

1 **Tracing the origin of a new organ by inferring the genetic basis of rumen evolution**

2 Xiangyu Pan<sup>1†</sup>, Yu Wang<sup>1†</sup>, Zongjun Li<sup>1†</sup>, Xianqing Chen<sup>2†</sup>, Rasmus Heller<sup>3†</sup>, Nini Wang<sup>1†</sup>, Chen Zhao<sup>1</sup>,  
3 Yudong Cai<sup>1</sup>, Han Xu<sup>1</sup>, Songhai Li<sup>4</sup>, Ming Li<sup>1</sup>, Cunyuan Li<sup>5</sup>, Shengwei Hu<sup>5</sup>, Hui Li<sup>1</sup>, Kun Wang<sup>2</sup>, Lei  
4 Chen<sup>2</sup>, Bin Wei<sup>1</sup>, Zhuqing Zheng<sup>1</sup>, Weiwei Fu<sup>1</sup>, Yue Yang<sup>2</sup>, Tingting Zhang<sup>1</sup>, Zhuoting Hou<sup>2</sup>, Yueyang  
5 Yan<sup>1</sup>, Xiaoyang Lv<sup>6</sup>, Wei Sun<sup>6,7</sup>, Xinyu Li<sup>8</sup>, Shisheng Huang<sup>9</sup>, Lixiang Liu<sup>10</sup>, Shengyong Mao<sup>10</sup>, Wenqing  
6 Liu<sup>11</sup>, Jinlian Hua<sup>11</sup>, Zhipeng Li<sup>12</sup>, Guojie Zhang<sup>13,14,15,16</sup>, Yulin Chen<sup>1</sup>, Xihong Wang<sup>1</sup>, Qiang Qiu<sup>2,17</sup>, Brian  
7 P. Dalrymple<sup>18</sup>, Wen Wang<sup>2,15,16\*</sup>, Yu Jiang<sup>1\*</sup>

8 1 Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology,  
9 Northwest A&F University, Yangling 712100, China

10 2 School of Ecology and Environment, Northwestern Polytechnical University, Xi'an 710072, China

11 3 Section for Computational and RNA Biology, Department of Biology, University of Copenhagen, DK-2100 Copenhagen, Denmark

12 4 Marine Mammal and Marine Bioacoustics Laboratory, Institute of Deep-sea Science and Engineering, Chinese Academy of  
13 Sciences, Sanya 572000, China.

14 5 College of Life Sciences, Shihezi University, Shihezi, Xinjiang 832003, China

15 6 College of Animal Science and Technology, Yangzhou University, Yangzhou 225009, China;

16 7 Joint International Research Laboratory of Agriculture and Agri-Product Safety of Ministry of Education of China, Yangzhou  
17 University, Yangzhou 225009, China

18 8 Key Laboratory for Major Obstetric Diseases of Guangdong Province, The Third Affiliated Hospital of Guangzhou Medical  
19 University, Guangzhou 510150, China

20 9 School of Life Science and Technology, Shanghai Tech University, Shanghai 201210, China

21 10 College of Animal Science and Technology, Nanjing Agricultural University, Nanjing 210095, China

22 11 College of Veterinary Medicine, Shaanxi Centre of Stem Cells Engineering & Technology, Northwest A&F University, Yangling,  
23 Shaanxi 712100, China

24 12 Department of Special Economic Animal Nutrition and Feed Science, Institute of Special Animal and Plant Sciences, Chinese  
25 Academy of Agricultural Sciences, Changchun 130112, China

26 13 Section for Ecology and Evolution, Department of Biology, University of Copenhagen, DK-2100 Copenhagen, Denmark

27 14 China National GeneBank, BGI-Shenzhen, Shenzhen 518083, China

28 15 State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming  
29 650223, China

30 16 Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China

31 17 State Key Laboratory of Grassland Agro-Ecosystem, College of Life Sciences, Lanzhou University, Lanzhou 730000, China

32 18 School of Animal Biology and Institute of Agriculture, The University of Western Australia, 35 Stirling Highway, Crawley WA  
33 6009, Australia

34 <sup>†</sup>These authors contributed equally to this work.

35 \*Corresponding author. E-mail: [yu.jiang@nwafu.edu.cn](mailto:yu.jiang@nwafu.edu.cn) (Y.J.), [wwang@mail.kiz.ac.cn](mailto:wwang@mail.kiz.ac.cn) (W.W.);

36 **Abstract**

37 The rumen is the hallmark organ of ruminants and hosts a diverse ecosystem of  
38 microorganisms that facilitates efficient digestion of plant fibers. We used 897  
39 transcriptomes from three Cetartiodactyla lineages: ruminants, camels and cetaceans,  
40 as well as data from ruminant comparative genomics and functional assays to explore  
41 the genetic basis of rumen origin and evolution. Comparative analyses reveal that the  
42 rumen and the first-chamber stomachs of camels and cetaceans shared a common  
43 tissue origin from the esophagus. The rumen recruited genes from other tissues/organs  
44 and up-regulated many esophagus genes to acquire functional innovations involving  
45 epithelium absorption, improvement of the ketone body metabolism and regulation of  
46 microbial community. These innovations involve such genetic changes as  
47 ruminant-specific conserved elements, newly evolved genes and positively selected  
48 genes. Our *in vitro* experiments validate the functions of one enhancer, one  
49 positively selected gene and two newly evolved antibacterial genes. Our study  
50 provides novel insights into the origin and evolution of a complex organ.

51 Evolutionary biology has a long history of trying to understand how complex organs  
52 evolve<sup>1</sup>. The origin of some notable organs has been central to animal evolution, e.g.  
53 the eyes of animals<sup>2,3</sup>, electric organs of fishes<sup>4</sup>, mammalian placenta<sup>5,6</sup> and ruminant  
54 headgear<sup>7</sup>. Another remarkable organ innovation found in mammals are the  
55 multi-chambered stomachs found in the Cetartiodactyla lineages, including Tylopoda  
56 (e.g. camels), Tayassuidae (e.g. peccaries), Hippopotamidae (e.g. hippos), Cetacea  
57 (e.g. whales) and Ruminantia (**Fig. 1**). Among these, ruminants have the most complex  
58 digestive system in herbivores, allowing efficient uptake of nutrients from plant  
59 material by providing a microbial fermentation ecosystem in the highly specialized  
60 rumen<sup>8</sup>. Camels (Tylopoda) have three-chambered stomachs and are also sometimes  
61 called "pseudo-ruminants" due to their similar ruminating behavior and microbial  
62 fermentation taking place in their first-chamber (FC) stomach<sup>9</sup>. The whales (Cetacea)  
63 form the sister group of the Ruminantia<sup>10</sup>, however the FC of their four-chambered  
64 stomach is mainly used as a temporary storage chamber for ingested food and for  
65 mechanical grinding of food items<sup>11</sup>. With the rumen, ruminants obtained a unique  
66 evolutionary advantage through superior utilization of short chain fatty acids (SCFAs)  
67 from microbial fermentation, which significantly promoted the expansion and  
68 diversification of ruminant taxa<sup>12</sup>. The evolutionary innovation of the rumen is  
69 therefore interesting not only in its functional complexity and uniqueness, but also  
70 because it has greatly benefited humans by providing high-quality nutrition in the shape  
71 of highly productive ruminant livestock species<sup>13,14</sup>.

72 The anatomical predecessor from which the rumen evolved has been proposed to

73 be the esophagus<sup>15</sup>, yet the two organs are highly divergent in morphology and  
74 physiology. The stratified squamous epithelium of the esophagus is smooth and  
75 non-keratinized, and mainly serves a barrier function, but in contrast the rumen  
76 stratified squamous epithelium is keratinized and lined with papillae, which facilitates  
77 nutrient uptake and antibacterial peptide production<sup>16,17</sup>. These features allow the  
78 absorption of SCFAs and sustain the homeostasis of microorganisms. The origin and  
79 evolution of new organs involve structural and functional innovations that were  
80 proposed to be driven by several types of genetic reprogramming: recruitment of  
81 genes usually expressed in other organs, transformation of regulatory elements such  
82 as promoters and enhancers, mutations in protein-coding genes and  
83 post-transcriptional mechanisms<sup>1,5</sup>. Given the substantial structural and physiological  
84 changes involved in the transition from esophagus to rumen, significant genetic  
85 reprogramming must have occurred during the process.

86 Usually, it is challenging to obtain detailed insights into the genetic  
87 reprogramming associated with organ evolution due to the rarity of such occurrences  
88 and the lack of intermediate evolutionary states<sup>5</sup>. However, in the case of the rumen,  
89 we can take advantage of two important points allowing “triangulation” of the  
90 changes leading to the rumen: the availability of synapomorphic stomach chambers in  
91 Cetartiodactyla and the likely ancestral relation between the esophagus and the rumen.  
92 Here, we conducted a comprehensive comparison using 897 transcriptomes of  
93 different tissues from three Cetartiodactyla lineages and multiple genomes to  
94 investigate the genetic basis of gene programming evolution and functional

95 innovations in rumen, together with validation of some cases using *in vitro*  
96 experiments.

97 **Results**

98 **Gene expression features of the rumen**

99 We sequenced transcriptomes of 33 samples across 14 adult tissues from Bactrian  
100 camels, eight adult tissues from one species in Mysticeti (Bryde's whale) and one  
101 species in Odontoceti (Indo-Pacific Finless Porpoise) from Cetacea, 852 samples (210  
102 sequenced in this study and 642 published in previous studies<sup>7,18,19</sup>) from 50 tissues of  
103 two representative ruminants (sheep and roe deer) within Ruminantia (**Supplementary**  
104 **Table 1**). The global gene expression patterns of all the FC stomachs are consistently  
105 most similar to the esophagus in all species (**Fig. 2a, Fig. S1**). To investigate the  
106 specifically expressed genes in the three types of FC stomachs, we defined those that  
107 the rank of expression is less than or equal to a E50 index threshold with type I error  
108 less than 0.05 (**Supplementary Note**) in the FC stomachs of ruminants, camels, and  
109 cetaceans compared to other conspecific tissues/organs. We identified 655, 593, and  
110 375 such specifically expressed genes in the FC stomachs of ruminants, camels, and  
111 cetaceans, respectively (**Supplementary Table 2-4; Supplementary Note**).

112 *Comparisons of gene expression profiles between rumen and the first-chamber stomach  
113 of camels and cetaceans*

114 Among these FC-specific genes, the three FC stomachs shared 18 genes which are  
115 co-expressed in the esophagus in all species (**Supplementary Table 5**). The 18 genes  
116 were significantly enriched in keratinocyte differentiation (**Supplementary Table 6**,  
117 Fisher's exact test, adjusted  $P$  value =  $9.85 \times 10^{-3}$ ). This is consistent with the fact that  
118 the FC stomachs all share a basic stratified squamous epithelium with the

119 esophagus<sup>20-22</sup>, which is markedly different from other stomach chambers (e.g. the  
120 abomasum of the ruminants, the third-chamber stomachs of camels and cetaceans).  
121 Notably, *PAX9*<sup>23</sup>, a known key transcription factor during esophagus differentiation, is  
122 highly expressed in all three FC stomachs and may play a role in the origin of the FC  
123 stomachs from their anatomic origin (**Supplementary Table 5**). Our results therefore  
124 indicate that the FC stomachs in Cetartiodactyla share a common developmental origin  
125 from the esophagus, and that changes in epidermis development may be an ancestral  
126 feature in this proto-rumen.

127 Despite the shared features of epithelial histology found in all Cetartiodactyla FC  
128 stomachs, the rumen also has a series of unique structural and functional innovations.  
129 Among the 655 rumen specifically expressed genes, we identified 448 up-regulated and  
130 79 down-regulated genes when compared to the FC stomachs of camels (**Fig. 2b**;  
131 **Supplementary Table 7**), and 563 up-regulated and 29 down-regulated genes when  
132 compared to the FC stomachs of cetaceans (**Fig. 2b**; **Supplementary Table 8**;  
133 **Supplementary Note**). Among these, the majority (427, 65.2%) are up-regulated in  
134 rumen relative to both the FC stomach of camels and cetaceans (**Fig. 2b**;  
135 **Supplementary Table 9**). These exclusively rumen-specific (i.e., not specifically  
136 expressed in other FC stomachs) genes are significantly associated with the synthesis  
137 and degradation of ketone bodies (Fisher's exact test, adjusted *P* value =  $1.21 \times 10^{-3}$ )  
138 (**Fig. 2c**; **Supplementary Table 10**). Unlike monogastric animals, in which  
139 ketogenesis mainly occurs in the liver and the intestinal tract<sup>24,25</sup>, the rumen is the main  
140 site of ketogenesis in adult ruminants, and the occurrence of ketogenesis is regarded as

141 a diagnostic feature of rumen maturity<sup>26</sup>. In addition to ketogenesis genes , seven genes  
142 from the KEGG pathway *Staphylococcus aureus* infection were also highly expressed  
143 in the rumen compared to the FC stomachs of camels and cetaceans (Fisher's exact test  
144 for KEGG pathway enrichment, adjusted  $P$  value =  $1.35 \times 10^{-2}$ ) (**Supplementary Table**  
145 **10**). These results indicate that improved ketone body metabolism and microbial  
146 regulation were important features in the evolution of the rumen from a proto-rumen  
147 origin shared with other Cetartiodactyls.

148 *Gene recruitment by the rumen*

149 Among the 655 rumen specifically expressed genes, the rumen co-expressed 96  
150 (14.7%) genes with the esophagus (**Fig. 2d; Supplementary Table 2**). The 96 genes  
151 were enriched in the cornified envelope (adjusted  $P = 4.11 \times 10^{-14}$ ) and epidermal cell  
152 differentiation processes (adjusted  $P = 3.77 \times 10^{-25}$ ) (**Supplementary Table 11**).  
153 Meanwhile, we also found that the rumen recruited genes from a range of other tissues  
154 and biological pathways (**Fig. 2d**), e.g. keratinocyte differentiation (**Supplementary**  
155 **Table 12**, 88 genes co-expressed with keratinization-associated tissues), urea cycle  
156 (**Supplementary Table 13**, 24 genes co-expressed with liver), monocarboxylic acid  
157 transport (**Supplementary Table 14**, 61 genes co-expressed with intestine), skeletal  
158 muscle contraction (**Supplementary Table 15**, 23 genes co-expressed with muscle),  
159 urea transport (**Supplementary Table 16**, 19 genes co-expressed with kidney) and  
160 saliva secretion (**Supplementary Table 17**, 10 genes co-expressed with salivary  
161 gland). These pathways are all strongly associated with known rumen functions. For  
162 instance, enhanced urea recycling is an important characteristic of the rumen leading to

163 increased nitrogen utilization for ruminants<sup>27</sup>. Collectively, these results suggest that  
164 the rumen—in addition to up-regulating genes expressed in the esophagus—recruited  
165 genes from different tissues to evolve its unique structure and complex functions.

166 *Identification of genes functioning in early rumen development*

167 The above rumen specifically expressed genes are identified in postnatal rumen,  
168 but the development of the rumen structure mainly occurs during early embryo  
169 stages<sup>28,29</sup>. In order to identify genes functioning in this critical stage, we performed  
170 five RNA sequencing from the ruminal and esophageal epithelium cells of four 60  
171 days' sheep embryos, the stage at which the ruminal epithelium starts to  
172 differentiate<sup>28,29</sup> (**Supplementary Table 1**). We identified 285 rumen up-regulated  
173 differentially expressed genes (DEGs) compared to the esophagus (**Supplementary**  
174 **Table 18**). These are enriched in cell-cell junction (adjusted *P* value =  $8.33 \times 10^{-3}$ ) and  
175 desmosome organization (adjusted *P* value =  $1.47 \times 10^{-3}$ ) (**Supplementary Table 19**).  
176 We also found 1,840 rumen down-regulated DEGs which are enriched in anatomical  
177 structure morphogenesis (adjusted *P* value =  $1.39 \times 10^{-15}$ ) (**Supplementary Table 18**,  
178 **20**). These results indicate that the specific epithelial histology of the rumen wall  
179 constitutes the most significant developmental genetic reprogramming as the organ  
180 forms and grows in the embryo. After filtering redundancy, we combined the 655  
181 rumen specifically expressed genes with the 285 rumen up-regulated DEGs compared  
182 to the esophagus at the key development stage and eventually obtain 846 rumen key  
183 genes which we consider crucial for rumen development and evolution.

184

185 **Evolutionary analyses on the rumen key genes**

186 Based on the data from ruminant comparative genomics<sup>30</sup>, we employed evolutionary  
187 genomic analyses on the 846 rumen key genes in the evolutionary context of 51  
188 ruminants and 12 other mammals, by identifying ruminant-specific conserved  
189 nonexonic elements (RSCNEs) ( $\geq 20$  bp), newly evolved genes and positively selected  
190 genes (PSGs) to systematically investigate the genetic changes associated with these  
191 rumen key genes. In the common ancestor of Ruminantia, we identified 657 genes with  
192 RSCNEs (**Supplementary Table 21**), two newly evolved genes and 28 PSGs  
193 (**Supplementary Table 22**) among the 846 rumen key genes. They are mainly  
194 involved in keratin filament binding, serine-type peptidase activity, ketone body  
195 metabolism and detection of bacterium.

196 *Improved ketone body synthesis in rumen*

197 In the pathway of synthesis and degradation of ketone bodies, *HMGCS2* and  
198 *SLC16A1* were under positive selection in the common ancestor of ruminants (**Fig. 2c**,  
199 **3a; Supplementary Table 9, 10, 22**), and had ruminant-specific mutations when  
200 compared to non-ruminant mammals (**Fig. 3b**). Of the five ruminant-specific amino  
201 acid changes in the HMGCS2 protein, four are located in the HMG-CoA synthase  
202 domain (PF01154) (**Fig. 3b**). To further examine the effects of these mutations on the  
203 enzyme structure, we conducted three-dimensional (3D) structure simulations, and  
204 found that mutations in HMG-CoA synthase domain could induce a change of the  
205 protein 3D structure when compared to the human HMGCS2 protein (**Fig. 3c**). We also  
206 noted that the *SLC16A1* gene, which participates in the transportation of ketone bodies

207 into the blood<sup>24</sup>, exhibited seven ruminant-specific mutations, six of which are located  
208 in the MFS\_1 domain (PF07690), resulting in a domain structure change as revealed by  
209 protein structure homology-modeling (**Fig. S2, S3**). We therefore hypothesized that the  
210 changes in *HMGCS2* and *SLC16A1* may result in a more efficient ketone body  
211 metabolism in ruminants. This is supported by *HMGCS2* being the key rate-limiting  
212 enzyme in the ketogenesis pathway<sup>24</sup>. To explore the functional relevance of these  
213 mutations, we synthesized sheep and human *HMGCS2* orthologs *in vitro* and tested  
214 their enzyme synthetic activities by measuring the activities in a reconstituted system  
215 consisting of the enzyme and substrate (**Supplementary Note**). The sheep *HMGCS2*  
216 (S) protein variant exhibits significantly higher metabolic efficiency than human  
217 proteins (H) (~2-fold increase, t-test,  $P < 0.001$ ) (**Fig. 3d**). The enzyme activity of  
218 human *HMGCS2* containing the five ruminant-specific amino acids replacements  
219 (H-5R) is also significantly higher than the regular human protein (~1.5-fold increase,  
220  $P < 0.01$ ), while sheep *HMGCS2* with the corresponding five human amino acid  
221 replacements (S-5H) exhibits significantly lower enzymatic activities than the sheep  
222 protein (~2-fold decrease,  $P < 0.001$ ) (**Fig. 3d**). These results confirm that ruminants  
223 have evolved a more efficient ketogenesis than that of other mammals.

224 *Immune system and microbial regulation*

225 We identified one PSG (*NOD2*) (**Supplementary Table 22**) and two newly  
226 evolved genes (*DEFB1* and *LYZ1*) in the rumen key gene list that are involved in  
227 immune functions. Among these, our transcriptomic data show that *NOD2* was  
228 co-expressed with the macrophage cells, and highly expressed in the rumen compared

229 to both the FC stomachs of camels and cetaceans (**Supplementary Table 2, 9**). We  
230 detected 11 ruminant-unique amino acid changes in NOD2, resulting in domain  
231 structure changes as revealed by protein structure homology-modeling (**Fig. S4, S5**).  
232 This gene functions in the upstream part of IL17 signaling pathway, activating the  
233 Th17 cells to produce IL17F as part of the gastrointestinal immune system<sup>31</sup> (**Fig. 4a**).  
234 The IL17 signaling pathway protects the host against extracellular pathogens via  
235 activating downstream pathways to induce the expression of antimicrobial peptides<sup>32</sup>.  
236 Among the newly evolved genes in the ancestor of ruminants, we identified a  
237 rumen key gene, *DEFB1*, which belongs to the beta-defensin family that have  
238 important roles as antimicrobial peptides in the resistance of epithelial surfaces to  
239 microbial colonization (**Supplementary Table 2**). In addition, we identified one  
240 newly evolved rumen key gene *LYZ1* in the lysozyme c family (**Supplementary Table**  
241 **2**), which may protect the rumen epithelium from the activity of pathogenic bacteria<sup>18</sup>.  
242 We predicted that the *LYZ1* contains a ruminant-specific 20 amino-acid-chain that  
243 encodes a probable transmembrane anchor (**Fig. S6, S7**), suggesting that the *LYZ1* gene  
244 encodes a secreted membrane-anchored protein, which may act on the rumen  
245 environment.  
246 To validate the functions of these two newly evolved genes, we synthesized  
247 *DEFB1* and *LYZ1* *in vitro* and tested their antibacterial ability by performing an  
248 inhibition zone assay on agarose plates with *Escherichia coli* (American Type Culture  
249 Collection, ATCC 25922) and *Staphylococcus aureus* (ATCC 29213) as representative  
250 of Gram-negative and -positive bacteria (**Supplementary Note**). The *DEFB1* (**Fig. 4b**)

251 and LYZ1 (**Fig. 4c**) protein both showed antibacterial activity to *S. aureus*, but not *E.*  
252 *coli*. This characteristic of selective inhibition of Gram-positive bacteria is similar to  
253 that of monensin, which is commonly used as an antibiotic drug that regulates the  
254 microbiome and increases ruminant feed conversion efficiency<sup>33,34</sup>. Taken together,  
255 these results highlight that several important antibacterial functions are uniquely  
256 evolved in the rumen relative to other similar organs, and that some of these may  
257 work by specifically managing the microbiome composition.

258 *New regulatory elements related to rumen epithelium absorbtion function*

259 We searched among 221,166 RSCNEs to identify candidate regulatory regions in  
260 the vicinity of rumen key genes. We found that 657 of the 846 rumen key genes have  
261 nearby RSCNEs (**Supplementary Table 21**). To assess the regulatory role of these  
262 RSCNEs in the recruitment of increased gene expression in the rumen, we performed  
263 eight ATAC-seq libraries of the ruminal and esophageal epithelium cells from four 60  
264 days' sheep embryos (**Supplementary Table 23; Supplementary Note**). Our analysis  
265 indicates that 243 rumen key genes have nearby RSCNEs overlapping with identified  
266 open accessible peaks (**Supplementary Table 24**), and these genes are enriched in  
267 epidermal cell differentiation (adjusted *P* value =  $4.82 \times 10^{-19}$ ) (**Supplementary Table**  
268 **25**). In the comparison of ATAC-seq between the rumen and esophagus, we identified  
269 3,904 rumen-specific and 5,531 esophagus-specific open differentially accessible  
270 peaks (DAPs) (**Fig. S8; Supplementary Table 26**). Interestingly, we found 267 and  
271 478 RSCNEs ( $\geq 20$  bp) overlapping with rumen-specific and esophagus-specific  
272 DAPs, which is highly statistically significant (Fisher's exact test, both *P* value = 0.00).

273 Rumen-specific DAP-associated RSCNEs are physically near 22 rumen key genes  
274 (**Supplementary Table 27**). Among these genes, *CRNN* is one of the genes in the  
275 epidermal differentiation complex (EDC) locus, which is essential for the cornified cell  
276 envelope in rumen<sup>15</sup>, and is implicated in several epithelial malignancies in human<sup>35</sup>. A  
277 rumen-specific DAP-associated RSCNE with six ruminant-specific mutations was  
278 found at the 5' upstream of *CRNN* of ruminants, which might play a role in regulating  
279 its expression in rumen. Concordantly, *DMRT2* is a key transcriptional factor in the  
280 dermomyotome organization and *DMRT2*-deficient mice have epithelial morphology  
281 abnormalities<sup>36</sup>. We observed that *DMRT2* has five rumen-specific DAP-associated  
282 RSCNEs in its 3' downstream region, potentially causing high *DMRT2* expression in  
283 rumen.

284 Interestingly, *WDR66* is not only highly expressed in the rumen compared with  
285 both the FC stomachs of camels and cetaceans but also under positive selection in the  
286 common ancestor of Ruminantia (**Fig. 5a; Supplementary Table 9, 22**). It regulates  
287 the expression of occludin, which tightens the intercellular space and enables epithelial  
288 permeability<sup>37</sup>. We observed 10 ruminant-specific non-synonymous mutations and one  
289 rumen-specific DAP-associated RSCNE in the intronic region of *WDR66* (**Fig. 5b; Fig.**  
290 **S9; Supplementary Table 27**). In order to assess the regulatory activity of this  
291 particular RSCNE, we cloned it into a luciferase reporter vector (pGL3-Promoter) and  
292 transfected it into both sheep and goat fibroblasts *in vitro*. The RSCNE showed  
293 significantly higher luciferase transcriptional activation compared to the  
294 pGL3-Promoter control (t-test,  $P < 0.05$ ) (**Fig. 5c**), confirming that it acts as an

295 enhancer. Therefore, these DAP-associated RSCNEs might plausibly have exerted  
296 novel *cis*-regulation of the rumen key genes, thus providing a mechanistic explanation  
297 of how the rumen might have recruited these genes from other tissues. Hence, we  
298 propose a central role of such regulatory elements in the development and evolution of  
299 rumen structure and function.

300 *Positively selected genes involved in rumen epithelium absorption*

301 We observed that eight rumen key genes involved in the cell junction biological  
302 process (*WDR66*, *COL7A1*, *EVPL*, *KRT14*, *CLDN23*, *F2RL1*, *TMPRSS13* and  
303 *TMPRSS11A*) were under positive selection in ruminants (**Fig. 5a**; **Fig. S9-S16**;  
304 **Supplementary Table 22**). Non-synonymous changes in these genes may result in the  
305 change of cell junctions, which may break the epithelium barrier and increase the  
306 epithelium absorption properties<sup>38-42</sup>. *COL7A1* is highly expressed in the rumen of fetal  
307 sheep, but not in the esophagus (**Supplementary Table 18**). We detected 17 unique  
308 amino acid (aa) changes in *COL7A1* in ruminants (**Fig. S10**). *COL7A1* is an anchoring  
309 fibril between the external epithelia and the underlying basal lamina<sup>39</sup>. Amino acid  
310 mutations in this gene are associated with epidermolysis bullosa, a condition in which  
311 tissue fluid diffuses through the intercellular space into the epidermis<sup>39</sup>. In addition,  
312 *TMPRSS13*, a membrane-anchored serine protease gene<sup>41</sup>, is highly expressed in rumen  
313 compared to esophagus (**Supplementary Table 18**). Interestingly, we identified five  
314 ruminant-specific aa changes in *TMPRSS13*, four of which are located in the  
315 trypsin-like serine protease domain (**Fig. S15**). It is reported that the deficiency of  
316 *TMPRSS13* in mice impairs stratum corneum formation and epidermal barrier

317 acquisition, accompanied by trans-epidermal fluid loss<sup>41</sup>. In normal epithelium cells  
318 (e.g., epithelium cells of skin), the epithelium barrier is produced by strong intracellular  
319 protein filaments crossing the cytoplasm and attaching to specialized junctions, which  
320 in turn ties the surfaces of adjacent cells either to each other or to the underlying basal  
321 lamina<sup>43</sup> (**Fig. 5a**). Given that the epithelium transportation and absorption functions  
322 are affected by the epithelium barrier, mutations in these cell junction-related genes  
323 may be related to metabolite uptaking function of the rumen.

324 **Discussion**

325 Our large quantity of transcriptomic data in adults and an early embryo rumen  
326 development stage provide a detailed comparative insight into the distinct gene  
327 expression profile of the rumen. Although there has been no consensus about the  
328 evolutionary relationship between the FC stomachs of camels, peccaries, cetaceans and  
329 ruminants<sup>21,44</sup>, it is unlikely that the multi-chambered stomach evolved independently  
330 four times in Cetartiodactyla exclusively. Therefore, the most parsimonious  
331 explanation is that they may have a single evolutionary origin, followed by  
332 specialization in the different lineages of the Cetartiodactyla due to their specific diets  
333 and niches. For instance, the FC stomachs of camels have evolved the ability to store  
334 water<sup>21,45</sup>, the FC stomachs of cetaceans has the capacity to mechanically grind food<sup>11</sup>,  
335 and the rumen provides efficient fermentation and metabolism of plant material. The  
336 gene expression profiles of the FC stomachs in ruminants, camels and cetaceans show  
337 that they are all highly similar to the esophagus, suggesting these organs share an  
338 anatomical origin from the esophagus (**Fig. 2a; Fig. S1**).

339 Based on our comparative genomic and functional data, we outline the genetic  
340 mechanisms underlying the origin, development and evolution of the rumen from the  
341 ancestral esophagus tissue. These genetic innovations are mainly related to epithelium  
342 absorption, ketone body metabolism and microbial regulation. Among the 846 rumen  
343 key genes (**Supplementary Table 2, 18**), we found that 657 (77.7%) genes have nearby  
344 RSCNEs (**Supplementary Table 21**), 28 genes are under positive selection  
345 (**Supplementary Table 22**) and two genes newly evolved in the common ancestor of

346 ruminants, suggesting these three types of genetic reprogramming all contributed to the  
347 structural and functional evolution of rumen. Notably, the majority of rumen key genes  
348 have RSCNEs nearby and our ATAC-seq validated that 243 rumen key genes had  
349 nearby RSCNEs overlapping with highly accessible chromatin (**Supplementary Table**  
350 **24**), suggesting the RSCNEs as regulatory elements may play a crucial role in rumen  
351 gene recruitment. The highly significant association between RSCNEs, rumen key  
352 genes and open accessible peaks is a strong indication of this, although there were also  
353 many RSCNEs that did not overlap with open accessible peaks in our ATAC-seq  
354 analysis. While this suggests that RSCNEs play other roles besides being regulatory  
355 elements, it is also possible that some were false negatives due to the limitations of  
356 development stages sampled in this study, which might have omitted some associations  
357 between rumen key genes and regulatory RSCNEs. Hence, a denser sampling of  
358 different developmental time points might expand the rumen key gene list and reveal  
359 novel regulatory roles of RSCNEs. Nevertheless, our study has revealed the important  
360 genetic mechanisms underlying the key evolutionary innovations of the rumen. The  
361 identified rumen key genes and their specific mutations provide a starting point for  
362 future studies of rumen development, and for understanding the interactions between  
363 rumen and microbiota. This will be key to further improvement of ruminant livestock,  
364 e.g. by providing a framework for manipulating the rumen fermentation process.

365 **Data availability**

366 The raw reads for all RNA-seq data, the ATAC-seq data from the rumen and the  
367 esophagus have been deposited at the Sequence Read Archive (SRA) under project  
368 number PRJNA485657.

369 **Acknowledgments**

370 This project was supported by the National Natural Science Foundation of China  
371 (31822052, 31572381), the National Thousand Youth Talents Plan to Y.J., National  
372 Natural Science Foundation of China (31660644) to S.H, National Natural Science  
373 Foundation of China (41422604) to S.L. We thank the members of the FANNG project  
374 for sharing their transcriptome data. We thank Yongchuan Li, Zhengzhi Wei, Zixin  
375 Yang, and Haiyu Gao from Institute of Deep-sea Science and Engineering, Chinese  
376 Academy of Sciences, for helping to collect samples from the porpoise and whale. We  
377 thank High-Performance Computing (HPC) of Northwest A&F University (NWAFU)  
378 for providing computing resources.

379 **Author contributions**

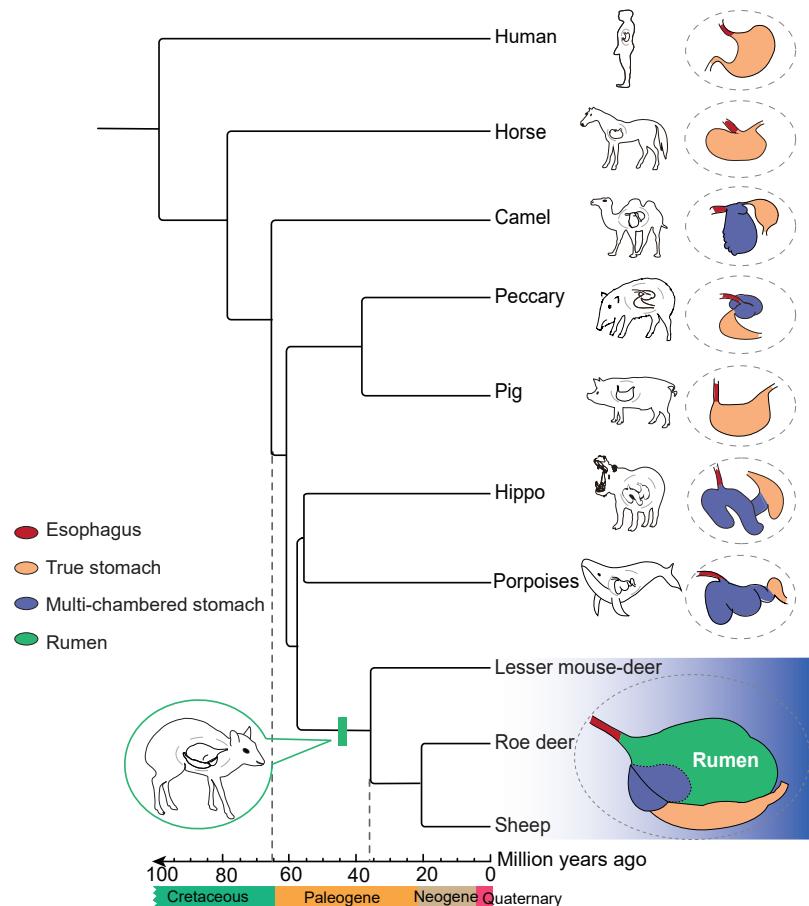
380 Y.J. and W.W. conceived the project and designed the research. X.P., Y.C., N.W., C.  
381 Z., and X.H. performed the majority of analysis with contributions from K.W., L.C.,  
382 Z.L., Z.Z., B.W., S.H.; Q.Q., S.M., X.L., WF., LL., Y.L., W.S., W.L., T.Z., J.H.,  
383 M.L., S.L., S.H., M.L., C.L., and Y.C. prepared the sheep, camels and cetaceans  
384 samples for transcriptomics and rumen and esophagus epithelium cells for ATAC-seq.  
385 H.L. performed the luciferase reporter assay. X.C., Y.Y. and Z.H. performed the  
386 inhibition zone assay and the enzyme synthetic activities assay. X.P., Z.L. and Y.C.

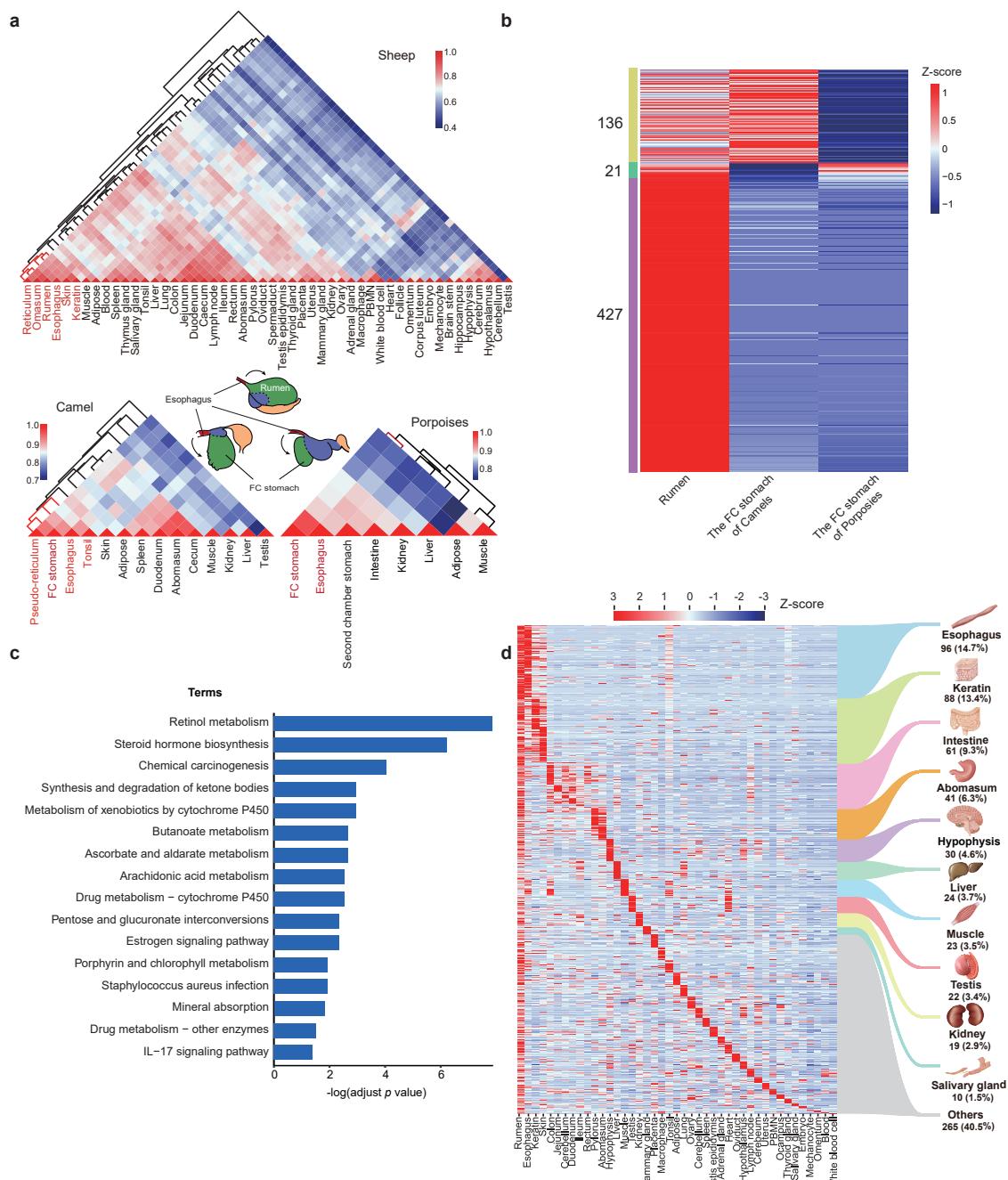
387 drafted the manuscript with input from all authors, whereas Y.J., W.W., R.H., B.PD.,

388 G.Z., X.W. and Y.W. revised the manuscript.

389 **Competing interests**

390 Two provisional Chinese patent applications on potential application in the antimicrobial and  
antibiotic substitute by way of the *DEFB1* gene and *LYZ1* gene have been filed by Northwest A&F  
University (application number 202010100677.8 and 202010097562.8), where Y.J., X.P., X.C, and  
W.W. are listed as inventors. The authors declare no competing interests.



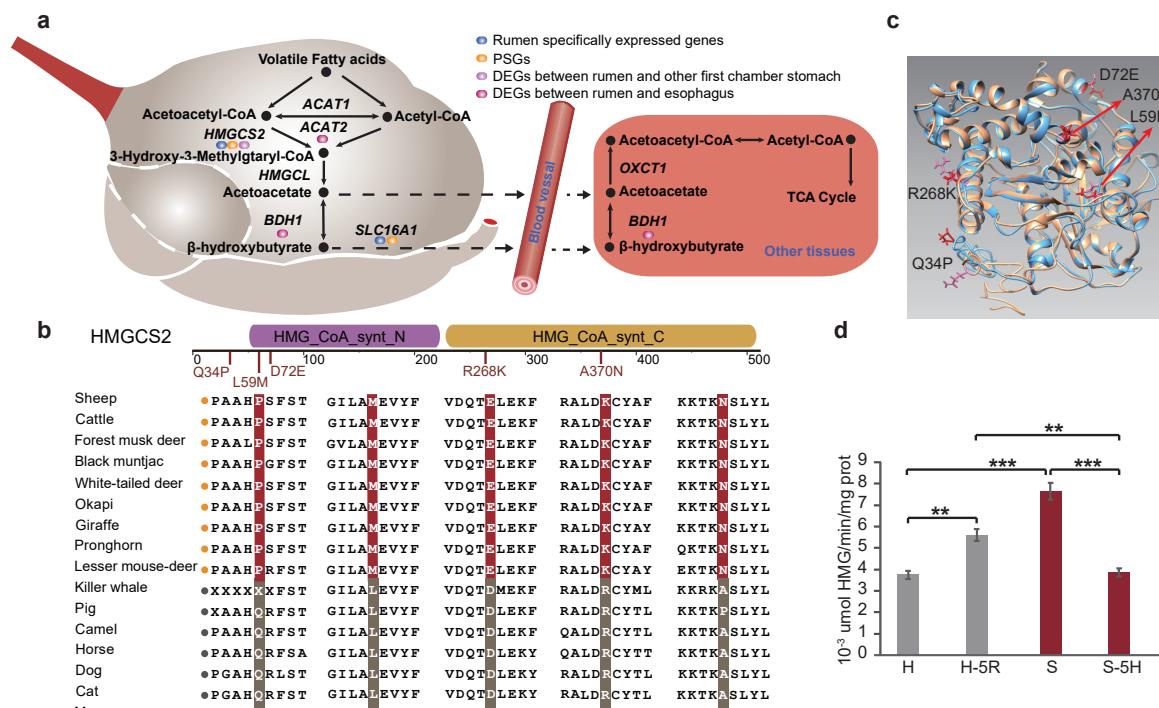


397

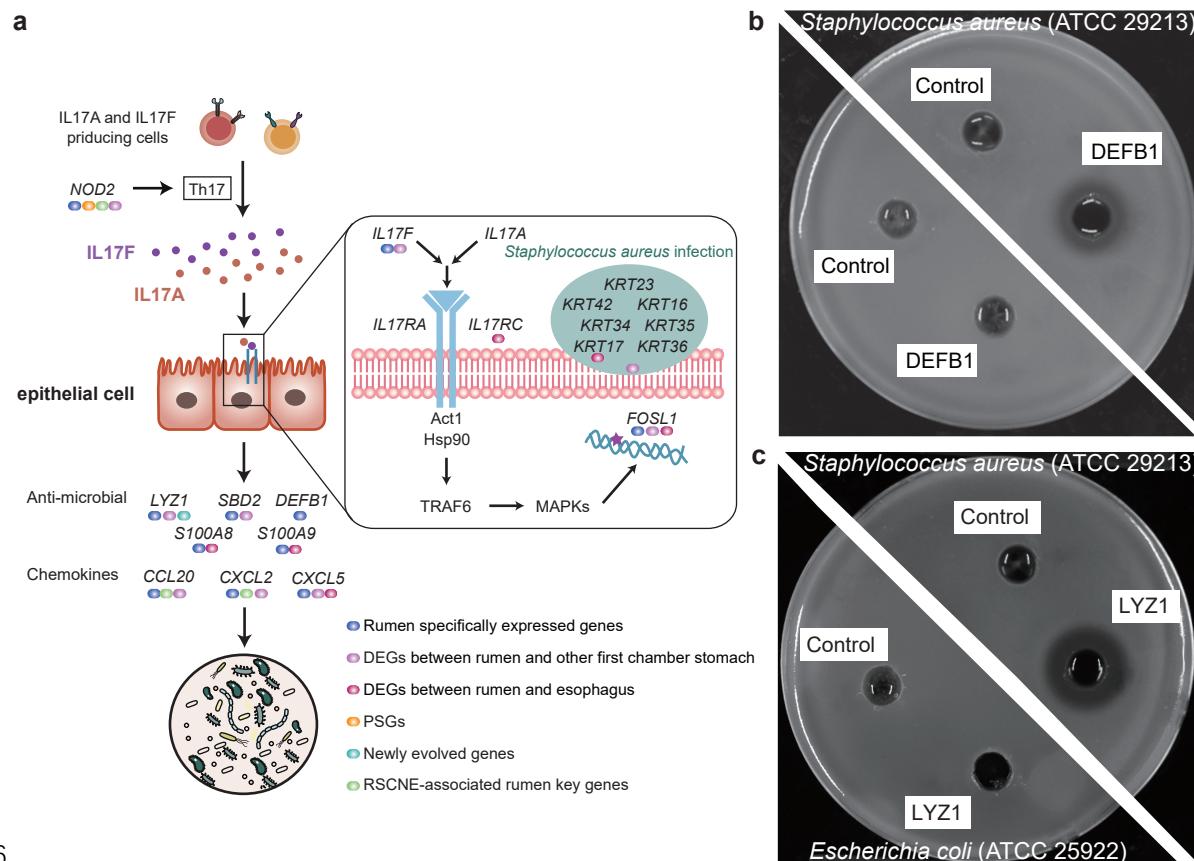
398 Fig. 2 | Comparisons of gene expression profile among rumen and other tissues. a,

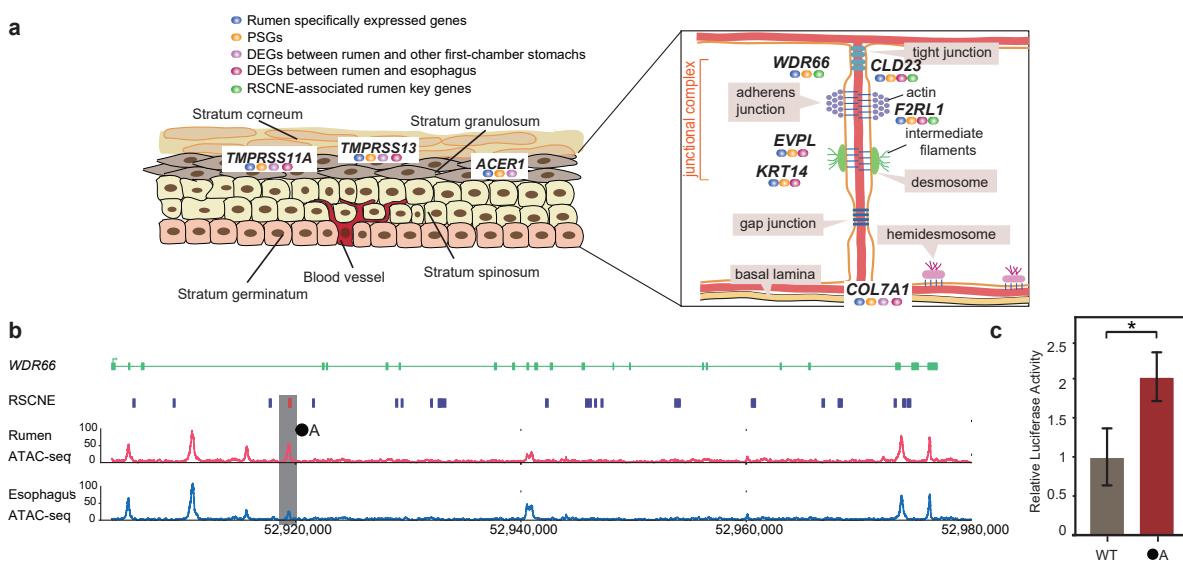
399 Hierarchical clustering results showing the relationships among 50 tissues of sheep and a heatmap  
400 showing the pairwise Spearman correlations between sheep tissues (the top triangle), between 14  
401 tissues of camels (lower left triangle) and between eight tissues of two cetaceans (lower right  
402 triangle). **b**, Heatmap of differentially expressed rumen specifically expressed genes among the  
403 rumen and other FC stomachs. The color bars on the left present 136 DEGs of the rumen relative to

404 the FC stomach of cetaceans (yellow), 21 DEGs relative to the FC stomach of camels (green), and  
405 427 DEGs relative to the FC stomach of both species (purple). The expression levels were  
406 normalized by Z-scores. **c**, KEGG pathway analysis of 427 rumen up-regulated DEGs relative to  
407 both the FC stomach of camels and cetaceans. **d**, Heatmap showing the gene expression profiles of  
408 all 655 rumen specifically expressed genes across 43 tissues of sheep. Different colored lines  
409 represent the tissues from which the rumen specifically expressed genes were recruited. Number of  
410 genes from each tissue is shown below the tissue name with the percentage of total genes recruited  
411 in parentheses.



413 **Fig. 3 | Genetic changes in the rumen ketone body metabolism genes and pathways. a,** Genes  
414 annotated in the ketone body metabolism are labeled with different color to indicate rumen  
415 specifically expressed genes (blue), positively selected genes in ruminant (orange) and  
416 differentially expressed genes between rumen and other FC stomachs (purple). The solid arrows  
417 represent ketone body metabolism pathways. The dashed arrows indicate the process of material  
418 transport from rumen to other tissues. **b,** Top panels: Structural domains of the HMGCS2 protein  
419 and the location of the ruminant specific mutations. Lower panel: Peptide sequence alignment of  
420 HMGCS2. The species is followed a yellow circle belonging to the ruminant. The red highlighting  
421 indicates ruminant-specific amino acid mutations. **c,** Predicted tertiary structures of the HMGCS2  
422 of ruminant (blue) and other mammals (orange), respectively. **d,** Enzyme activities of HMGCS2  
423 compared with those of sheep and human in vitro. H: human, H-5R: human HMGCS2 with five  
424 ruminant aa replacements, S: sheep, S-5H: sheep HMGCS2 with five human aa replacements. \*\*  
425  $p$  value  $< 0.01$ , \*\*\*  $p$  value  $< 0.001$  calculated from the t test. Data are shown as mean  $\pm$  s.d.





434

435 **Fig. 5 | Genetic changes related to rumen epithelium transportation and absorption. a,**  
436 Diagram of rumen epithelial cell proteins involved in epithelium permeability identified in the  
437 common ancestor of the ruminants. Rumen specifically expressed genes (blue), positively selected  
438 genes in ruminant (orange), differentially expressed genes between rumen and other FC stomachs  
439 (purple), differentially expressed genes between rumen and esophagus (red), and  
440 RSCNE-associated rumen key genes (green). Note the junction structure (desmosome) between  
441 keratinocytes of the ruminal epithelium has been degraded, instead the enlarged intercellular space  
442 with copious blood supply enables metabolites absorption in the ruminal epithelium<sup>47</sup>. **b**, Gene  
443 structure of *WDR66* based on the NCBI Oar\_v4.0 annotation shown above. Green boxes represent  
444 exons. Purple bars indicate ruminant-specific conserved non-exonic elements (RSCNEs). Red and  
445 blue bars indicate ATAC-seq peaks of the ruminal and esophageal epithelium cell, respectively.  
446 The grey rectangle box is the overlapping element of RSCNE and ATAC-seq which is located in  
447 the intron region. **c**, The luciferase activity of the pGL3-Promoter (WT) and the pGL3-Promoter  
448 with the RSCNE (●A). \*  $p$  value < 0.05 calculated from the t test. Data are shown as mean  $\pm$  s.d.

449 **References**

1. Gregory, T. R., The evolution of complex organs. *Evo. Edu. Outreach* **1**, 358-389 (2008).
2. Land, M. F. & Fernald, R. D., The evolution of eyes. *Annu. Rev. Neurosci.* **15**, 1-29 (1992).
3. Land, M. F., The optics of animal eyes. *Contemp. Phys.* **29**, 435-455 (1988).
4. Gallant, J. R. *et al.*, Genomic basis for the convergent evolution of electric organs. *Science* **344**, 1522-1525 (2014).
5. Griffith, O. W. & Wagner, G. P., The placenta as a model for understanding the origin and evolution of vertebrate organs. *Nat. Ecol. Evol.* **1**, 0072 (2017).
6. Lynch, V. J., Leclerc, R. D., May, G. & Wagner, G. P., Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat. Genet.* **43**, 1154-1159 (2011).
7. Wang, Y. *et al.*, Genetic basis of ruminant headgear and rapid antler regeneration. *Science* **364**, v6335 (2019).
8. Janis, C., The evolutionary strategy of the equidae and the origins of rumen and cecal digestion. *Evolution* **30**, 757-774 (1976).
9. Dehority, B. A., Gastrointestinal tracts of herbivores, particularly the ruminant: Anatomy, physiology and microbial digestion of plants. *J. Appl. Anim. Res.* **21**, 145-160 (2002).
10. Novacek, M. J., Mammalian phylogeny: Shaking the tree. *Nature* **356**, 121-125 (1992).
11. Mathiesen, S. D. *et al.*, Digestive physiology of minke whales. *Developments in Marine Biology* **4**, 351-359 (1995).
12. Cantalapiedra, J. L. *et al.*, Dietary innovations spurred the diversification of ruminants during the Caenozoic. *Proc. Biol. Sci.* **281**, 20132746 (2014).
13. De Tarso, S. G. D. S., Oliveira, D. & Afonso, J. A. B., Ruminants as part of the global food system: How evolutionary adaptation and diversity of the digestive system brought them to the future. *J. Dairy Vet. Anim. Res.* **3**, 171-176 (2016).
14. Zeder, M. A., Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 11597-11604 (2008).
15. Xiang, R., Oddy, V. H., Archibald, A. L., Vercoe, P. E. & Dalrymple, B. P., Epithelial, metabolic and innate immunity transcriptomic signatures differentiating the rumen from other sheep and mammalian gastrointestinal tract tissues. *PeerJ* **4**, e1762 (2016).
16. Millen, D. D., Mario, D. B. A. & Rodrigo, D. L. P., *Rumenology*. (Springer, Switzerland, 2016).
17. Huttner, K. M., Brezinski-Caliguri, D. J., Mahoney, M. M. & Diamond, G., in *Molecular and Cellular Studies of Rumen Epithelial Metabolism* (1998).
18. Yu, J. *et al.*, The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* **344**, 1164-1168 (2014).
19. Clark, E. L. *et al.*, A high resolution atlas of gene expression in the domestic sheep (*Ovis aries*). *Plos Genet.* **13**, e1006997 (2017).
20. Stevens, C. E. & Hume, I. D., *Comparative physiology of the vertebrate digestive system*. (Cambridge University Press, 2004)
21. Vallenas, A., Cummings, J. F. & Munnell, J. F., A gross study of the compartmentalized stomach of two new-world camelids, the llama and guanaco. *J. Morphol.* **134**, 399-423 (1971).
22. Tarpley, R. J., Sis, R. F., Albert, T. F., Dalton, L. M. & George, J. C., Observations on the

492 anatomy of the stomach and duodenum of the bowhead whale, *Balaena mysticetus*. *Am. J. Anat.*  
493 **180**, 295-322 (1987).

494 23. Xiong, Z. *et al.*, PAX9 regulates squamous cell differentiation and carcinogenesis in the  
495 oro-oesophageal epithelium. *J. Pathol.* **244**, 164-175 (2018).

496 24. Newman, J. C. & Verdin, E., Ketone bodies as signaling metabolites. *Trends in Endocrinology &*  
497 *Metabolism* **25**, 42-52 (2014).

498 25. Thumelin, S., Forestier, M., Girard, J. & Pegorier, J. P., Developmental changes in mitochondrial  
499 3-hydroxy-3-methylglutaryl-CoA synthase gene expression in rat liver, intestine and kidney.  
500 *Biochem. J.* **292**, 493-496 (1993).

501 26. Baldwin, R. L., McLeod, K. R., Klotz, J. L. & Heitmann, R. N., Rumen development, intestinal  
502 growth and hepatic metabolism in the pre- and postweaning ruminant. *J. Dairy Sci.* **87**, E55-E65  
503 (2004).

504 27. Leng, R. A. & Nolan, J. V., Nitrogen metabolism in the rumen. *J. Dairy Sci.* **67**, 1072-1089  
505 (1984).

506 28. Wardrop, I. D., Some preliminary observations on the histological development of the  
507 fore-stomachs of the lamb I. Histological changes due to age in the period from 46 days of foetal  
508 life to 77 days of post-natal life. *J. Agric. Sci.* **57**, 335-341 (1961).

509 29. Fath, E. M., Schwarz, R. & Ali, A. M., Micromorphological studies on the stomach of sheep  
510 during prenatal life. *Anat. Histol. Embryol.* **12**, 139-153 (1983).

511 30. Chen, L. *et al.*, Large-scale ruminant genome sequencing provides insights into their evolution  
512 and distinct traits. *Science* **364**, (2019).

513 31. Gaffen, S. L., An overview of IL-17 function and signaling. *Cytokine* **43**, 402-407 (2008).

514 32. van Beelen, A. J. *et al.*, Stimulation of the intracellular bacterial sensor NOD2 programs dendritic  
515 cells to promote interleukin-17 production in human memory T cells. *Immunity* **27**, 660-669  
516 (2007).

517 33. Thornton, J. H. & Owens, F. N., Monensin supplementation and in vivo methane production by  
518 steers. *J. Anim. Sci.* **52**, 628-634 (1981).

519 34. Duffield, T. F., Merrill, J. K. & Bagg, R. N., Meta-analysis of the effects of monensin in beef  
520 cattle on feed efficiency, body weight gain, and dry matter intake. *J. Anim. Sci.* **90**, 4583-4592  
521 (2012).

522 35. Kyriouliou, M., Huber, M. & Hohl, D., The human epidermal differentiation complex: Cornified  
523 envelope precursors, S100 proteins and the 'fused genes' family. *Exp. Dermatol.* **21**, 643-649  
524 (2012).

525 36. Seo, K. W. *et al.*, Targeted disruption of the DM domain containing transcription factor Dmrt2  
526 reveals an essential role in somite patterning. *Dev. Biol.* **290**, 200-210 (2006).

527 37. Wang, Q., Ma, C. & Kemmner, W., Wdr66 is a novel marker for risk stratification and involved  
528 in epithelial-mesenchymal transition of esophageal squamous cell carcinoma. *Bmc Cancer* **13**,  
529 137 (2013).

530 38. Risk, J. M. *et al.*, Envoplakin, a possible candidate gene for focal NEPPK/esophageal cancer  
531 (TOC): The integration of genetic and physical maps of the TOC region on 17q25. *Genomics* **59**,  
532 234-242 (1999).

533 39. Sheth, J., Mistri, M., Patel, H., Ankleshwaria, C. & Parikh, A., Autosomal dominant mutation in  
534 COL7A1 gene causing epidermolysis bullosa dystrophica. *Mol. Cytogenet.* **7**, P58 (2014).

535 40. Bousquet, O. *et al.*, The nonhelical tail domain of keratin 14 promotes filament bundling and

536        enhances the mechanical properties of keratin intermediate filaments in vitro. *J. Cell Biol.* **155**,  
537        747-754 (2001).

538        41. Madsen, D. H., Szabo, R., Molinolo, A. A. & Bugge, T. H., TMPRSS13 deficiency impairs  
539        stratum corneum formation and epidermal barrier acquisition. *Biochem. J.* **461**, 487-495 (2014).

540        42. Gareus, R. *et al.*, Normal epidermal differentiation but impaired skin-barrier formation upon  
541        keratinocyte-restricted IKK1 ablation. *Nat. Cell Biol.* **9**, 461-469 (2007).

542        43. Alberts, B., Johnson, A. & Lewis, J., *Molecular biology of the cell*. (Garland Pub., 1983).

543        44. Mead, J. G., in *Encyclopedia of Marine Mammals (Second Edition)* (Academic Press, 2009).

544        45. von Engelhardt, W., Dycker, C. & Lechner-Doll, M., Absorption of short-chain fatty acids,  
545        sodium and water from the forestomach of camels. *J. Comp. Physiol. B* **177**, 631-640 (2007).

546        46. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B., TimeTree: A resource for timelines,  
547        timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812-1819 (2017).

548        47. Phillipson, A. T., *Physiology of digestion and metabolism in the ruminant*. (Oriel Press Ltd.,  
549        1970).