

Published in final edited form as:

Science. 2014 June 6; 344(6188): 1168–1173. doi:10.1126/science.1252806.

## The Sheep Genome Illuminates Biology of the Rumen and Lipid Metabolism

A full list of authors and affiliations appears at the end of the article.

# These authors contributed equally to this work.

### Abstract

Sheep (*Ovis aries*) are a major source of meat, milk and fiber in the form of wool, and represent a distinct class of animals that have a specialized digestive organ, the rumen, which carries out the initial digestion of plant material. We have developed and analyzed a high quality reference sheep genome and transcriptomes from 40 different tissues. We identified highly expressed genes encoding keratin cross-linking proteins associated with rumen evolution. We also identified genes involved in lipid metabolism that had been amplified and/or had altered tissue expression patterns. This may be in response to changes in the barrier lipids of the skin, an interaction between lipid metabolism and wool synthesis, and an increased role of volatile fatty acids in ruminants, compared to non-ruminant animals.

Sheep and goats are thought to be the first domesticated livestock species and thus integral to animal husbandry. Sheep are ruminants, digesting plant material in a four chambered stomach (1), the largest compartment is the rumen, which uses microbial flora to ferment the feed facilitating the conversion of ligno-cellulose rich plant materials, of low value in the human diet, to animal protein (2). The rumen is thought to have evolved around 35-40 million years ago (3) coinciding with the emergence of grasslands, in a cooler climate and atmosphere containing lower CO<sub>2</sub> than today (4, 5). Ruminants are now the dominant terrestrial herbivores. The rumen microbial flora also generate volatile fatty acids (VFAs)

\*Correspondence to: brian.dalrymple@csiro.au, wwang@mail.kiz.ac.cn, xuxun@genomics.cn, alan.archibald@roslin.ed.ac.uk, kworley@bcm.edu, noelle.cockett@usu.edu.

‡Present address: 16 Park Square, Port Melbourne, Vic 3207, Australia.

§Present address: Huntsman Cancer Institute, Salt Lake City, UT 84112, USA.

•Present address: Cobb-Vantress, Siloam Springs, AR USA.

\*Present address: School of Molecular and Biomedical Science, University of Adelaide, Adelaide, South Australia 5005, Australia.

●Present address: Genomic Strategies, PO Box 123 Lawson, NSW 2783, Australia.

||Present address: Department of Animal Science, Oklahoma State University, Stillwater, OK 74078, USA.

○Present address: to be confirmed

†Present address: Illumina Cambridge Ltd, Cambridge, UK.

The authors declare no competing financial interests.

The genome assemblies have been deposited in GenBank, Oar v1.0 (GCA\_000005525.1), and Oar v3.1 (GCA\_000298735.1) and in GigaDB, Oar v2.0 (<http://dx.doi.org/10.5524/100023>). The Ensembl annotation of the Oar v3.1 assembly is available at [http://www.ensembl.org/Ovis\\_aries/Info/Annotation#assembly](http://www.ensembl.org/Ovis_aries/Info/Annotation#assembly) and the gene builds are available from [ftp://ftp.ensembl.org/pub/release-74/bam/ovis\\_aries/genebuild/](ftp://ftp.ensembl.org/pub/release-74/bam/ovis_aries/genebuild/). The RNA-Seq datasets have been deposited in public databases; 83 samples from The Roslin Institute (ENA study accession PRJEB6169), seven tissues from the genome sequenced Texel ewe and Gansu alpine fine wool sheep skin (GenBank accession GSE56643), 3 blood samples (GenBank BioProject accession PRJNA245615). The MeDIP-seq raw reads from the Texel ewe liver have been deposited in GenBank (GSE56644). The BAC sequence assembly has been deposited in GenBank (KJ735098). The raw reads of the genome sequencing projects have been deposited in public nucleotide databases; Texel ewe (GenBank accession SRA059406), Texel ram (ENA study accession PRJEB6251, GenBank accession SRP015759), 0.5 fold coverage 454 sequencing of six animals (GenBank accessions SRP000982, SRP003883, SRP006794).

(6) requiring specialized energy and lipid metabolism in ruminants, and produce the greenhouse gas methane, which may be relevant to climate change (7). Another feature of sheep is wool, which has a significant proportion of its weight made up of lanolin, formed primarily from wax esters (8, 9). Thus, synthesis of wool may be linked to fatty acid metabolism. Given these unusual evolutionary traits, sheep provide a novel area for exploration of the genetic underpinnings of digestion and fatty acid metabolism.

We assembled the reference genome sequence of the sheep (Fig. 1) from two Texel individuals totaling ~150 fold sequence coverage (table S1) using linkage and radiation hybrid maps (tables S2, S3) to order and orientate the super-scaffolds (10). The final sheep genome assembly, Oar v3.1, has a contig N50 length of ~40 kb and a total assembled length of 2.61 Gb, with ~99% anchored onto the 26 autosomes and the X chromosome (table S4) (10). About 0.2% of each Texel's genome are heterozygous loci, i.e. SNPs (Single Nucleotide Polymorphisms) (Figs. 1C, D), a quarter of which are heterozygous in both animals. Due to selection for a beneficial muscle hypertrophy phenotype in the Texel breed (11) (OMIA 001426-9940) both sheep share a long run of homozygosity spanning the *MSTN* gene (Figs. 1C, D). The SNPs also enabled us to identify allele-specific gene expression (Fig 1E) (supplementary online text) (10).

We identified segmental duplications in sheep (Fig. 1A, table S5) (supplementary online material) (10) and compared the genome assemblies of sheep, goats and cattle (fig. S1) identifying 141 breakpoints between sheep and cattle (Fig. 1B, table S6) (supplementary online text). We compared the sequences of sheep, goat, cattle, yak, pig, camel, horse, dog, mouse, opossum and human proteins and identified 4,850 single-copy orthologous genes from which we constructed a phylogenetic tree (Figs. 2, S2) (10). The separation of sheep from goats and the diversification of the bovids occurred contemporaneously with the expansion of the C4 grasses in the late Neogene (4). RNASeq transcriptome data were generated from 94 tissue samples (from 40 tissues), including 83 from four additional Texel individuals (table S7) (supplementary online text) (10). A protein clustering analysis among the 11 mammalian species identified 321 expanded sub-families in the ruminant branch, of which 73 were ruminant specific (tables S8, S9) (10). We identified sheep genes exhibiting changes in copy number (e.g. lysozyme C-related proteins, prolactin-related proteins, pregnancy-associated glycoproteins. *RNASE1*, *ASIP*, *MOGAT2*, *MOGAT3*) and changes in tissue specificity of gene expression (e.g. *MOGAT2*, *MOGAT3*, *FABP9*) (Figs. 1, 2) (supplementary online text) (10).

The mammalian epidermal development complex (EDC) region contains up to 70 genes encoding proteins involved in keratinized epidermal structure development, including the rumen (1), skin and wool (12). The sheep EDC region (Figs. 1, 3) included several previously unidentified, or poorly annotated, genes in any mammalian genome (table S10). One such gene in the top 0.1% of all genes expressed in the rumen, but not expressed in the skin (Fig. 3A), is predicted to encode a large S100 fused-type protein (12). This protein has homology to trichohyalin (TCHH) (12), which is highly expressed in the skin, and we designated it Trichohyalin-like 2 (TCHHL2) (figs. S3, S4). Expressed sequence tag (EST) data indicate that *TCHHL2* is also expressed in the rumen of cattle (supplementary online text). Although not previously annotated, syntenically conserved orthologous genes to

*TCHHL2* were detected in many mammalian genomes, including a marsupial, the Tasmanian devil, and a monotreme, the platypus (fig. S3), but not in the chicken, suggesting that *TCHHL2* may be specific to mammals. All *TCHHL2* orthologs encode a protein containing up to 70 tandem copies of a highly variable 15 amino acid repeat, rich in arginine, glutamic acid, aspartic acid, and proline, which does not appear to be rumen specific (fig. S3). A short array of 7 copies of the 15 amino acid repeat unit has been duplicated in the common ancestor of ruminants (Figs. 3B, S4). S100 fused-type proteins are substrates for transglutaminase-mediated cross-linking to themselves, and to other proteins including keratins, during epidermal cornification and hair/wool development (12), suggesting *TCHHL2* may play a role in cross-linking the keratins at the rumen surface. *TCHHL2* was expressed in other sheep epidermal tissues, but >1000x lower than in rumen; thus, it may have a similar, but less extensive, role in these tissues.

*PRD-SPRR11* (13) is also in the EDC region and the top 0.1% of genes expressed in the rumen, but not in any other tissue sampled including skin (Fig. 3A). *PRD-SPRR11* is homologous to the *SPRR2* gene family (12, 14), but encodes a protein with a distinctive proline and histidine rich sequence which disrupts the glutamine rich amino-terminus present in *SPRR2* proteins, potentially affecting its transglutamination sites (Fig. 3C) (15). We identified four additional genes related to *PRD-SPRR11* in the sheep EDC region, two of which were also highly expressed in the rumen, but not in any other tissue sampled including the skin (Fig. 3A), and eight related genes in the cattle EDC region also expressed in the rumen, but none in non-ruminants (Fig. 3C) (supplementary online text). Thus it appears that the ruminant specific *PRD-SPRR11* family genes, resulting from the amplification and sequence divergence of an *SPRR2* gene, have gained a new expression pattern, altered amino-terminal sequence, and modified function during rumen evolution. By analogy with *SPRR2* (12, 16) the *PRD-SPRR11* family proteins are predicted to be major structural proteins and may function in the cornification of the keratin-rich surface of the rumen.

The primary role of the skin and wool, an important economic product of sheep, is to form a barrier between the organism and the external environment, reducing water and heat loss and pathogen entry (17). The sheep EDC gene, LOC101122906, which we designated *LCE7A*, represents a previously unrecognized subfamily of the late cornified envelope (LCE) genes (12, 18) (fig. S5). We identified *LCE7A* coding sequence in the genomes of most mammals, although it has not been previously annotated (fig. S5). *LCE7A* is expressed in sheep (Fig. 3A, table S10), cattle and goat skin (supplementary online text), but not in the rumen, or any other tissue examined (Fig. 3A). In situ hybridization showed *LCE7A* expression in Merino wool follicles including the inner root sheath (fig. S6). *LCE7A* contains the transglutaminase target site present in most LCE proteins (fig. S5) and is likely to be a substrate for transglutaminase mediated cross-linking of proteins in the epidermis, inner root sheath or wool (18). *LCE7A* also appears to be under positive selection in the sheep lineage, with a sheep vs. cattle *ka/ks* ratio of 2.5 ( $P < 0.005$ ), possibly reflecting an association of *LCE7A* with wool development.

Wool grease (lanolin), secreted from the sebaceous glands attached to the wool follicles, constitutes 10-25% of the wool weight (9). The wool follicles are located in the dermal layer

between the surface keratinocytes, which synthesize surface lipids (19), and the subcutaneous adipose tissue. We identified the genes encoding lipid metabolic enzymes expressed in the skin (table S11) and positioned them on known and putative lipid metabolic pathways likely to be involved in the storage and mobilization of long chain fatty acid components of the sebum and epidermal lipids (Fig. 4A). Unexpectedly, the skin transcriptome revealed high expression of *MOGAT2* and *MOGAT3*, members of the acylglycerol O-acyltransferase (*DGAT2/MOGAT*) gene family which are involved in the synthesis of diacylglyceride (DAG) and triacylglyceride (TAG) from monoacylglyceride (MAG) (Fig. 4A). In humans, *MOGAT3* is an essential and rate-limiting step for the absorption of dietary fat in the small intestine (20), and an important liver enzyme (21). *MOGAT2* and *MOGAT3* are single copy genes in almost all mammals with available data. However, in ruminants both genes have undergone tandem gene expansions, indicative of evolving functionality. (Figs. 4B, S7, S8). *MOGAT2* has more than five tandemly duplicated copies in sheep with the first copy expressed in the duodenum and the last copy expressed in the skin, with no expression of any copy detected in the liver (fig. S7). Three nearly identical *MOGAT3* copies were highly expressed in sheep skin, and at a much lower level in white adipose and omentum (Fig. 4B). In contrast to humans, we detected no expression of functional *MOGAT3* in sheep duodenum or liver. The skin had two *MOGAT3* splice variants (fig. S9): the most common transcript encodes a predicted protein orthologous to the typical mammalian *MOGAT3*; the second contains a predicted alternate start codon and amino-terminal sequence that is missing the probable membrane anchor (fig. S10) (supplementary online text), predicted to be uncoupled from the endoplasmic reticulum membrane bound TAG synthesis pathway (Fig. 4A).

The presence of *MOGAT2* and *MOGAT3* in sheep skin indicates that there may be an alternative pathway for DAG synthesis, either from recycling MAG generated from the mobilization of TAG stored within a cell to generate fatty acids for incorporation into other products, or from external sources of MAG (Fig. 4A). The *MOGAT* pathway does not generate glycerol, which requires phosphorylation in the liver prior to reuse for TAG synthesis in the skin via the glycerol-3-phosphate (G3P) pathway (Fig. 4A), potentially increasing the efficiency of recycling of the glycerol backbone in sheep skin. The *MOGAT* pathway also bypasses 1-acyl-lysophosphatidic acid (LPA) and phosphatidic acid (PA) synthesis (Fig. 4A). Skin produces a lipase (*LIPH*) to cleave PA into 2-acyl LPA, which has a role in controlling hair follicle development (22). *LIPH* mutations in several mammalian species result in wool-like hair due to changes in follicle shape (23). Thus the *MOGAT* pathway in sheep skin may also reduce the coupling between TAG and PA synthesis, skin barrier lipid synthesis and follicle development signaling, facilitating wool production.

In ruminants, the liver is primarily a gluconeogenic organ utilizing propionate (a VFA) as the source of carbon. It contributes little to the synthesis of lipids, or to the uptake of lipids from circulation, and is inefficient at exporting TAG (24, 25). The apparent loss of *MOGAT3* expression in the intestines and both *MOGAT2* and *MOGAT3* in the liver may reflect the greater importance of VFAs and reduced importance of the liver in long chain fatty acid metabolism in ruminants compared to non-ruminants.

In conclusion, we identified major genomic signatures associated with interactions between diet, the digestive system and metabolism in ruminants. These include two extensions of their biochemical capabilities that have been extensively exploited by humans: the production of wool by sheep and the evolution of an organ that houses a diverse community of microorganisms which enable efficient digestion of plants.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Authors

Yu Jiang<sup>#1,2,3</sup>, Min Xie<sup>#4</sup>, Wenbin Chen<sup>#4</sup>, Richard Talbot<sup>5</sup>, Jillian F. Maddox<sup>6,‡</sup>, Thomas Faraut<sup>7</sup>, Chunhua Wu<sup>8,§</sup>, Donna M. Muzny<sup>9</sup>, Yuxiang Li<sup>4</sup>, Wenguang Zhang<sup>1,10,11</sup>, Jo-Ann Stanton<sup>12</sup>, Rudiger Brauning<sup>13</sup>, Wesley C. Barris<sup>2,■</sup>, Thibaut Hourlier<sup>14,21</sup>, Bronwen L. Aken<sup>14,21</sup>, Stephen M.J. Searle<sup>14</sup>, David L. Adelson<sup>2,▲</sup>, Chao Bian<sup>4</sup>, Graham R. Cam<sup>2,●</sup>, Yulin Chen<sup>3</sup>, Shifeng Cheng<sup>4</sup>, Udaya DeSilva<sup>2,||</sup>, Karen Diken<sup>15</sup>, Yang Dong<sup>1</sup>, Guangyi Fan<sup>4</sup>, Ian R. Franklin<sup>2,○</sup>, Shaoyin Fu<sup>10</sup>, Rui Guan<sup>4</sup>, Margaret A. Highland<sup>16,17</sup>, Michael E. Holder<sup>9</sup>, Guodong Huang<sup>4</sup>, Aaron B. Ingham<sup>2</sup>, Shalini N. Jhangiani<sup>9</sup>, Divya Kalra<sup>9</sup>, Christie L. Kovar<sup>9</sup>, Sandra L. Lee<sup>9</sup>, Weiqing Liu<sup>4</sup>, Xin Liu<sup>4</sup>, Changxin Lu<sup>4</sup>, Tian Lv<sup>4</sup>, Tittu Mathew<sup>9</sup>, Sean McWilliam<sup>2</sup>, Moira Menzies<sup>2</sup>, Shengkai Pan<sup>4</sup>, David Robelin<sup>7</sup>, Bertrand Servin<sup>7</sup>, David Townley<sup>2,†</sup>, Wenliang Wang<sup>4</sup>, Bin Wei<sup>4,18</sup>, Stephen N. White<sup>16,17</sup>, Xinhua Yang<sup>4</sup>, Chen Ye<sup>4</sup>, Yaojing Yue<sup>19</sup>, Peng Zeng<sup>4</sup>, Qing Zhou<sup>4</sup>, Jacob B. Hansen<sup>15</sup>, Karsten Kristensen<sup>20</sup>, Richard A. Gibbs<sup>9</sup>, Paul Flicek<sup>21</sup>, Christopher C. Warkup<sup>22</sup>, Huw E. Jones<sup>22</sup>, V. Hutton Oddy<sup>23</sup>, Frank W. Nicholas<sup>24</sup>, John C. McEwan<sup>13</sup>, James Kijas<sup>2</sup>, Jun Wang<sup>4,20,25,26</sup>, Kim C. Worley<sup>9,\*</sup>, Alan L. Archibald<sup>27,\*</sup>, Noelle Cockett<sup>8,\*</sup>, Xun Xu<sup>4,\*</sup>, Wen Wang<sup>1,\*</sup>, and Brian P. Dalrymple<sup>2,\*</sup>

## Affiliations

<sup>1</sup>State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China. <sup>2</sup>CSIRO Animal Food and Health Sciences, St Lucia, QLD 4067, Australia. <sup>3</sup>College of Animal Science and Technology, Northwest A&F University, Yangling 712100, China. <sup>4</sup>BGI-Shenzhen, Shenzhen 518083, China. <sup>5</sup>Edinburgh Genomics, University of Edinburgh, Easter Bush, Midlothian EH 25 9RG, UK. <sup>6</sup>Department of Veterinary Science, University of Melbourne, Victoria 3010, Australia. <sup>7</sup>INRA, Laboratoire de Génétique Cellulaire, UMR 444, Castanet-Tolosan F-31326, France. <sup>8</sup>Utah State University, Logan, UT 84322-1435-1435, USA. <sup>9</sup>Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA <sup>10</sup>Inner Mongolia Agricultural University, Hohhot 010018, China. <sup>11</sup>Institute of ATCG, Nei Mongol Bio-Information, Hohhot, China. <sup>12</sup>Department of Anatomy, University of Otago, Dunedin 9054, New Zealand. <sup>13</sup>AgResearch, Invermay Agricultural Centre, Mosgiel 9053, New Zealand. <sup>14</sup>Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK. <sup>15</sup>Department of Biology, University of Copenhagen, DK-2100 Copenhagen Ø, Denmark. <sup>16</sup>USDA-ARS Animal Disease

Research Unit, Pullman, WA 99164 USA. <sup>17</sup>Department of Veterinary Microbiology & Pathology, Washington State University, Pullman, WA 99164 USA. <sup>18</sup>Maize Research Institute, Sichuan Agricultural University, Chengdu 611130, China. <sup>19</sup>Lanzhou Institute of Husbandry and Pharmaceutical Science, Lanzhou, 730050, China. <sup>20</sup>Department of Biology, University of Copenhagen, DK-2200 Copenhagen N, Denmark. <sup>21</sup>European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, United Kingdom. <sup>22</sup>Biosciences KTN, The Roslin Institute, Easter Bush, Midlothian, EH25 9RG, UK. <sup>23</sup>School of Environmental and Rural Science, University of New England, Armidale, NSW 2351, Australia. <sup>24</sup>Faculty of Veterinary Science, University of Sydney, NSW 2006, Australia. <sup>25</sup>Princess Al Jawhara Center of Excellence in the Research of Hereditary Disorders, King Abdulaziz University, Jeddah 21589, Saudi Arabia. <sup>26</sup>Macau University of Science and Technology, Macau 999078, China. <sup>27</sup>The Roslin Institute and R(D)SVS, University of Edinburgh, Easter Bush, Midlothian EH 25 9RG, UK.

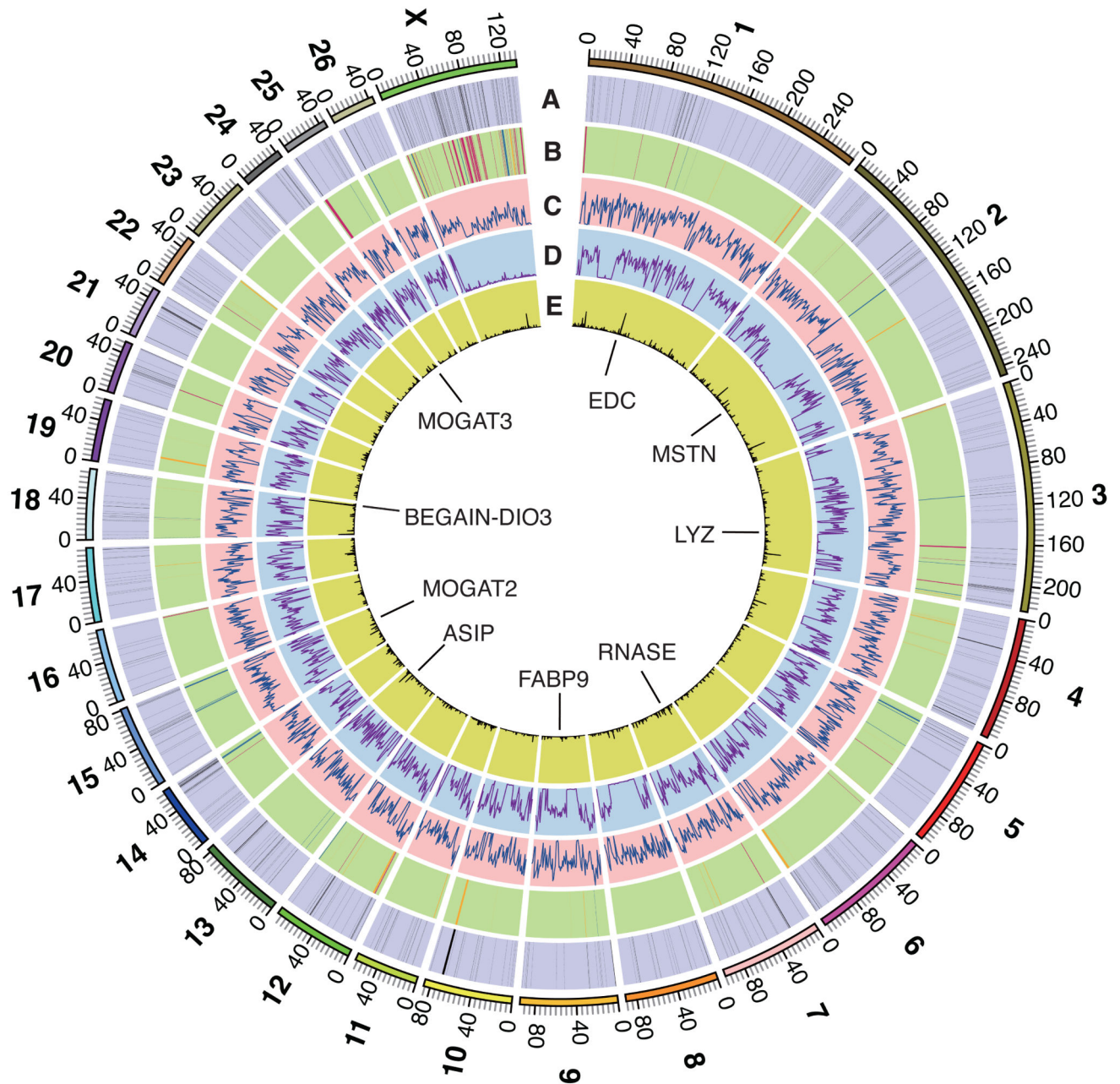
## Acknowledgments

The International Sheep Genomics Sequencing Consortium is grateful to the following for funding support for the sheep genome sequencing project, two 973 Programs (No. 2013CB835200, No. 2007CB815700) to Kunming Institute of Zoology, China; BGI-Shenzhen (ZYC200903240077A, ZYC200903240078A); China National GeneBank-Shenzhen for support for the storage of samples and data; Inner Mongolia Agricultural University (30960246,31260538); The Roslin Institute, University of Edinburgh and Biotechnology and Biological Sciences Research Council, U.K. (BBSRC): BB/1025360/1, BB/1025328/1, Institute Strategic Programme and National Capability Grants; The Roslin Foundation; The Scottish Government, U.K.; Defra/HEFC/SHEFC Veterinary Training and Research Initiative, U.K.; USDA-ARS, USA, USDA-NRICGP, USA (grant numbers 2008-03923 and 2009-03305); Wellcome Trust (grant numbers WT095908 and WT098051), BBSRC (grant numbers BB/1025506/1) and EMBL; USDA-NRSP-8, USA; USDA-ARS grant 5348-32000-031-00D; Meat and Livestock Australia and Australian Wool Innovation Limited through sheepGENOMICS, Australia; Australian Government International Science Linkages Grant (CG090143), Australia; University of Sydney, Australia; CSIRO, Australia; AgResearch, NZ, Beef + Lamb NZ through Ovita, New Zealand; INRA and ANR project SheepSNPQTL, France; European Union through the Seventh Framework Programme Quantomics (KBBE222664) and 3SR (KBBE245140) projects; the Ole Rømer grant from Danish Natural Science Research Council, BGI-Shenzhen, China; Australian Department of Agriculture Food and Fisheries, Filling the Research Gap project, “Host control of methane emissions”. We thank Brad Freking (USDA-ARS-USMARC) for provision of Texel ram tissue samples for DNA extraction and sequencing. We thank the sequencing teams and other contributors, full details in the acknowledgements section of the supplementary material. We thank SheepGENOMICS and Utah State University for access to the genotyping data for the SheepGENOMICS and LSU flocks respectively.

## References and Notes

- Hofmann RR. *Oecologia*. 1989; 78:443–457.
- Wolin MJ. *Science*. 1981; 213:1463–1468. [PubMed: 7280665]
- Hackmann TJ, Spain JN. *J. Dairy Sci.* 2010; 93:1320–1334. [PubMed: 20338409]
- Stromberg CAE. *Annu. Rev. Earth Planet. Sci.* 2011; 39:517–544.
- Edwards EJ, et al. *Science*. 2010; 328:587–591. [PubMed: 20431008]
- Bergman EN. *Physiol. Rev.* 1990; 70:567–590. [PubMed: 2181501]
- Johnson KA, Johnson DE. *J. Anim. Sci.* 1995; 73:2483–2492. [PubMed: 8567486]
- Stewart, ME. chap. 43. In: Bereiter-Hahn, J.; Matoltsy, AG.; Richards, KS., editors. *Biology of the integument*. Vol. 2. Springer; Berlin Heidelberg: 1986. p. 824-832.
- Schlossman M, McCarthy J. *J. Am. Oil Chem. Soc.* 1978; 55:447–450.
- Material and methods are available as supplementary material on *Science* Online

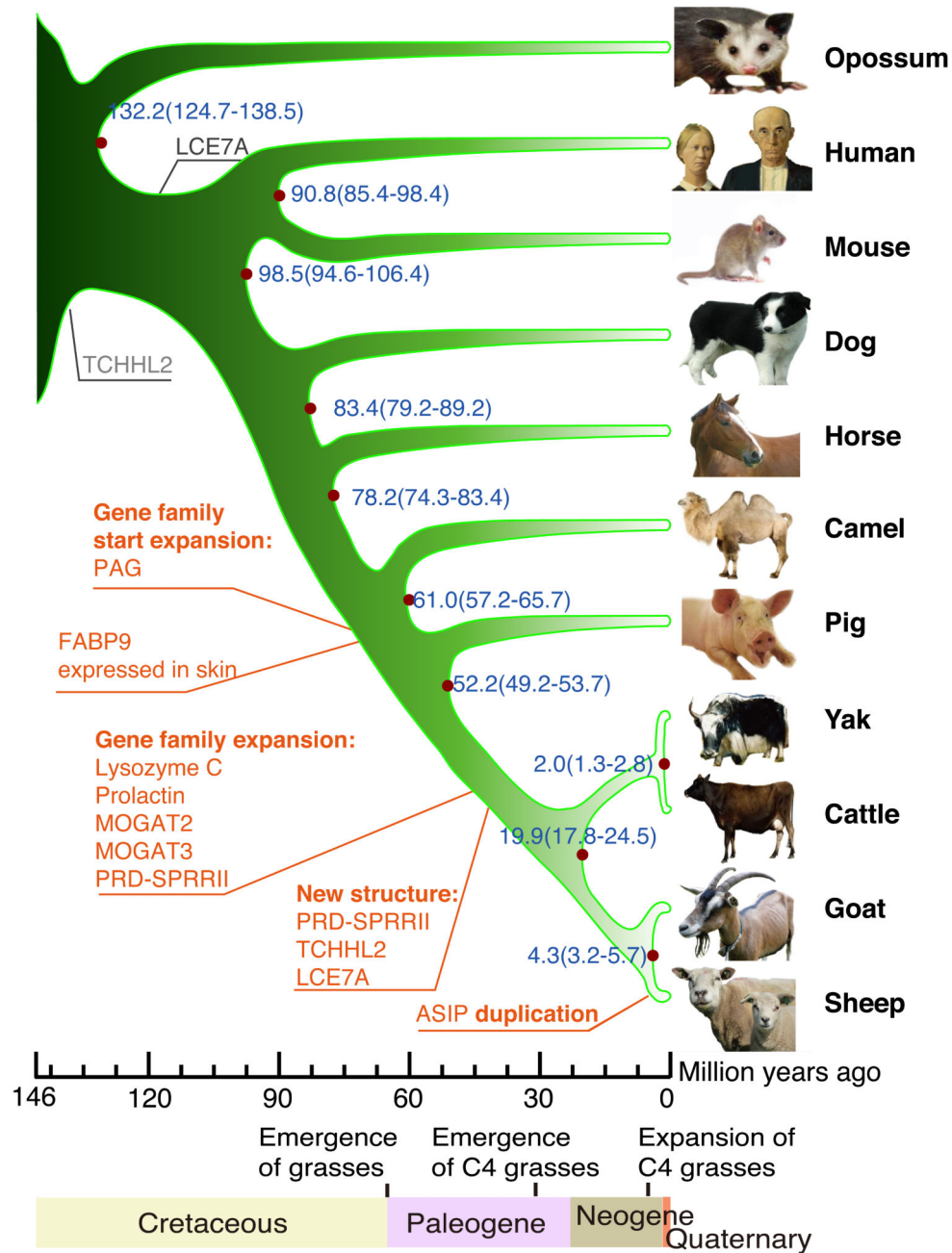
11. Clop A, et al. *Nat. Genet.* 2006; 38:813–818. [PubMed: 16751773]
12. Kypriotou M, Huber M, Hohl D. *Exp. Dermatol.* 2012; 21:643–649. [PubMed: 22507538]
13. Wang L, Baldwin R. L. t. Jesse BW. *Biochem. J.* 1996; 317(Pt 1):225–233. [PubMed: 8694768]
14. Song H, et al. *Genomics.* 1999; 55:28–42. [PubMed: 9888996]
15. Deng J, Pan R, Wu R. *J. Biol. Chem.* 2000; 275:5739–5747. [PubMed: 10681560]
16. Steinert P, Candi E, Kartasova T, Marekov L. J. *Struct. Biol.* 1998; 122:76–85. [PubMed: 9724607]
17. Feingold K. J. *Lipid Res.* 2007; 48:2531–2546. [PubMed: 17872588]
18. Marshall D, Hardman M, Nield K, Byrne C. *Proc. Natl. Acad. Sci. U.S.A.* 2001; 98:13031–13036. [PubMed: 11698679]
19. Radner F, Grond S, Haemmerle G, Laser A, Zechner R. *Dermatoendocrinol.* 2011; 3:77–83. [PubMed: 21695016]
20. Cheng D, et al. *J. Biol. Chem.* 2003; 278:13611–13614. [PubMed: 12618427]
21. Hall AM, et al. *J. Lipid. Res.* 2012; 53:990–999. [PubMed: 22394502]
22. Kazantseva A, et al. *Science.* 2006; 314:982–985. [PubMed: 17095700]
23. Inoue A, et al. *EMBO J.* 2011; 30:4248–4260. [PubMed: 21857648]
24. Bobe G, Young J, Beitz D. *J. Dairy Sci.* 2004; 87:3105–3124. [PubMed: 15377589]
25. Ingle D, Bauman D, Garrigus U. *J. Nutr.* 1972; 102:617–623. [PubMed: 5022195]



**Fig. 1. The genome of sheep**

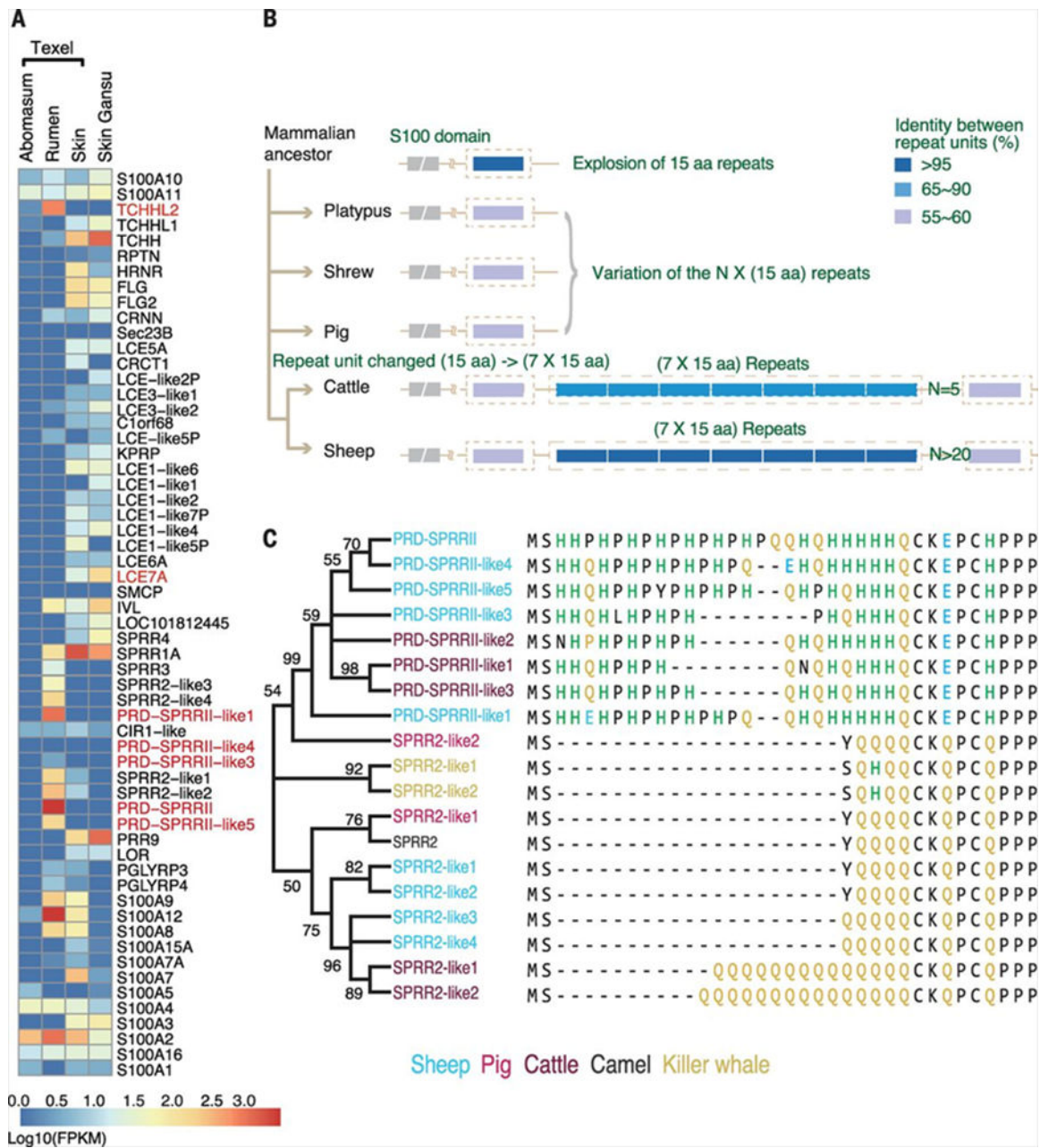
(A) 1,097 segmental duplications (length >5 kb) in the two Texel sheep (10). (B) Sheep vs. cattle chromosome break points (10), gaps in Oar v3.1 are red, probable misassemblies in UMD 3.1 are yellow and probable true structural differences are blue. (C) Distribution of SNPs in the Texel ewe in 1 Mb non-overlapping windows, range of values 1 to 4954. (D) Distribution of SNPs in the Texel ram in 1 Mb non-overlapping windows, range of values 3 to 5676. (E) Distribution of mono-allelically expressed SNPs in 500 kb sliding windows, range of values 0 to 42 (10). The scale is in Mb. Positions of loci discussed in the text are indicated.



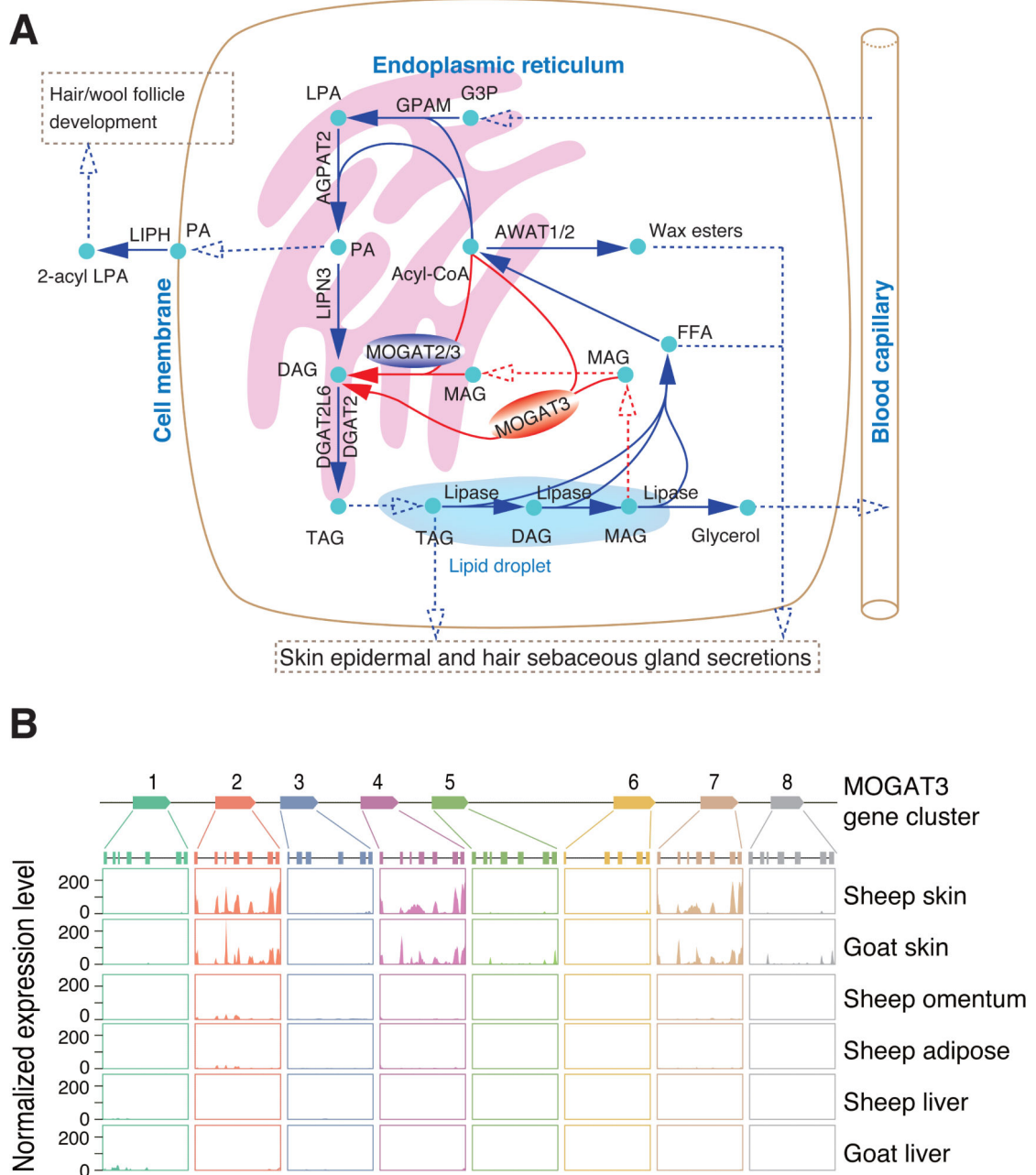


**Fig. 2. Sheep relative to human and livestock evolution**

A phylogenetic tree generated using single-copy orthologous genes (10). The origins (black) and amplifications (red) of genes discussed in the text are marked. The scale is in millions of years ago (Mya). Blue numbers on the nodes are the divergence time from present (Mya) and its confidence interval. Dates for major events in the evolution of grasses are from (4) and (5).



**Fig. 3. Epidermal Differentiation Complex (EDC) gene expression and sequence organization**  
**(A)** Expression levels of the genes in the EDC region in selected sheep tissues. The abomasum is equivalent to the stomach of a non-ruminant. Genes are ordered according to their genomic order, with pseudogenes omitted. Genes discussed in the text are highlighted.  
**(B)** Proposed evolution patterns of the core 15 amino acid repeat of predicted TCHHL2 proteins.  
**(C)** Maximum likelihood phylogenetic tree of PRD-SPRRII and SPRR2 proteins. For clarity three representative cattle sequences shown, Bootstrap values 50% (500 replicates) are shown.



**Fig. 4. Proposed sheep skin lipid metabolic pathways and the expression of the *MOGAT3* region**  
**(A)** Proposed sheep skin lipid metabolic pathways. The proposed non-membrane form of *MOGAT3* is red; the membrane form of *MOGAT2/3* is blue. The glycerol-3-phosphate and *MOGAT* pathways to DAG are in blue and red arrows respectively. Dashed lines represent transport or diffusion of products. **(B)** Expression of eight *MOGAT3* genes in sheep skin (Gansu fine wool), and sheep omentum, white adipose and liver (Texel), goat skin (Shanbei Cashmere), and goat liver (Yunling Black).