predicted 4 complete lysozyme *c* genes (and included all of the gene sequences found in the Ensembl alpaca genome assembly) at one end of a contig sequence (Table 3 and Figure 3). The *Yeats*4 gene was found to be adjacent to one side of the lysozyme *c* gene cluster, however no genes were found on the other side of the lysozyme *c* gene cluster in this genomic contig (Figure 3). The presence of the *Yeats*4 gene adjacent to the alpaca lysozyme *c* genes suggests that a similar genomic neighborhood exists in alpaca, but since the lysozyme *c* genes were at one end of the genomic contig it is possible that additional unsequenced lysozyme *c* genes may exist in the alpaca genome. The pig (*Sus scrofa*) is a representative of the family Suidea, which is the next diverging lineage within artiodactyls (Morgan et al, 2013; Romiguier et al, 2013). As expected, and previously reported (Irwin et al, 1989; Yu & Irwin, 1996), only a single lysozyme gene is found in this species (Table 3, Figures 3 and S1). As previously reported (Irwin et al, 2011), the genomic neighborhood surrounding the pig lysozyme gene differs from that of other mammals, raising the possibility that this genomic area has experienced recombination (Figure 3). Cetaceans (e.g., whales and dolphins) fall within artiodactyls, thus yielding cetartiodactyla (Morgan et al, 2013; Romiguier et al, 2013). A single lysozyme *c* gene was identified in all five cetacean (bottlenose dolphin, *Tursiops truncates*; minke whale, *Balenoptera acutorostrata scammoni*; killer whale, *Orinus orca*; Yangtze river dolphin, *Lipotes vexilllifer*; and sperm whale, *Physeter catodon*) genomes (Table 3 and Figure S1), which is found in genomic location consistent with the ancestral genomic organization (Figure 3).

Pecoran artiodactyls (cow, sheep, deer, and relatives) are true ruminants with a stomach composed of four chambers (Clauss et al, 2010; Janis, 1976; Mackie, 2002). In addition to the sheep (*Ovis aires*) genome (Jiang et al, 2014), which is available from Ensembl, genome sequences of 5 other pecoran ruminant species (yak, *Bos grunniens* (Qiu et al,

Table 3 Locations of lysozyme *c* genes in diverse artiodactyls and relatives

Gene	Chromosome / scaffold	Strand	Bases	Ensembl gene ID / NCBI accession	Missing exons
Yak (*Bos grunniens*)					
Lyz1	NW_005394307	Minus	5 211-9 033	XM_005901148	
Lyz2	NW_005394307	Plus	75 876-81 020	NA[a]	1[b]
Lyz3	NW_005394198	Plus	7 980-12 653	XM_005900299	
Lyz4	NW_005394198	Minus	27 041-32 422	XM_005900300	
Lyz5	NW_005394198	Minus	53 712-61 005	XM_005900301	
Lyz6	NW_005394198	Minus	83 939-90 419	XM_005900302	
Lyz7	NW_005394198	Minus	104 585-108 706	NA[a]	
Lyz8	NW_005392857	Minus	16 134-18 709	NA[a]	2, 3
Lyz9	NW_005392857	Minus	61 437-66 594	XM_005886999	
Lyz10	NW_005394307	Minus	5 211-9 033	XM_005901148	
Zebu (*Bos indicus*)					
Lyz1	AGFL01046860	Minus	920-2 950	NA[a]	3, 4[b]
Lyz2a	AGFL01046876	Plus	10 835-14 065	NA[a]	
Lyz2b	AGFL01046877	Plus	11 353-15 933	NA[a]	
Lyz3	AGFL01046880	Minus	12 055-17 431	NA[a]	
Lyz2c	AGFL01046883	Plus	3 743-8 620	NA[a]	
Lyz4	AGFL01046890	Plus	880-5 642	NA[a]	
Lyz5	AGFL01046890	Minus	19 872-25 024	NA[a]	
Lyz6	AGFL01046892	Minus	13 424-20 718	NA[a]	
Lyz7	AGFL01046895	Minus	668-7 157	NA[a]	
Lyz8	AGFL01046896	Minus	13 855-17 933	NA[a]	
Lyz9	AGFL01046897	Minus	889-1 024	NA[a]	2, 3, 4[b]
Lyz10	AGFL01046900	Minus	2 248-9 934	NA[a]	
	AGFL01046860	Minus	920-2 950	NA[a]	3, 4[b]
Water buffalo (*Bubalus bubalis*)					
Lyz1	NW_005784949	Minus	16 204-27 136	XM_006058377	
Lyz2	NW_005785126	Plus	13 244-16 121	NA[a]	4[b]
Lyz3	NW_005785126	Minus	148 364-153 473	NA[a]	
Lyz4	NW_005785126	Plus	173 780-178 474	XM_006064264	
Lyz5	NW_005785126	Minus	194 055-199 224	XM_006064265	
Lyz6	NW_005785126	Minus	220 532-227 898	XM_006064266	
Lyz7	NW_005785126	Minus	245 385-151 911	XM_006064267	
Lyz8	NW_005785126	Minus	269 603-271 738	NA[a]	
Lyz9	NW_005785126	Minus	283 880-286 454	NA[a]	2, 3
Lyz10	NW_005785126	Minus	323 066-328 240	XM_006064268	
	NW_005784949	Minus	16 204-27 136	XM_006058377	
Tibetan antelope / Chiru (*Pantholops hodgsonii*)					
Lyz1	NW_005806187	Minus	948 632-954 099	XM_005957102	
Lyz2	NW_005806187	Plus	1 005 582-1 010 034	XM_005957103	
Lyz3	NW_005806187	Minus	1 045 148-1 051 559	XM_005957104	
Lyz4	NW_005806187	Plus	1 074 291-1 079 230	NA[a]	2, 3[b]
Lyz5	NW_005811703	Minus	9 411-10 957	XM_005968144	3, 4[b]
Lyz6	NW_005811703	Minus	31 570-36 493	XM_005968125	
Lyz7	NW_005811703	Minus	64 104-69 686	XR_318952	
Lyz8	NW_005811703	Minus	83 106-87 330	NA[a]	
Lyz9	NW_005811703	Minus	102 887-105 253	NA[a]	2, 3
Lyz10	NW_005811703	Minus	168 707-174 027	XM_005968126	
Goat (*Capra hircus*)					
Lyz1	NW_005100667	Minus	6 548 409-6 554 516	XM_005680191	
Lyz2	NW_005100667	Plus	6 607 855-6 612 306	XM_005680189	
Lyz3	NW_005100667	Minus	6 638 119-6 644 658	NA[a]	

Continued

Gene	Chromosome / scaffold	Strand	Bases	Ensembl gene ID / NCBI accession	Missing exons
Lyz4	NW_005100667	Plus	6 670 972-6 67 2304	NA[a]	1, 2[b]
Lyz5	NW_005100667	Minus	6 689 705-6 694 702	XM_005680192	
Lyz6	NW_005100667	Minus	6 715 083-6 719 993	NM_001287566	
Lyz7	NW_005100667	Minus	6 741 973-6 747 565	NA[a]	
Lyz8	NW_005100667	Minus	6 761 023-6 765 243	XM_005680235	
Lyz9	NW_005100667	Minus	6 783 196-6 785 752	NA[a]	2, 3
Lyz10	NW_005100667	Minus	6 834 500-6 839 640	NM_001285711	
Sheep (*Ovis aries*)					
Lyz1	3	Minus	150 165 176-150 170 352	ENSOARG00000020393	
Lyz2	3	Plus	150 225 205-150 229 630	ENSOARG00000020417	
Lyz3	JH921983.1	Minus	4 380-1 032	ENSOARG00000000543	
Lyz4	3	Plus	150 266 228-150 270 937	ENSOARG00000020429	
Lyz5	3	Minus	150 288 480-150 293 529	ENSOARG00000020393	
Lyz6	3	Minus	150 313 875-150 318 810	ENSOARG00000020439	
Lyz7	3	Minus	150 342 914-150 348 498	NA[a]	
Lyz8	3	Minus	150 362 122-150 366 351	ENSOARG00000020476	
Lyz9	3	Minus	150 385 062-150 387 578	NA[a]	2, 3
Lyz10	3	Minus	150 434 372-150 439 510	ENSOARG00000020515	
Pig (*Sus scrofa*)					
Lyz	5	Minus	36 179 949-36 185 488	ENSSSCG00000000492	5
Alpaca (*Vicugna pacos*)					
Lyz1	NT_167289.2	Minus	1 670 090-1 675 333	NA[a]	
Lyz2	NT_167289.2	Minus	1 707 719-1 713 452	NA[a]	
Lyz3	NT_167289.2	Plus	1 722 393-1 728 285	NA[a]	
Lyz4	NT_167289.2	Plus	1 760 703-1 766 608	NA[a]	
Bottlenose dolphin (*Tursiops truncates*)					
Lyz	scaffold_114746	Plus	182 136-187 936	ENSTTRG00000013948	
Minke whale (*Balaenoptera acutorostrata scammoni*)					
Lyz	NW_006733011	Minus	27 704 269-27 709 850	XM_007195043	
Killer whale (*Orcinus orca*)					
Lyz	NW_004438568	Plus	1 319 177-1 324 490	XM_004281877	
Yangtze River dolphin (*Lipotes vexillifer*)					
Lyz	NW_006790307	Minus	1 455 813-1 461 420	XM_007463554	
Sperm whale (*Physeter catodon*)					
Lyz	NW_006716048	Minus	6 880-11 985	XM_007118874	
Dog (*Canis lupus familiaris*)					
Lyz	10	Plus	11 346 500-11 350 639	ENSCAFG00000000426	
Panda (*Ailuropoda melanoleuca*)					
Lyz	GL192893.1	Plus	308 956-313 372	ENSAMEG00000011820	
Horse (*Equus caballus*)					
Lyz	6	Plus	84 276 158-84 280 173	ENSECAG00000018113	
Rhinoceros (*Ceratotherium simum simum*)					
Lyz	JH767750.1	Plus	25 463 742-25 467 903	ENSP00000261267_1	

[a] – Not annotated as a gene in Ensembl or NCBI.

[b] – Possibly missing due to incomplete gene.

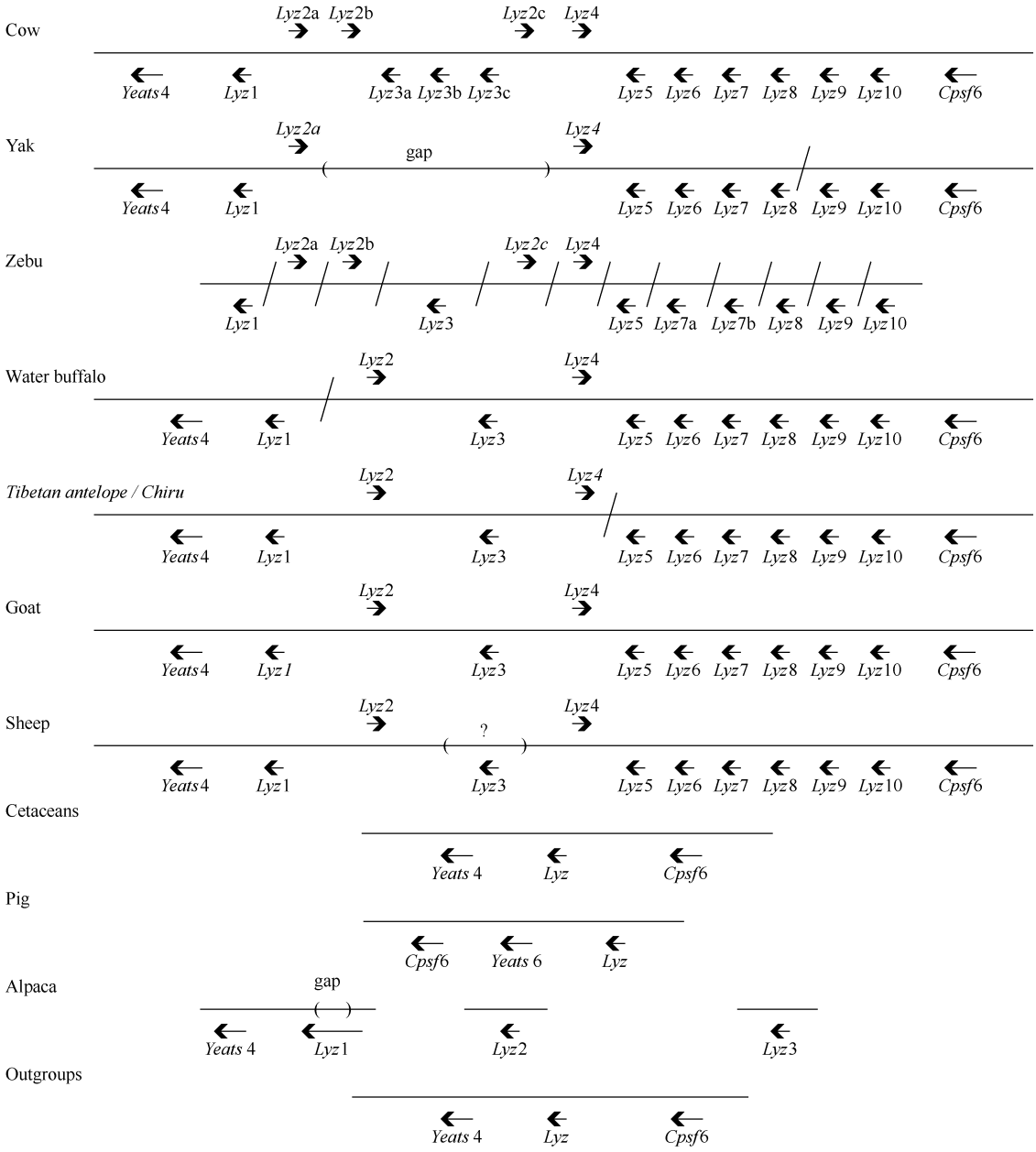

Figure 3 Organization of lysozyme *c* genes in diverse Artiodactyls and relatives

Schematic of the genomic arrangement of lysozyme *c* genes, and their neighbors, derived from genomic sequences in the Ensembl and NCBI databases. Sizes of genes, and distances between genes are not to scale. Genes shown above the lines are encoded by the plus strand, while those below are on the minus strand. Genomic sequences are listed in Table 3.

2012); zebu, *Bos indicus* (Canavez et al, 2012); water buffalo, *Bubalus bubalis*; Tibetan antelope (chiru) (Ge et al, 2013), *Pantholops hodgsonii*; and goat, *Capra hircus* (Dong et al, 2013)) are available in the NCBI database. The genomes of all pecoran ruminant species contained multiple lysozyme *c* genes (Table 3 and Figures 3 and S1), in accord with previous results (Irwin & Wilson, 1989; Irwin et al, 1989, 2011). For most pecoran species, lysozyme *c* genes could be mapped to large genomic contigs, or chromosomes, that show organizations similar to that seen in the cow (Table 3 and Figure 3). In the sheep, one gene (*Lyz*3) was not mapped to chromosome 3, but instead to an unmapped contig (Table 3). Since the goat genes all map to one contig (Table 3) it is possible that the sheep *Lyz*3 gene has been

misplaced (Figure 3), although movement to a new location through recombination cannot be excluded. The yak lysozyme *c* genes map to two contigs, with one containing a large gap that corresponds to the location where one or more missing lysozyme *c* genes might exist (Table 3 and Figure 3). The lysozyme *c* genes in both the Tibetan antelope and water buffalo map to two genomic contigs that might be adjacent in their genomes (Table 3 and Figure 3). Lysozyme *c* genes in the zebu are each on separate contigs, but could be arranged as seen in the cow and other pecoran species (Table 3 and Figure 3).

Mosaic evolutionary histories for exons of cow lysozyme *c* genes

To examine the evolutionary history of the cow lysozyme *c* genes, a phylogeny of the sequences was established. Phylogenetic trees were constructed for each exon of the lysozyme *c* genes (Figure 4) as previous analyses suggested that they might have experienced different histories (Irwin, 2004; Irwin & Wilson, 1990; Irwin et al, 1993, 1996; Wen & Irwin, 1999). As shown in Figure 4, different phylogenies were identified for each exons, with similar trees found if different outgroup species were used or if phylogenies were constructed using distance or parsimony methods or if only synonymous substitutions were used (results not shown). Some consistent phylogenetic patterns were observed across all exons, such as the clustering of the *Lyz*2a, *Lyz*2b, and *Lyz*2c genes and *Lyz*3a or *Lyz*3c being closest to *Lyz*3b (Figure 4). In contrast, the placement of some genes differed greatly between exons, such as the placement of *Lyz*1 or *Lyz*4 (Figure 4). To test whether there were statistically significant differences between the tree topologies estimated by each exon, we used Tree-puzzle (Strimmer & von Haeseler, 1996) to compare the four separate exon tree topologies with data for each exon. Despite the short lengths of some exons, at least two of the three alternative topologies could be excluded by all three of the KH, SH, and ELW statistical tests used by Tree-puzzle, with all three being excluded by at least one of the tests (Table 4). We cannot exclude the possibility that exons 2 and 3 share an identical evolutionary history, as these trees were not excluded by all three of the statistical tests, but exons 1 and 4 have evolutionary histories that are incompatible with each other and with exons 2 and 3 indicating that at least three different histories are represented by these four exons (Table 4). The differences in the topologies are unlikely to be due to convergent evolution acting on the lysozyme *c* protein sequences as the differences in the topologies were also seen when only synonymous differences were examined (results not shown).

Table 4 Phylogenies predicted from different cow lysozyme *c* gene exons are significantly different

Tree/Data	Log L	Difference	*SE*	KH[a]	SH[a]	ELW[a]	
Exon 1 Tree							
Exon 1 Data	–703.05	0.00		1.0000	1.0000	0.9425	BEST
Exon 2 Data	–738.42	35.37	13.1770	0.0080	0.0110	0.0001	EXCLUDED[b]
Exon 3 Data	–722.37	19.32	9.2124	0.0270	0.1170	0.0095	
Exon 4 Data	–733.42	30.37	14.3425	0.0280	0.0320	0.1540	EXCLUDED[b]
Exon 2 Tree							
Exon 1 Data	–844.36	113.94	22.9026	0.0000	0.0000	0.0000	EXCLUDED[b]
Exon 2 Data	–730.42	0.00		1.0000	1.0000	0.9438	BEST
Exon 3 Data	–746.39	15.97	8.3002	0.0400	0.3250	0.0136	
Exon 4 Data	–829.74	99.32	20.0394	0.0000	0.0000	0.0000	EXCLUDED[b]
Exon 3 Tree							
Exon 1 Data	–391.39	59.87	13.7897	0.0000	0.0000	0.0000	EXCLUDED[b]
Exon 2 Data	–343.01	11.49	6.7348	0.0580	0.3190	0.0136	
Exon 3 Data	–331.52	0.00		1.0000	1.0000	0.5948	BEST
Exon 4 Data	–365.58	34.06	13.4375	0.0070	0.0160	0.0002	EXCLUDED[b]
Exon 4 Tree							
Exon 1 Data	–1804.06	95.06	18.8423	0.0000	0.0000	0.0000	EXCLUDEDb
Exon 2 Data	–1780.35	71.35	20.9871	0.0010	0.0020	0.0000	EXCLUDED[b]
Exon 3 Data	–1767.23	58.23	20.2572	0.0020	0.0070	0.0000	EXCLUDED[b]
Exon 4 Data	–1709.00	0.00		1.0000	1.0000	0.9756	BEST

[a] – Probability of observing the tree, given the data, from the statistical one sided KH test based on pairwise SH tests (KH), the Shimodaira-Hasegawa test (SH), and the expected likelihood weight test (ELW) from Tree-puzzle (Strimmer & von Haesler, 1996).

[b] – Probability that the data is compatible with the tree is less than 0.05.

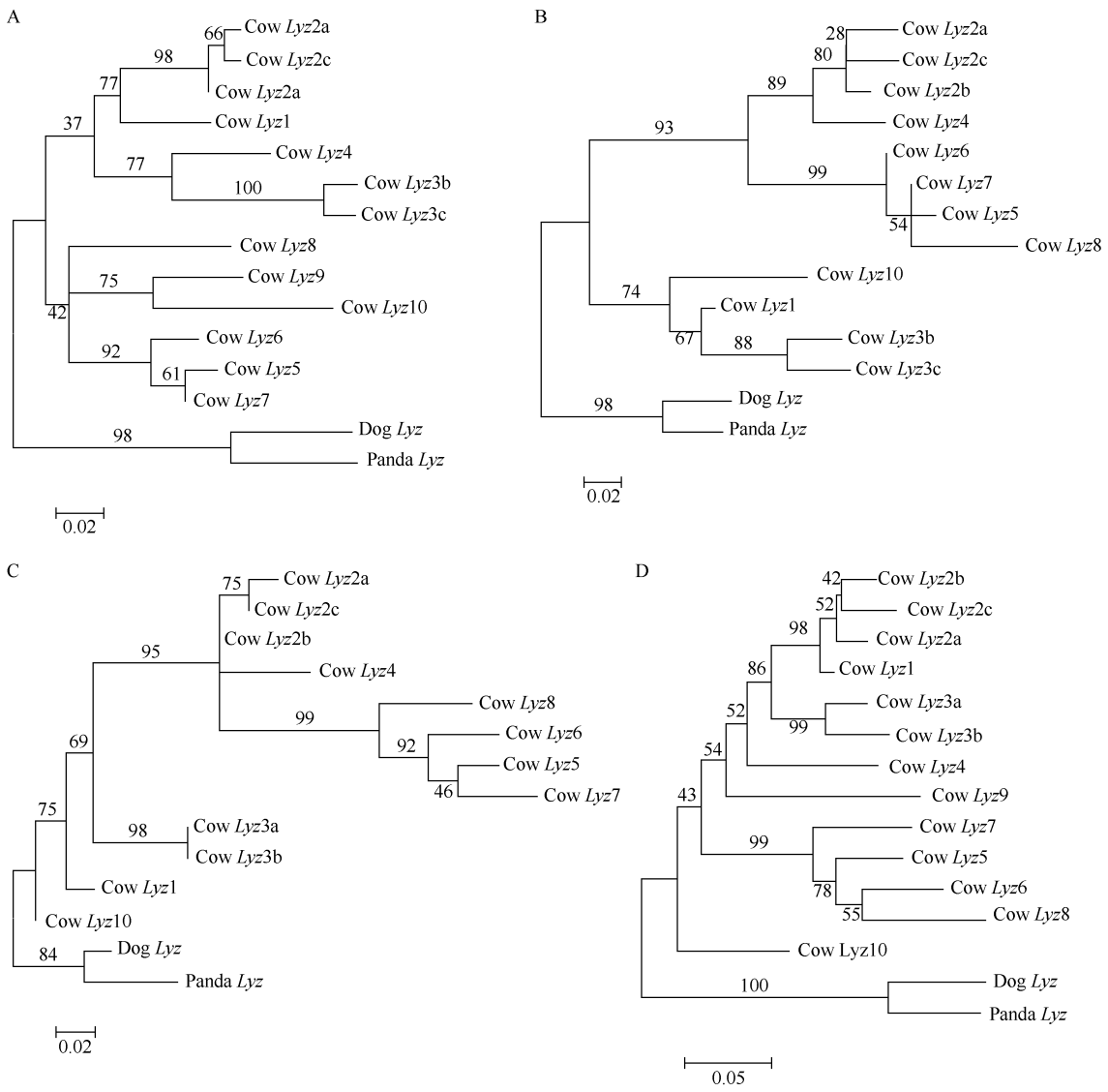

Figure 4 Phylogeny of cow lysozyme *c* genes derive from sequences of (A) exon 1, (B) exon 2, (C) exon 3, and (D) exon 4

Phylogenies for each of the 4 exons of the lysozyme *c* genes were estimated using maximum likelihood, as implemented in Mega6.06 (Tamura et al, 2013), using the Kimura 2-paramater model with a gamma distribution, which was the best fitting model for the sequence data. Similar results were obtained with the neighbor-joining method or parsimony, or the use of different outgroups. Phylogenies were generated from 152, 156, 74, and 306 aligned bases present in all sequences for exons 1, 2, 3, and 4, respectively. The presented phylogenies were bootstrapped 500 times.

These results are in agreement with previous conclusions that lysozyme *c* genes expressed in the abomasum of ruminants have experienced mosaic evolution due to gene conversion occurring between the coding exons (Irwin, 2004; Irwin & Wilson, 1990; Irwin et al, 1993, 1996; Wen & Irwin, 1999), and suggested that the 3' untranslated (exon 4) sequences likely best reflect the evolutionary history of the divergent genes, as this sequence appears to have experienced the fewest number of concerted evolution events.

Origin and evolutionary history of lysozyme *c* genes in ruminant genomes

To better examine the evolution of the duplicated lysozyme *c* genes in ruminant species, a phylogenetic tree was established for the lysozyme *c* sequences from the diverse ruminants (e.g., cow, sheep, and Tibetan antelope) and their close relatives (e.g., pig, cetaceans, and carnivores) (Figure 5). Exon 4 sequences were chosen to construct this phylogeny as they likely best reflect the divergence of the genes, and have experienced lower levels of concerted

evolution (see above). The phylogeny shown in Figure 5 was derived by maximum likelihood, and similar phylogenies were generated when neighbor-joining or parsimony was used (results not shown). The exon 4 phylogeny shown in Figure 5 of the lysozyme *c* genes yield strong evidence for the orthology of 8 of 10 types of lysozyme *c* genes found in ruminants (Figure 5). *Lyz*3, *Lyz*4, *Lyz*5, *Lyz*6, *Lyz*7, *Lyz*8, *Lyz*9, and *Lyz*10 orthology groups each have high (88%-100%) bootstrap support, with the species relationships within each group in general accord with the accepted species relationships (Figure 5). This observation implies that these 8 genes existed in the common ancestor of pecoran ruminants. The phylogenetic analysis did not resolve *Lyz*1 or *Lyz*2 genes as monophyletic groups, but instead suggested some intermixing of these genes (Figure 5). *Lyz*1 and *Lyz*2 sequences from species of tribe Bovine (cow, yak, zebu, and water buffalo) formed a moderately supported monophyletic group that had a primary divergence between the *Lyz*1 and *Lyz*2 sequences. The tribe Bovini *Lyz*1 and *Lyz*2 sequences were then grouped with *Lyz*1 sequences from the other pecoran ruminants (Tibetan antelope, goat and sheep), with the *Lyz*2 sequences from these same species being the outgoup to all of the *Lyz*1 and *Lyz*2 sequences. While it is possible that this distribution could be explained by an ancestor having four genes, and pairs of genes being lost in each species, an alternative explanation is that the pecoran ancestor possessed two genes, and that a concerted evolution event transferred sequences from the tribe Bovini *Lyz*1 exon 4 sequence to the tribe Bovini *Lyz*2 gene, resulting in the grouping of these sequences. Support for the monophyly of the *Lyz*1 and *Lyz*2 genes was found from phylogenies of exon 2 and exon 3 sequences (results not shown). These results suggest that the ancestor of pecoran ruminants possessed 10 lysozyme *c* genes.

While the ancestor of modern pecoran ruminants may have had 10 lysozyme *c* genes, several extant species have a higher number of genes, such as cow with 14 genes and the zebu with 12 genes (Tables 1 and 3). The increased numbers of lysozyme *c* genes in some ruminant species appear to be due to lineage-specific gene duplications. The phylogeny presented in Figure 5 implies lineage-specific duplications in three genes, *Lyz*2, *Lyz*3 and *Lyz*7, all of which occurred in species (cow and zebu) of the genus *Bos*. Both cow and zebu have three *Lyz*2 genes (*Lyz*2a, *Lyz*2b, and *Lyz*2c) (Tables 1 and 3). Only a single *Lyz*2 gene was found in the yak, however a gap in the genome assembly was found at this location (Table 3), thus it is possible that additional *Lyz*2 genes exist in this genome. Better assembly of the *Bos* genome sequences are needed to determine whether the triplicated genes have a single origin, or represent parallel duplication, a conclusion that does have some support from the phylogenetic analysis (Figure 5). Duplicated *Lyz*3 genes were only found in the cow, although the lack of this gene in the yak, potentially due to a gap in the assembly, and the poor assembly of the zebu genome do not rule out the possibility that multiple *Lyz*3 genes exist in these species (Tables 1 and 3). The duplications of the *Lyz*2 and *Lyz*3 genes in *Bos* likely represent products of segmental duplications (Liu et al, 2009; Seo et al, 2013). It is possible that segmental duplications may also exist in other pecoran ruminant species, but were collapsed as single genes during the genome sequence assembly process, and thus the increased numbers seen in the genus *Bos* simple reflect the better cow genome assembly.

The distribution of the numbers of lysozyme *c* genes in the genomes of ruminant species and their close relatives is consistent with an amplification of the lysozyme *c* gene on the lineage leading to true ruminants as previously proposed (Irwin & Wilson, 1989; Irwin et al, 1989, 1992, 2011; Yu & Irwin, 1996). In contrast, our current phylogenetic analysis of the 3' untranslated regions of lysozyme *c* genes suggests that the amplification of these genes was initiated very early in the artiodactyl lineage, before the divergence of the ruminants and tylopod (e.g., alpaca) lineages, and implying that the pig and cetaceans have lost genes, however these early divergences are very poorly supported (Figure 5). Indeed, phylogenetic analysis of exon 1, exon 2, or exon 3 sequences by themselves yielded differing conclusions concerning these earliest duplications, although again, none of these analysis yielded strong conclusions (results not shown). Analysis of larger amounts of genomic sequences (e.g., intronic and flanking sequence) potentially could resolve the order of the earliest divergences of the paralogous lysozyme *c* genes and cetartiodactyl species. While the alpaca has multiple lysozyme *c* genes (Table 3), our phylogenetic analysis suggests that they originated through a parallel series of lineage-specific independent duplications.

Rates of evolution in ruminant lysozyme *c* genes

Duplication of the lysozyme *c* gene on the ruminant lineage has allowed the specialization of gene expression in distinct tissues, such as different chambers of the stomach, and thus evolution of novel gene function (Callewaert & Michiels, 2010; Jiang et al, 2014; Irwin et al, 1992; Irwin, 1995, 2004; Prager & Jollès, 1996). Changes in the function of lysozyme *c* likely leads to changes in the evolutionary constraints acting upon these genes. To examine this issue we calculated the divergence at nonsynonymous and synonymous sites among lysozyme *c* genes, with the results from three divergent representatives of pecoran ruminants (cow, goat, and Tibetan antelope), and between genes in these three species and the single copy lysozyme *c* gene sequences found in pig and horse shown in Table 5 (similar results were seen with the other pecoran ruminant species). The relative rates of nonsynonymous to synonymous substitutions (dn/ds) varied between genes when compared among ruminants, from low values for the *Lyz*5 and *Lyz*6 genes, which imply that they are strongly constrained, to high values for the *Lyz*3 and *Lyz*9 genes, suggesting that there is little constraint on their protein sequences (Table 5). The cow *Lyz*3 and *Lyz*9 genes fail to predict intact open reading frames, suggesting that they are pseudogenes

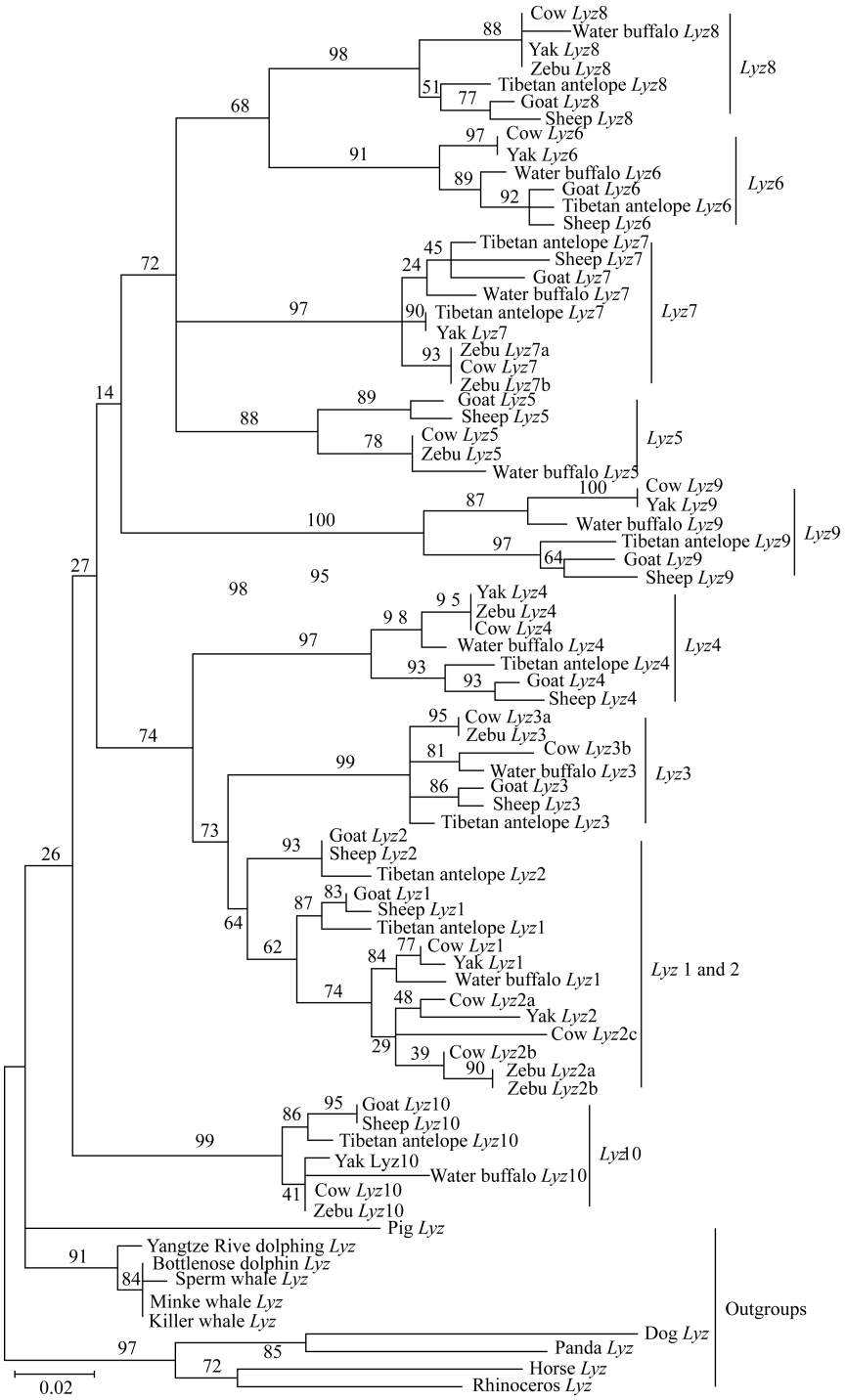

Figure 5 Phylogeny of ruminant lysozyme *c* genes derived from exon 4 sequences of predicted lysozyme *c* genes

The phylogeny of the lysozyme *c* genes was estimated from aligned exon 4 sequences (192 aligned bases in all sequences) using maximum likelihood, as implemented in Mega6.06 (Tamura et al, 2013), using the Kimura 2-paramater model with a gamma distribution, which was the best fitting model for the sequence data. Similar results were obtained with the neighbor-joining method or parsimony. The phylogeny was bootstrapped 500 times. Outgroups used to root the phylogeny are shown at the bottom. The ten types of lysozyme *c* genes are indicated on the right, with the bootstrap values that support 8 of these clades (all except the *Lyz*1 and *Lyz*2 clade) shown in bold.

Table 5 Rates of evolution of ruminant lysozyme *c* genes

	Lyz1			*Lyz2*			*Lyz3*			*Lyz4*[a]			*Lyz5*		
	dn	ds	dn/ds	dn	ds	dn/ds	dn	ds	dn/ds	dn	ds	dn/ds	dn	ds	dn/ds
Cow-Goat	0.026	0.043	0.611	0.052	0.081	0.640	0.073	0.054	1.350	0.031	0.048	0.647	0.022	0.114	0.197
Cow-Tibetan antelope	0.041	0.033	1.234	0.045	0.058	0.767	0.043	0.032	1.340	NA	NA	NA	NA	NA	NA
Goat-Tibetan antelope	0.026	0.043	0.610	0.007	0.066	0.105	0.028	0.021	1.314	NA	NA	NA	NA	NA	NA
Pecora average			0.818			0.504			1.334			0.647			0.197
Pig-Cow	0.184	0.327	0.564	0.246	0.342	0.720	0.268	0.268	0.998	0.230	0.264	0.872	0.224	0.354	0.632
Pig-Goat	0.192	0.317	0.604	0.231	0.318	0.726	0.271	0.259	1.048	NA	NA	NA	NA	NA	NA
Pig-Tibetan antelope	0.193	0.272	0.710	0.233	0.307	0.758	0.254	0.287	0.886	NA	NA	NA	NA	NA	NA
Pig average			0.626			0.735			0.977			0.872			0.632
Horse-Cow	0.099	0.438	0.226	0.180	0.467	0.386	0.129	0.459	0.281	0.193	0.345	0.560	0.197	0.394	0.501
Horse-Goat	0.083	0.451	0.184	0.171	0.441	0.388	0.144	0.379	0.380	NA	NA	NA	NA	NA	NA
Horse-Tibetan antelope	0.101	0.372	0.272	0.171	0.482	0.356	0.127	0.404	0.315	NA	NA	NA	NA	NA	NA
Horse average			0.227			0.377			0.325			0.560			0.501
	Lyz6			*Lyz7*			*Lyz8*			*Lyz9*			*Lyz10*		
	dn	ds	dn/ds	dn	ds	dn/ds	dn	ds	dn/ds	dn	ds	dn/ds	dn	ds	dn/ds
Cow-goat	0.025	0.128	0.195	0.043	0.098	0.441	0.058	0.067	0.867	0.052	0.047	1.097	0.033	0.058	0.573
Cow-Tibetan antelope	0.019	0.073	0.258	0.049	0.078	0.627	0.054	0.083	0.652	0.038	0.031	1.215	0.046	0.072	0.635
Goat-Tibetan antelope	0.025	0.065	0.385	0.049	0.071	0.692	0.023	0.028	0.825	0.029	0.014	1.996	0.012	0.026	0.446
Pecora average			0.279			0.587			0.781			1.436			0.551
Pig-Cow	0.215	0.411	0.523	0.229	0.382	0.599	0.256	0.389	0.658	0.247	0.352	0.701	0.191	0.270	0.708
Pig-Goat	0.238	0.427	0.557	0.278	0.445	0.625	0.265	0.316	0.840	0.272	0.393	0.691	0.187	0.248	0.753
Pig-Tibetan antelope	0.218	0.351	0.621	0.261	0.386	0.677	0.260	0.330	0.788	0.253	0.392	0.645	0.190	0.271	0.700
Pig average			0.567			0.634			0.762			0.679			0.721
Horse-Cow	0.185	0.432	0.429	0.193	0.419	0.461	0.221	0.449	0.491	0.129	0.368	0.349	0.113	0.351	0.323
Horse-Goat	0.204	0.411	0.497	0.227	0.457	0.498	0.241	0.416	0.579	0.169	0.406	0.415	0.101	0.366	0.277
Horse-Tibetan antelope	0.189	0.366	0.516	0.220	0.443	0.498	0.230	0.434	0.531	0.152	0.400	0.380	0.094	0.363	0.259
Horse average			0.481			0.485			0.534			0.381			0.286

[a] – Sheep *Lyz*4 used to replace the incomplete goat *Lyz*4 for comparisons.

(Table 1 and Figure 2) and thus should have no evolutionary constraints on their protein sequences.

The cow genome contains three *Lyz*3-like genes, with only one being a full-length gene sequence (*Lyz*3b), and a single copy of this gene was found in most of the other pecoran ruminant species (Tables 1 and 3). Cow *Lyz*3b gene was previously identified as the cow lysozyme *c* pseudogene NS4 (Irwin, 1995, 2004). While the *Lyz*3 gene sequences from tribe Bovini (cow, zebu and water buffalo) all share a frame shift mutation in exon 3 (amino acid residue 100 in Figure 2), which would prevent translation of a functional product, and additional mutations that potentially disrupt functions found in some sequences at other locations, the sequences from sheep, goat, and Tibetan antelope all predict a full-length open reading frame (Figure S1). This observation might suggest that the *Lyz*3 gene became a pseudogene, due to a frame-shifting mutation, on the lineage leading to tribe Bovini, after divergence from the other pecoran ruminant lineages. However, a high rate of nonsynonymous substitutions is also observed between the goat and Tibetan antelope *Lyz*3 gene sequences (Table 5) and between the sheep and both the goat and Tibetan antelope sequences (results not shown) suggesting that few evolutionary constraints were acting on this sequence and that this gene may have been non-functional in the common ancestor of all pecoran ruminants. It is possible that a mutation that prevented expression, or an amino acid substation that that prevented function, rather than a mutation that prevents translation of an intact product, was the initial mutation that created this pseudogene.

The second gene with a very high dn/ds ratio is the *Lyz*9 gene, which is composed of only 2 exons, exons 1 and 4, in the cow, due to the loss of exons 2 and 3 (Table 1 and

Figure 2). Orthologs of the *Lyz*9 genes in other pecoran ruminant species also have similar gene structures (Tables 1 and 3 and Figure S1), suggesting that this structure exited in the Lyz9 gene in the ancestor of all pecoran ruminants. The loss of exon 2 and 3 sequences from *Lyz*9 prevents the translation of a functional lysozyme, thus it can be concluded that this pseudogene originated before the radiation of the pecoran ruminants. Consistent with this conclusion, a high rate of divergence at nonsynonymous sites is observed in the *Lyz*9 gene sequence among all pecoran ruminant species (Table 5 and results not shown).

The *Lyz*3 and *Lyz*9 genes account for 4 of the five predicted lysozyme *c* pseudogenes in the cow genome (Table 1 and Figure 2). In addition to a pair of inframe stop codons (located between amino acid residues 24 and 25, and residue 26, Figure 2), the initiation codon for the cow *Lyz*8 gene is valine rather than methionine (amino acid -18 in Figure 2 and Figure S1). Orthologs of the *Lyz*8 gene from members of the tribe Bovini (yak, zebu, and water buffalo) share the inframe stop codons, as well other mutations such as a 9 base deletion in exon 2 (3 codons – residues 66-68 in Figure 2), while *Lyz*8 sequences from other pecoran species (sheep, goat and Tibetan antelope) do not possess any obvious harmful amino acid substitution, other than the valine substitution at the initiation codon (Figure S1). In contrast to the *Lyz*3 and *Lyz*9 pseudogenes, a much lower dn/ds ratio was observed in the pairwise comparisons among cow, goat, and Tibetan antelope (Table 5), which would be consistent with functional constraints acting on some, but not necessarily all, of the Lyz8 protein sequences. These observations appear to suggest, that despite the replacement of the initiator methionine with valine, the Lyz8 protein sequences in the sheep, goat and Tibetan antelope is functional, while a mutation occurred on the lineage leading to tribe Bovini to producing the *Lyz*8 pseudogene. How a functional protein can be translated from the *Lyz*8 gene, or evolutionary constraints that mirror protein function, is unclear. A downstream ATG, at codon 85 of the mature protein sequence (Figure 2), would be predicted to yield a protein of only 45 amino acid residues, far shorter than the typical 145 amino acid long protein lysozyme *c* precursor, with most of the sequence not being translated and thus not under evolutionary constraint for protein function.

Episodic evolution of ruminant lysozyme *c* genes

The cow lysozyme *c* genes displaying the lowest dn/ds ratios among ruminant species, and thus implying the strongest evolutionary constraints, are the *Lyz*5 and *Lyz*6 genes (Table 5), which are expressed in the abomasum (Table 1). Lysozyme *c* genes expressed predominantly in non-stomach tissues (Irwin, 2004), such as *Lyz*1, in milk, *Lyz*2, in the trachea, and *Lyz*10, in macrophages, have intermediate dn/ds ratios, but ratios that lower than those seen for the *Lyz*3, *Lyz*8, and *Lyz*9 pseudogenes (Table 5). However, when the dn/ds ratios are calculated between the ruminant genes (cow, goat, and Tibetan antelope) and an outgroup sequence (pig or horse), the stomach expressed *Lyz*5 and *Lyz*6 genes are seen to have dn/ds ratios that are either similar (when pig is the outgroup) or higher (horse being the outgroup) than those seen for the non-stomach (*Lyz*1, *Lyz*2, and *Lyz*10) genes (Table 5). To obtain this pattern of results, these observations suggest that the dn/ds ratio on the common ancestral lineage leading to the ruminants, after divergence from pig or horse, but before radiation of the pecoran ruminants, was higher for the lysozyme *c* genes expressed in the abomasum (*Lyz*5 and *Lyz*6) than for those expressed in non-stomach tissues (*Lyz*1, *Lyz*2, and *Lyz*10). This suggests that the rates of evolution of lysozyme *c* genes expressed in the abomasum display an episodic pattern, with more rapid evolution on the early ruminant lineage, and a slower rate within the pecoran ruminants. These results are consistent with previous findings of accelerated evolution of lysozyme *c* protein sequences obtained from the abomasum of ruminant species (Jollès et al, 1989; Irwin & Wilson, 1990; Irwin et al, 1992, 1993).

CONCLUSIONS

Genome sequences have advanced our understanding of the evolution of the lysozyme *c* gene family in ruminant species. Genomic sequences from seven divergent pecoran ruminant species allowed us to demonstrate that the genome of the pecoran ruminant common ancestor possessed at least 10 lysozyme *c* genes, and that these genes have largely been retained by extant ruminant species. More recent gene duplication, likely via segmental duplications (Liu et al, 2009; Seo et al, 2013), have resulted in increases in the number of lysozyme *c* genes on some lineages, with 14 genes found in the cow, but we can not exclude the possibility that some duplications may have been missed during assembly of some genomes. Lysozyme *c* genes have not evolved in a simple divergent manner, but rather by concerted evolution acting independently on each exon, yielding differing phylogenetic relationships for the ten types of lysozyme *c* genes. Some lysozyme *c* genes have become pseudogenes, either due to mutations in their coding sequence (e.g., *Lyz*3 and *Lyz*8) or by deletion of exon sequences (e.g., *Lyz*9). Some pseudogenes may have been generated by incomplete duplication of genes, such as *Lyz*3a and *Lyz*3c in the cow. Despite being presumably non-functional, at least two pseudogenes that exited in the ancestral pecoran ruminant (*Lyz*3 and *Lyz*9) have been retained in diverse ruminant species. A third lysozyme *c* gene has its initiation codon mutated to valine (from methionine), yet shows evidence that its coding sequence is evolutionary constrained on some ruminant lineages. This suggests that some lysozyme *c* pseudogenes may retain biological functions, however, how protein function in this sequence is maintained is unclear. Changes in the rates of nonsynonymous substitutions suggest that changes have occurred in the functional constraints acting on lysozyme *c* protein sequences, and these changes have occurred in an episodic fashion.

ACKNOWLEDGEMENTS

I thank the editor and two reviewers for helpful comments that have improved the manuscript.

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology,* **215**(3): 403-410.

Callewaert L, Michiels CW. 2010. Lysozymes in the animal kingdom. *Journal of Biosciences,* **35**(1): 127-160.

Canavez FC, Luche DD, Stothard P, Leite KR, Sousa-Canavez JM, Plastow G, Meidanis J, Souza MA, Feijao P, Moore SS, Camara-Lopes LH. 2012. Genome sequence and assembly of *Bos indicus*. *Journal of Heredity,* **103**(3): 342-348.

Clauss M, Hume ID, Hummel J. 2010. Evolutionary adaptations of ruminants and their potential relevance for modern production systems. *Animal,* **4**(7): 979-992.

Dobson DE, Prager EM, Wilson AC. 1984. Stomach lysozymes of ruminants. I. Distribution and catalytic properties. *Journal of Biological Chemistry,* **259**(18): 11607-11616.

Dong Y, Xie M, Jiang Y, Xiao NQ, Du XY, Zhang WG, Tosser-Klopp G, Wang JH, Yang S, Liang J, Chen WB, Chen J, Zeng P, Hou Y, Bian C, Pan SK, Li YX, Liu X, Wang WL, Servin B, Sayre B, Zhu B, Sweeney D, Moore R, Nie W, Shen Y, Zhao R, Zhang G, Li J, Faraut T, Womack J, Zhang Y, Kijas J, Cockett N, Xu X, Zhao S, Wang J, Wang W. 2013. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nature Biotechnology,* **31**(2): 135-141.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research,* **32**(5): 1792-1797.

Gallagher DS Jr, Threadgill DW, Ryan AM, Womack JE, Irwin DM. 1993. Physical mapping of the lysozyme gene family in cattle. *Mammalian Genome,* **4**(7): 368-373.

Ge RL, Cai Q, Shen YY, San A, Ma L, Zhang Y, Yi X, Chen Y, Yang L, Huang Y, He R, Hui Y, Hao M, Li Y, Wang B, Ou X, Xu J, Zhang Y, Wu K, Geng C, Zhou W, Zhou T, Irwin DM, Yang Y, Ying L, Bao H, Kim J, Larkin DM, Ma J, Lewin HA, Xing J, Platt RN 2nd, Ray DA, Auvil L, Capitanu B, Zhang X, Zhang G, Murphy RW, Wang J, Zhang YP, Wang J. 2013. Draft genome sequence of the Tibetan antelope. *Nature Communications,* **4**: 1858.

Irwin DM. 1995. Evolution of the bovine lysozyme gene family: changes in gene expression and reversion of function. *Journal of Molecular Evolution,* **41**(3): 299-312.

Irwin DM. 1996. Molecular evolution of ruminant lysozymes. *In*: Jollès P. Lysozymes: Model Enzymes in Biochemistry and Molecular Biology. Basel: Birkhäuser Verlag, 347-361.

Irwin DM. 2004. Evolution of cow nonstomach lysozyme genes. *Genome,* **47**(6): 1082-1090.

Irwin DM, Wilson AC. 1989. Multiple cDNA sequences and the evolution of bovine stomach lysozyme. *Journal of Biological Chemistry,* **264**(19): 11387-11393.

Irwin DM, Wilson AC. 1990. Concerted evolution of ruminant stomach lysozymes. Characterization of lysozyme cDNA clones from sheep and deer. *Journal of Biological Chemistry,* **265**(9): 4944-4952.

Irwin DM, Sidow A, White RT, Wilson AC. 1989. Multiple genes for ruminant lysozymes. *In*: Smith-Gill SJ, Sercarz EE. The Immune Response to Structurally Defined Proteins. NY: Adenine Press, 73-85.

Irwin DM, Prager EM, Wilson AC. 1992. Evolutionary genetics of ruminant lysozymes. *Animal Genetics,* **23**(3): 193-202.

Irwin DM, White RT, Wilson AC. 1993. Characterization of the cow stomach lysozyme genes: repetitive DNA and concerted evolution. *Journal of Molecular Evolution,* **37**(4): 355-366.

Irwin DM, Yu M, Wen Y. 1996. Isolation and characterization of vertebrate lysozyme genes. *In*: Jollès P. Lysozymes: Model Enzymes in Biochemistry and Molecular Biology. Basel: Birkhäuser Verlag, 225-241.

Irwin DM, Biegel JM, Stewart CB. 2011. Evolution of the mammalian lysozyme gene family. *BMC Evolutionary Biology,* **11**: 166.

Janis C. 1976. The evolutionary strategy of Equidae and the origins of rumen and cecal digestion. *Evolution,* **30**(4): 757-774.

Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, Wu C, Muzny DM, Li Y, Zhang W, Stanton JA, Brauning R, Barris WC, Hourlier T, Aken BL, Searle SM, Adelson DL, Bian C, Cam GR, Chen Y, Cheng S, DeSilva U, Dixen K, Dong Y, Fan G, Franklin IR, Fu S, Fuentes-Utrilla P, Guan R, Highland MA, Holder ME, Huang G, Ingham AB, Jhangiani SN, Kalra D, Kovar CL, Lee SL, Liu W, Liu X, Lu C, Lü T, Mathew T, McWilliam S, Menzies M, Pan S, Robelin D, Servin B, Townley D, Wang W, Wei B, White SN, Yang X, Ye C, Yue Y, Zeng P, Zhou Q, Hansen JB, Kristiansen K, Gibbs RA, Flicek P, Warkup CC, Jones HE, Oddy VH, Nicholas FW, McEwan JC, Kijas JW, Wang J, Worley KC, Archibald AL, Cockett N, Xu X, Wang W, Dalrymple BP. 2014. The sheep genome illuminates biology of the rumen and lipid metabolism. *Science,* **344**(6188): 1168-1173.

Jollès J, Jollès P, Bowman BH, Prager EM, Stewart CB, Wilson AC. 1989. Episodic evolution in the stomach lysozymes of ruminants. *Journal of Molecular Evolution,* **28**(6): 528-535.

Kornegay JR, Schilling JW, Wilson AC. 1994. Molecular adaptation of a leaf-eating bird: stomach lysozyme of the hoatzin. *Molecular Biology and Evolution,* **11**(6): 921-928.

Kornegay JR. 1996. Molecular genetics and evolution of stomach and nonstomach lysozymes in the hoatzin. *Journal of Molecular Evolution,* **42**(6): 676-684.

Liu GE, Ventura M, Cellamare A, Chen L, Cheng Z, Zhu B, Li C, Song J, Eichler EE. 2009. Analysis of recent segmental duplications in the bovine genome. *BMC Genomics,* **10**: 571.

Mackie RI. 2002. Mutualistic fermentative digestion in the gastrointestinal tract: diversity and evolution. *Integrative and Comparative Biology,* **42**(2): 319-326.

Morgan CC, Foster PG, Webb AE, Pisani D, McInerney JO, O'Connell MJ. 2013. Heterogeneous models place the root of the placental mammal phylogeny. *Molecular Biology and Evolution,* **30**(9): 2145-2156.

Néron B, Ménager H, Maufrais C, Joly N, Maupetit J, Letort S, Carrere S, Tuffery P, Letondal C. 2009. Mobyle: a new full web bioinformatics framework. *Bioinformatics,* **25**(22): 3005-3011.

Prager EM. 1996. Adaptive evolution of lysozyme: changes in amino acid sequence, regulation of expression and gene number. *In*: Jollès P. Lysozymes: Model Enzymes in Biochemistry and Molecular Biology. Basel: Birkhäuser Verlag, 323-345.

Prager EM, Jollès P. 1996. Animal lysozymes *c* and *g*: an overview. *In*:

Jollès P. Lysozymes: Model Enzymes in Biochemistry and Molecular Biology. Basel: Birkhäuser Verlag, 9-31.

Qiu Q, Zhang GJ, Ma T, Qian WB, Wang JY, Ye ZQ, Cao CC, Hu QJ, Kim J, Larkin DM, Auvil L, Capitanu B, Ma J, Lewin HA, Qian XJ, Lang YS, Zhou R, Wang LZ, Wang K, Xia JQ, Liao SG, Pan SK, Lu X, Hou HL, Wang Y, Zang XT, Yin Y, Ma H, Zhang J, Wang ZF, Zhang YM, Zhang DW, Yonezawa T, Hasegawa M, Zhong Y, Liu WB, Zhang Y, Huang ZY, Zhang SX, Long RJ, Yang HM, Wang J, Lenstra JA, Cooper DN, Wu Y, Wang J, Shi P, Wang J, Liu JQ. 2012. The yak genome and adaptation to life at high altitude. *Nature Genetics,* **44**(8): 946-949.

Romiguier J, Ranwez V, Delsuc F, Galtier N, Douzery EJ. 2013. Less is more in mammalian phylogenomics: AT-rich genes minimize tree conflicts and unravel the root of placental mammals. *Molecular Biology and Evolution,* **30**(9): 2134-2144.

Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W. 2000. PipMaker--a web server for aligning two genomic DNA sequences. *Genome Research,* **10**(4): 577-586.

Schwartz S, Elnitski L, Li M, Weirauch M, Riemer C, Smit A; NISC Comparative Sequencing Program, Green ED, Hardison RC, Miller W. 2003. MultiPipMaker and supporting tools: Alignments and analysis of multiple genomic DNA sequences. *Nucleic Acids Research,* **31**(13): 3518-3524.

Seo S, Larkin DM, Loor JJ. 2013. Cattle genomics and its implications for future nutritional strategies for dairy cattle. *Animal,* **7**(S1): 172-183.

Short ML, Nickel J, Schmitz A, Renkawitz R. 1996. Lysozyme gene expression and regulation. *In*: Jollès P. Lysozymes: Model Enzymes in Biochemistry and Molecular Biology. Basel: Birkhäuser Verlag, 243-257.

Stevens CE, Hume ID. 1998. Contributions of microbes in vertebrate gastrointestinal tract to production and conservation of nutrients. *Physiological Reviews,* **78**(2): 393-427.

Stewart CB, Wilson AC. 1987. Sequence convergence and functional adaptation of stomach lysozymes from foregut fermenters. *Cold Spring Harbor Symposium on Quantitative Biology,* **52**: 891-899.

Stewart CB, Schilling JW, Wilson AC. 1987. Adaptive evolution in the stomach lysozymes of foregut fermenters. *Nature,* **330**(6146): 401-404.

Strimmer K, von Haeseler A. 1996. Quartet puzzling: A quartet maximum likelihood method for reconstructing tree topologies. *Molecular Biology and Evolution,* **13**(7): 964-969.

Swanson KW, Irwin DM, Wilson AC. 1991. Stomach lysozyme gene of the langur monkey: tests for convergence and positive selection. *Journal of Molecular Evolution,* **33**(5): 418-425.

Takeuchi K, Irwin DM, Gallup M, Shinbrot E, Kai H, Stewart CB, Basbaum C. 1993. Multiple cDNA sequences of bovine tracheal lysozyme. *Journal of Biological Chemistry*, **268**(36): 27440-27446.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution,* **30**(12): 2725-2729.

Wen Y, Irwin DM. 1999. Mosaic evolution of ruminant stomach lysozyme genes. *Molecular Phylogenetics and Evolution,* **13**(3): 474-482.

Yu M, Irwin DM. 1996. Evolution of stomach lysozyme: the pig lysozyme gene. *Molecular Phylogenetics and Evolution,* **5**(2): 298-308.

Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, Hanrahan F, Pertea G, Van Tassell CP, Sonstegard TS, Marçais G, Roberts M, Subramanian P, Yorke JA, Salzberg SL. 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biology,* **10**(4): R42.

Zoological Research Editorial Board

ZOOLOGICAL RESEARCH

动物学研究 DONGWUXUE YANJIU

Bimonthly, Since 1980

中国精品科技期刊 中国科技核心期刊 动物学类核心期刊

Editor-in-Chief: Yong-Gang YAO

Executive Editor-in-Chief: Yun ZHANG

Editors: Su-Qing LIU Long NIE Andrew Willden

Edited by Editorial Office of Zoological Research

(Kunming Institute of Zoology, Chinese Academy of Sciences, 32 Jiaochang Donglu, Kunming, Yunnan, Post Code: 650223 Tel: +86 871 65199026 E-mail: zoores@mail.kiz.ac.cn

Sponsored by Kunming Institute of Zoology, Chinese Academy of Sciences; China Zoological Society©

Supervised by Chinese Academy of Sciences

Published by Science Press (16 Donghuangchenggen Beijie, Beijing 100717, China)

Printed by Kunming Xiaosong Plate Making & Printing Co, Ltd

Domestic distribution by Yunnan Post and all local post offices in China

International distribution by China International Book Trading Corporation (Guoji Shudian) P.O.BOX 399, Beijing 100044, China

Advertising Business License 广告经营许可证：滇工商广字66号

Domestic Postal Issue No.: 64-20

Price: 10.00 USD/60.00 CNY Post No.: BM358

ISSN 2095-8137