Article

Genomic Identification, Evolution, and Expression Analysis of Collagen Genes Family in Water Buffalo during Lactation

Xingrong Lu †, Anqin Duan †, Shasha Liang, Xiaoya Ma and Tingxian Deng *

Guangxi Provincial Key Laboratory of Buffalo Genetics, Breeding and Reproduction Technology, Buffalo Research Institute, Chinese Academy of Agricultural Sciences, Nanning 530001, China; luxingrong074@163.com (X.L.); duanaq321@163.com (A.D.); cangshucangshu@126.com (S.L.); maxiaoya8899@163.com (X.M.)

* Correspondence: dtx282000@163.com; Tel.: +86-1359-7111-864

† These authors contributed equally to this work.

Abstract: Collagens, as extracellular matrix proteins, support cells for structural integrity and contribute to support mammary basic structure and development. This study aims to perform the genomic identification, evolution, and expression analyses of the collagen gene family in water buffalo (*Bubalus bubalis*) during lactation. A total of 128 buffalo collagen protein sequences were deduced from the 45 collagen genes identified in silico from buffalo genome, which classified into six groups based on their phylogenetic relationships, conserved motifs, and gene structure analyses. The identified collagen sequences were unequally distributed on 16 chromosomes. The tandem duplicated genes were found within three chromosomes, while only one segmental event occurred between Chr3 and Chr8. Collinearity analysis revealed that a total of 36 collagen gene pairs were orthologous between buffalo and cattle genomes despite having different chromosome numbers. Comparative transcription analyses revealed that a total of 23 orthologous collagen genes were detected in the milk samples at different lactation periods between the two species. Notably, the duplicated gene pair of *COL4A1-COL4A2* during lactation had a higher mRNA expression level than that of cattle, while a higher expression level of *COL6A1-COL6A2* pair was found in cattle compared with that of buffalo. The present study provides useful information for investigating the potential functions of the collagen family in buffalo during lactation and helps in the functional characterization of collagen genes in additional research.

Keywords: buffalo; collagen family; evolutionary relationship; gene expression

1. Introduction

Collagens represent the most abundant protein of the extracellular matrix (ECM) in animals. To date, the knowledge about the molecular structure, biosynthesis, and function of the collagen family has emerged [1]. Collagens comprised 28 members in vertebrates, and are multidomain proteins that commonly possessed at least one triple-helical domain. Types I to IV of the collagen family are the most common, each serving different functions with the appropriate structure. For example, collagen IV is an important component of the ECM in the mammary glands [2,3]. Chen et al. [4] found the *COL4A1* was significantly down regulated in the inflammation-associated fibroblasts extracted from bovine mammary glands with clinical mastitis compared with normal fibroblasts from a slaughtered dairy cow, implying that it might involve the ECM remodeling or immune response. Crisà et al. [5] reported that *COL4A1* was higher expressed in mature milk than colostrum milk in goats. Undoubtedly, as the major component of ECM, collagens play a vital role in mammary gland development and

lactation. It is well known that the mammary gland is a dynamic organ of mammals that produce milk to meet nourishment requirements for the offspring [6,7]. Therefore, it is of great importance to reveal the putative function of the collagen family affecting the mammary gland development and lactation. Although the expression profiles and functions of collagens have been already reported, expression patterns and putative functions of the collagen gene family in water buffalo (*Bubalus bubalis*) during lactation are poorly characterized.

Water buffalo is the second largest source of milk production mainly distributed in tropical and subtropical areas, providing more than 5% of the world's milk supply [8]. Remarkably, buffalo milk is more abundant in fat and protein than cow's milk [9], which has attracted widespread attention from the dairy industry. In the past several decades, numerous high-throughput data were generated and utilized for identifying the candidate genes related to traits of interest in buffalo, such as the potential genes related to milk or productive traits [10,11], transcriptome profiles of buffalo embryos with normal and retarded growth [12], and maternally expressed proteins in buffalo oocyte [13]. All these data, along with the complete buffalo genome sequence [14], provide the possibility to perform the gene family analysis at a genome-wide level. In this regard, we identified the collagen family in the present study from the buffalo genome, and analyzed their evolutionary relationship, sequence features, chromosomal location, gene duplication, and dynamic expression patterns in response to different lactations (early lactation, mid-lactation, and late lactation) in milk. Our results provide some insights into the understanding of the buffalo collagen family affecting mammary gland development and lactation, and present vital evidence for future functional studies.

2. Materials and Methods

2.1. Identification of the Buffalo Collagen Genes

Whole-genome data of six representative mammals including the human (GRCh38.p12), cattle (ARS-UCD1.2), buffalo (UOA_WB_1), goat (ARS1), sheep (Oar_rambouillet_v1.0), and horse (EquCab3.0) were downloaded from the National Center for Biotechnology Information (NCBI) Genome database [15], aiming to identify the buffalo collagen genes. The HMM profile of the collagen domain (PF01391) from the Pfam database [16] was used to search the buffalo dataset using the HMMER [17,18] software with an E-value cut-off of $1.0 \times e^{-5}$. The identified buffalo collagen protein sequences were also validated by the BLAST with collagens from the other five mammals as queries. Further, the ClustalW algorithm was used for the multiple sequence alignment of collagen with the full-length protein sequences. The aligned sequences were used for constructing the neighbor-joining (NJ) phylogenetic tree of the collagen family using the MEGA7 [19] software with the Poisson model, pairwise deletion, and 1000 bootstrap resampling.

2.2. Sequence Analysis

The ExPASy proteomics server was utilized for the prediction of theoretical molecular weight (MW) and isoelectric points (pI) of the buffalo collagen family [20]. The gene structure was measured and visualized using the TBtools version 0.6657 [21]. Exon–intron structure analysis was conducted by the buffalo genome annotation file using the in-house scripts. We analyzed conserved protein motifs of collagen proteins in buffalo, using the MEME programs [22] at a maximum number of motifs, 10.

Chromosome locations of each collagen gene were obtained from their genome resources. Collagen duplication events were identified using the Multiple Collinearity Scan Toolkit (MCScanX) previously described by Wang et al. [23]. Overall, a total of 58,532 buffalo protein sequences were analyzed using the BLAST search with E-value $< 1.0 \times e^{-5}$ and then the corresponding gene positions files were obtained from the buffalo genome annotation file with the GFF format, which can be set as the input files of MCScanX program.

For the duplicated collagen genes, we further performed the divergence estimates and diversity analysis. In brief, MEGA7 software with the MUSCLE algorithm was first used for the pairwise

alignment of the homologous collagen gene pairs. Subsequently, the DnaSP v6.0 [24] software was employed to calculate the pairwise synonymous (Ks) and nonsynonymous (Ka) numbers of substitutions corrected for multiple hits.

2.3. Comparative Transcriptomic Analyses for the Milk Samples of Buffalo and Cattle

To explore the expression difference of collagens between buffalo and cattle in milk samples at different lactation, including early lactation, mid-lactation, and late lactation, two published RNA-seq data (BioProject: PRJNA419906 and PRJNA453843) were selected and employed further analyses. Briefly, the FASTQC [25] program was utilized to remove the adapter sequence from the raw sequence reads. We aligned the clean data from buffalo and those from cattle to their genomes (buffalo: UOA_WB_1 and cattle: ARS-UCD1.2) using HISAT2 [26] with default parameters, respectively. The StringTie [27] was utilized for calculating the gene or transcript count matrix. Transcripts per million (TPM) values for each gene were obtained using the DESeq2 [28] R-package. The one-to-one orthologous collagens in milk samples were selected and then merged, as described by Yu et al. [29]. Clustering and generation of a heat map of TPM values for the selected genes were performed using the pheatmap package in R.

2.4. Real-time Quantitative PCR

Murrah buffalo was used as the source of animal material in the present study, which was kept at the Buffalo Research Institute, Chinese Academy of Agricultural Sciences (BRI-CAAS). Biopsy samples of mammary gland tissue from eight buffaloes were collected on days 7 (D7), 140 (D140), and 280 (D280) after calving, which was further used for the real-time quantitative PCR (qRT-PCR). The biopsy procedure was performed based on the method reported by Schmitz et al. [30]. All fresh samples were immediately preserved in liquid nitrogen until use.

Total RNA for different buffalo tissues was isolated using the RNA Plus reagent (Tiangen, China). Quality of the total RNA was assessed using the NanoDrop2000 (Thermo Fisher Scientific, Wilmington, DE, USA) and gel electrophoresis. For 2 μg total RNA, the first-strand cDNA was synthesized using a reverse transcriptase kit (Takara, China). SYBR Premix Ex Taq (Takara) were used for qRT-PCR analysis, which was monitored on the LightCycler 480 (Roche, Switzerland), and each reaction was performed in triplicate. The expression levels of the buffalo collagen family were analyzed by the $2^{-\Delta\Delta Ct}$ method [31] and normalized using the expression of RPS9 and RPS15 analysis. Primers used for the qRT-PCR analysis are shown in Table S1.

3. Results

3.1. Genomic Identification of Buffalo Collagen Genes

To identify the collagen family members, a total of 128 non-redundant protein sequences encoded by 45 collagen genes were predicted from the buffalo whole genome using the BLAST and HMMER software (Table S2). The open reading frames (ORFs) of the collagen protein isoforms ranged from 1317 to 9657 bp in length encoding the protein of 438 to 3218 residues, with the predicted MW from 44.65 to 347.21 kDa. The pI values of these isoforms ranged from 4.48 to 10.51. Furthermore, the phylogenetic analysis revealed that all 45 collagen genes could be divided into six groups (Figure 1). Group I was the top one with the larger numbers of collagen genes ($n = 9$), while the group VI was the smaller one ($n = 2$). The constructed dendrogram further showed that the buffalo collagen family was usually the most closely evolutionary relationship with the other five representative mammals.
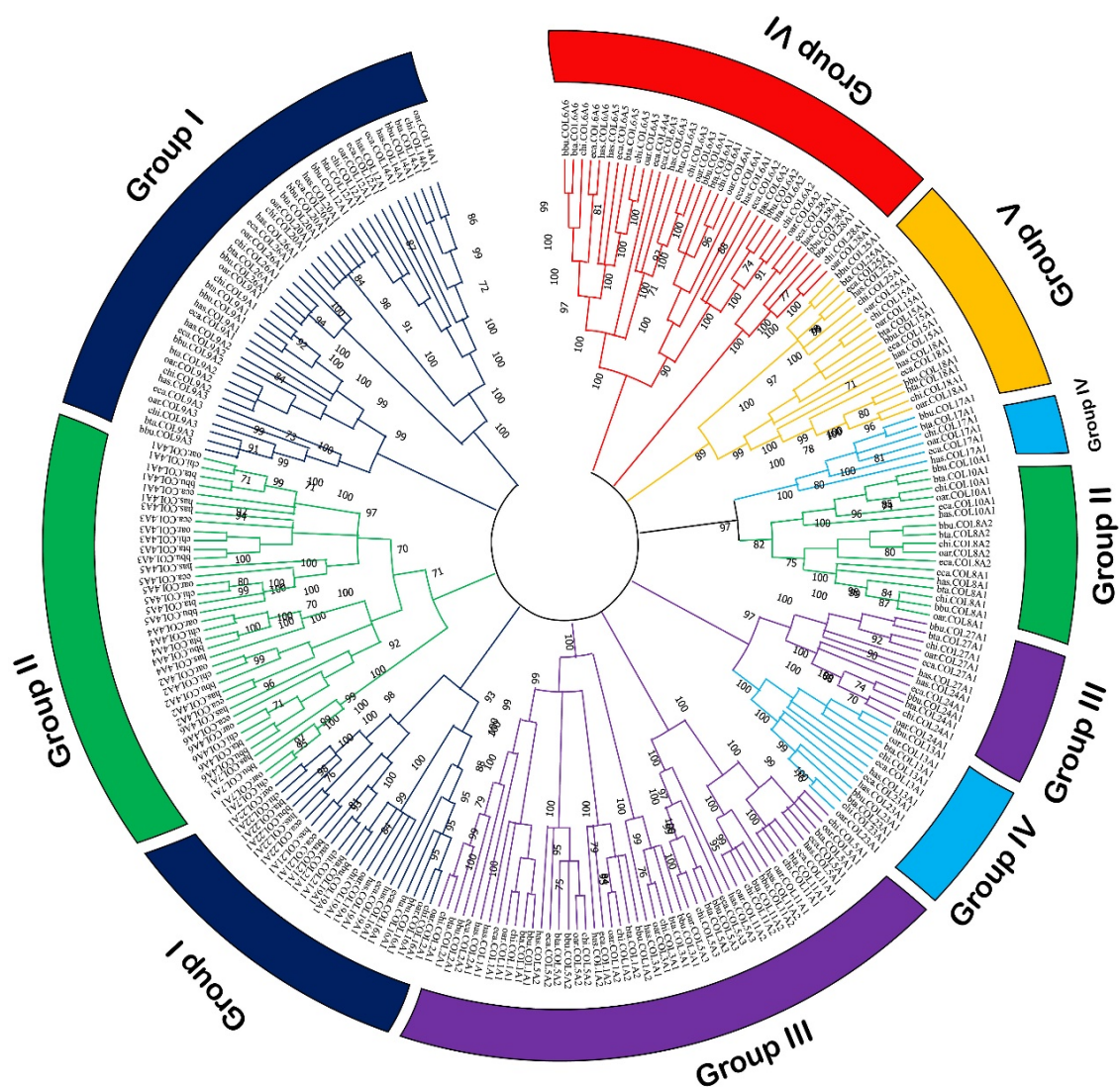
Figure 1. Phylogenetic relationship of collagen proteins in six mammals. Line with different colors indicates different groups. Circle of different color also indicates different groups. Buffalo: bbu; Cattle: bta; Goat: chi; Sheep: oar; Horse: eca; Human: has.

3.2. Sequence Analysis of Buffalo Collagen Genes

To explore the structural characteristics of buffalo collagens, the motifs pattern, gene structures, and conserved domains were performed taking into account their phylogenetic relationships (Figure 2). As showed in Figure 2B, a total of 10 conserved motifs were identified in the identified collagen genes. After the Pfam search, motifs 3 and 5, which both composited 41 amino acids, were annotated as the collagen domain (Table 1). The results were also supported by the identified collagens blasted against the conserved domain database (CDD) from NCBI (Figure 2D). Interestingly, the fibrillar collagen C-terminal domain (COLFI), fibronectin type 3 domain (FN3), C-terminal tandem repeated domain in type 4 procollagen (C4), von Willebrand factor (vWF) type A domain (VWA), von Willebrand factor type C domain (VWC), and EMI domain also was determined in some collagen genes. Moreover, although the introns and UTRs structure varied greatly, gene structural analysis indicated that buffalo collagen genes in the same groups had similar numbers of exon and intron (Figure 2C), suggesting that different collagen groups had different patterns of intron numbers, which verify our previous classification process.
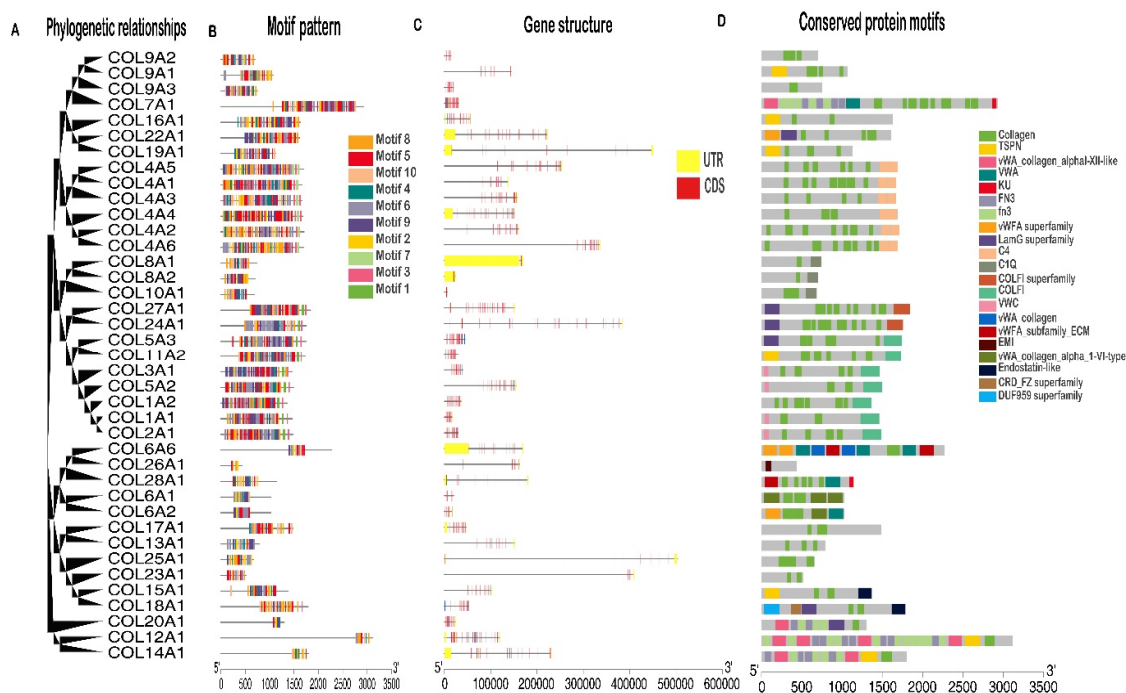
Figure 2. Phylogenetic relationships, motif pattern, gene structure, and conserved protein motifs of buffalo collagen gene family. (**A**) Phylogenetic tree of 45 collagen proteins. (**B**) Motif pattern of buffalo collagen gene family. Ten putative motifs are indicated in different colored boxes. For details of motifs refer to Table 1. (**C**) UTR/CDS organization of collagen genes. Yellow box, black line, and red box represent untranslated region (UTR), intron and coding sequencing (CDS), respectively. (**D**) Distributions of conserved protein motifs in collagen genes.

Table 1. Ten different motifs commonly observed in buffalo collagen family.

Motif	Protein Sequence	Length	Pfam Domain
MEME-1	GLPGLKGEKGEAGLPGFKGEKGVKGEKGE	29	-
MEME-2	KGEDGLPGLPGEKGEKGEKGDPGPPGPPG	29	-
MEME-3	GEKGERGLPGLPGKKGAKGEPGIPGAKGEKGPPGPPGPPGE	41	Collagen
MEME-4	GPPGPPGPPGPPGPPGLPGPPGPPGLPGPP	30	-
MEME-5	PGPPGPKGPRGEKGDPGSTGPPGEPGLPGLQGPPGEKGDKG	41	Collagen
MEME-6	GPKGERGPKGQKGEKGQPGEP	21	-
MEME-7	TGPPGPIGLPGLPGPKGEKGE	21	-
MEME-8	GEPGJPGEKGEPGLPGPPGLPGEKGPKGK	29	-
MEME-9	GEQGERGPKGEKGEA	15	-
MEME-10	RGEPGLPGPPGPPGP	15	-

3.3. Chromosomal Distribution and Collinearity Analysis of Collagen Genes

All identified buffalo collagen genes were randomly distributed on 16 chromosomes (Figure 3A), while the cattle collagens were randomly located on 19 chromosomes (Figure 3B). The majority of buffalo collagen genes were mainly located on the proximate or the distal ends of the chromosomes.
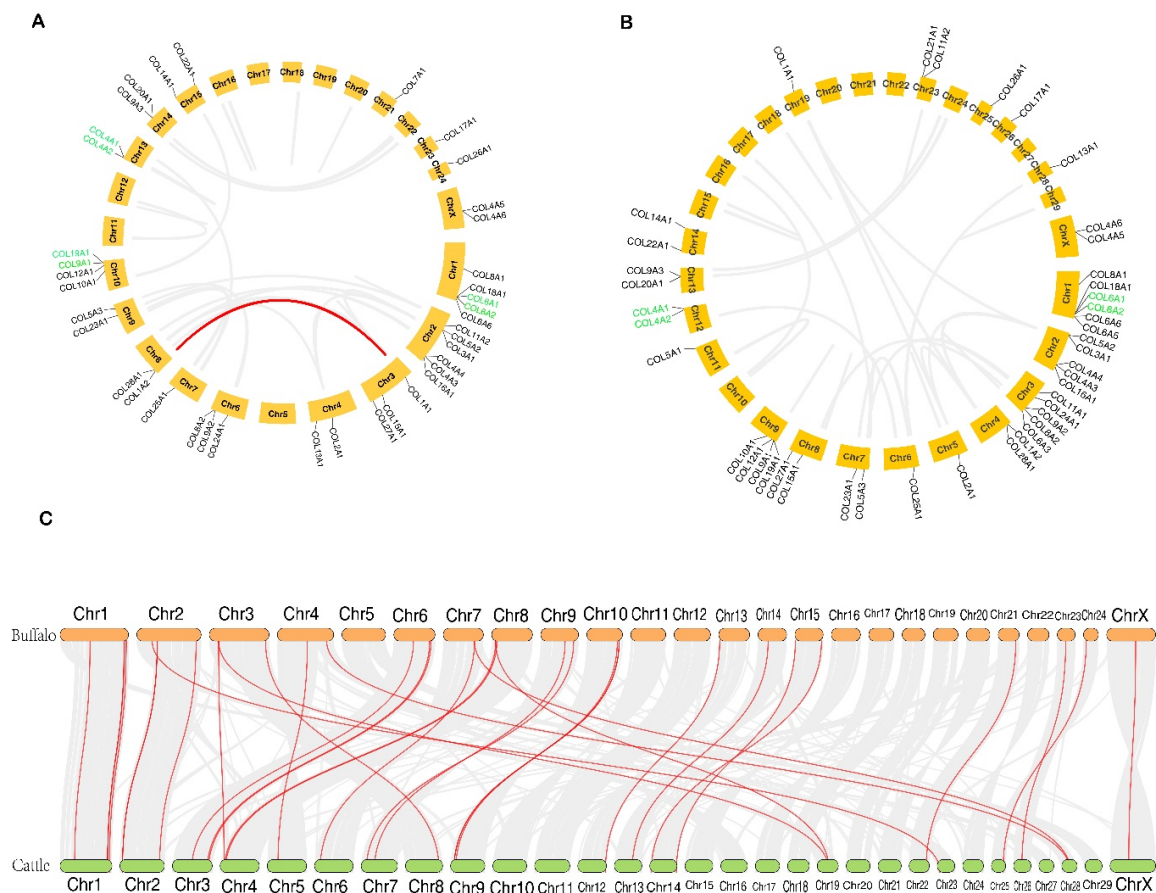
Figure 3. Gene duplication of buffalo (**A**) and cattle (**B**) collagen genes, as well as their collinear analysis (**C**). The tandem duplicated genes were marked by the green color, and segmentally duplicated genes are indicated by the red line.

To investigate the evolutionary progress of the collagen family, we analyzed the duplication events of buffalo and cattle (Figure 3A and B). Among these collagen genes, three pairs of genes, including *COL4A1-COL4A2*, *COL6A1-COL6A2,* and *COL9A1-COL19A1* exhibited the tandem duplication in buffalo. Two pairs of genes (*COL4A1-COL4A2* and *COL6A1-COL6A2*) were seen as tandem duplication genes in cattle. Although only one segmental duplication event (*COL1A1* and *COL1A2*) was discovered in the buffalo collagen family, the scenario was not published in cattle. Positive selection analyses further showed that a total of four duplicated collagen gene pairs in buffalo with the number of nonsynonymous substitutions per nonsynonymous site (Ka)/ the number of synonymous substitutions per synonymous site (Ks) ratios < 1 were identified, three of which had Ka/Ks ratios less than 0.5 (Table 2).

Table 2. Analysis of the Ka/Ks ratios for each pair of duplicated collagen genes in buffalo.

Gene Pairs	CHR	Ka	Ks	Ka/Ks Ratio
COL1A1/COL1A2	3/8	0.2872	1.2576	0.2283
COL4A2/ COL4A1	13	0.5336	1.9459	0.2742
COL6A1/ COL6A2	1	0.7068	1.0659	0.6631
COL9A1/ COL19A1	10	0.6218	2.6177	0.2375

CHR: Chromosome of buffalo; Ka: the number of nonsynonymous substitutions per nonsynonymous site; Ks: the number of synonymous substitutions per synonymous site.

Collinearity analysis between buffalo and cattle genomes showed that a total of 34,381 orthologous genes were determined, which account for 82.21% of the total genes (Figure 3C). Interestingly, a larger

chromosome homologous existed between river buffalo (2*n* = 50) and cattle (2*n* = 60) despite the fact that they have different chromosome numbers. Among the collagen family genes, a total of 36 pairs of collagen genes were orthologous between the two species, and their information was listed in Table S3.

3.4. Comparative Transcriptomic Analyses of Orthologous Collagens between Buffalo and Cattle

Using the RNA-seq data, we further dissect the expression difference of orthologous collagens between buffalo and cattle in milk samples at three lactation points (early, mid, and late lactation). The results showed that a total of 29 collagen genes were found in the milk samples of cattle (Figure 4A) and buffalo (Figure 4B), accounting for 64.44% and 69.05% of the total, respectively. We further found that a total of 23 orthologous collagen genes were detected in milk samples between the two species (Figure 5A). The expression profiles of orthologous collagen genes were more similar in the same species than the expression patterns for the same tissue in different species. The different expression trend of orthologous collagens was also found within the same species. Interestingly, the expression level of the duplicated *COL4A1-COL4A2* genes was higher in buffalo than that of cattle during lactation, while the higher expression level of the *COL6A1-COL6A2* pair was found in cattle compared with that of buffalo. Moreover, our results of qPCR analysis showed that the selected collagen genes in buffalo had similar mRNA expression tendency with that of RNA-seq data (Figure 5B).
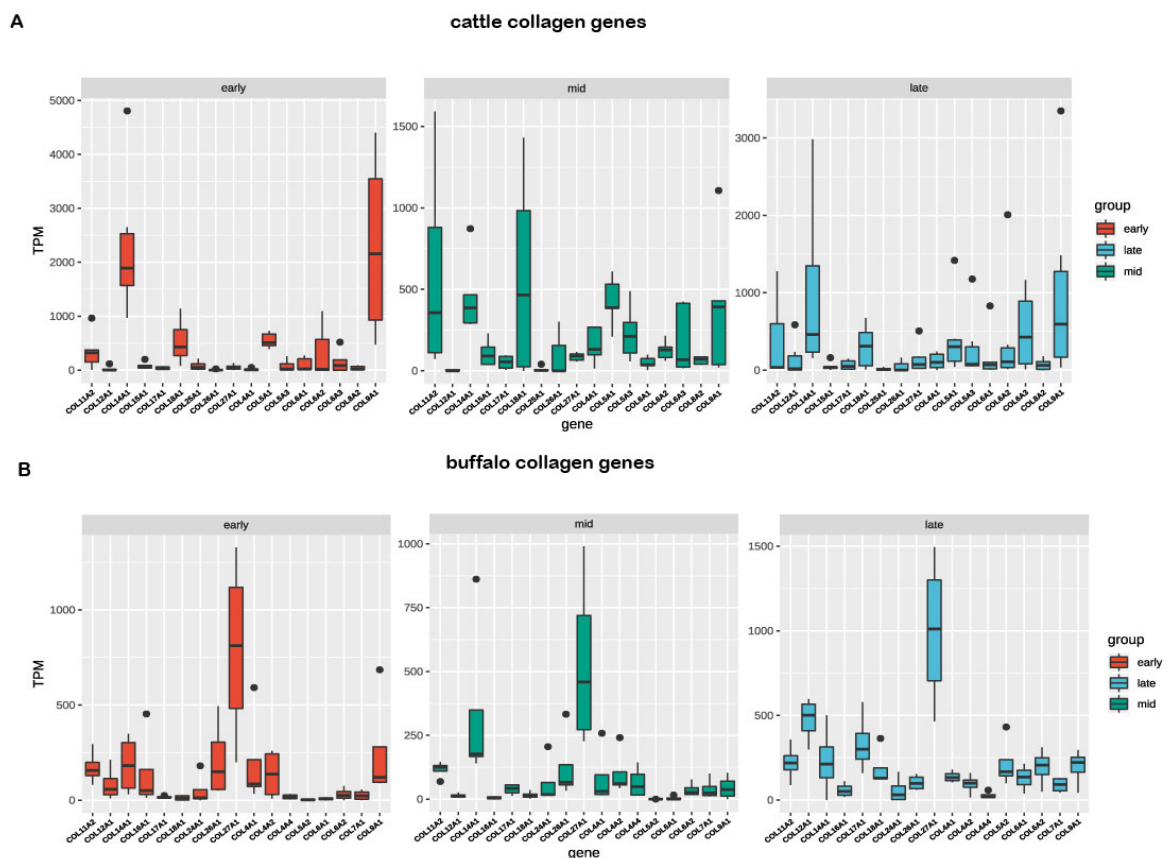

Figure 4. Collagen expression analysis of cattle (**A**) and buffalo (**B**) in milk at different lactations. Red box indicates early lactation; green box indicates mid-lactation; light blue box indicates late lactation.
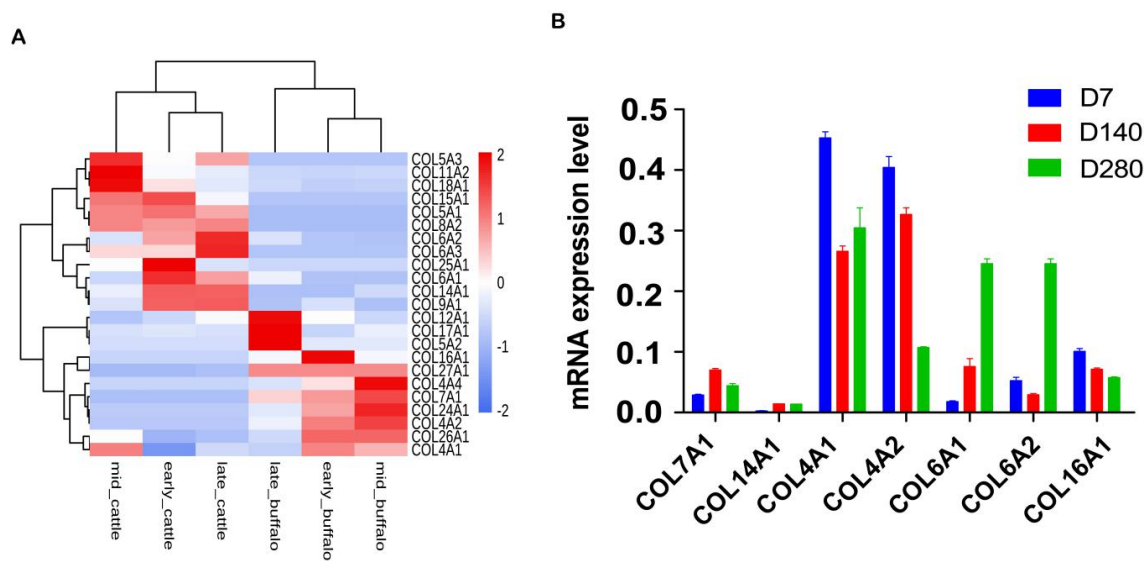
Figure 5. Heat map (**A**) of orthologous collagen genes of cattle and buffalo and validation of selected buffalo collagens by qRT-PCR (**B**). Distances, representing the relative similarity among genes and tissues, were calculated using Pearson's correlation coefficients. Color represents the TPM (transcripts per million) values of gene expression after scaling and centering. D7 indicates early lactation, D140 indicates mid-lactation, and D280 indicates late lactation.

4. Discussion

Collagens, as ECM molecules, support cells for structural integrity and a variety of other functions [32], thus contributing to the support of mammary basic structure and development [33]. Currently, our understanding of the functional role of the buffalo collagen family is limited. In the present study, we identified 128 collagen proteins sequences in buffalo based on its complete genome sequence. The identified collagen protein sequences corresponded to 45 collagen genes in buffalo, which were classified into six groups based on their evolutionary relationships. The result was in line with the previous classification of collagens described by Ricardblum [34]. The conserved motif and gene structure analyses of buffalo collagen genes also supported this classification perspective. Conserved motif analysis indicated that all the identified collagens harbored at least the collagen domain, which is also supported by the previous studies [35–37]. Interestingly, the collagen group I genes contained a fibrillar collagen C-terminal domain (*COLFI*). For the gene structure analysis, most of the collagen genes contained 40 to 60 exons. The intron number analysis revealed that the majority of collagen genes contained more than forty introns, whereas only the *COL8A1*, *COL8A2*, and *COL10A1* contained fewer than three introns. These results are consistent with the exon–intron structure of collagen genes from other representative mammals, suggesting that the collagen family had a conserved gene structure.

Four events during genetic evolution including chromosome doubling, chromosome fragment insertion mutation, tandem duplication, and transposition provide a possibility for the novel gene function acquisition [38,39]. Gene duplication including tandem duplication and segmental duplication can mainly help to accelerate the gene family expansion and genome evolutionary mechanisms [40–42]. In buffalo, all identified collagens were unevenly distributed on 16 chromosomes. Here, a total of three buffalo collagen gene pairs were confirmed to be tandem duplicated genes, but only one segmental duplication event was confirmed to be discovered, which revealed that tandem duplication had a predominant role in the expansion of buffalo collagen family. This finding was supported by Liu et al. [39], who highlighted that segmental duplications in most mammalian lineage are organized in a tandem configuration. For the duplicated collagen pairs, interestingly, we found that four pairs of duplicated collagens with the Ka/Ks ratios < 1 were identified; three of them with the Ka/Ks ratios were less than 0.5, which might experience strong purifying selection pressure.

Previously studies showed that the members of the collagen family are the most abundant proteins in ECM that are tightly regulated throughout the development of the mammary gland [34,42,43]. Therefore, to explore the expression pattern of collagen genes between buffalo and cattle in milk samples at different lactation points is necessary, which contribute to dissect the potential roles of these genes in the milk morphogenesis. In the present study, collinearity analysis revealed that a large number of homologous chromosomal regions were observed between buffalo and cattle. A similar result was also noted in other studies [8,44]. We observed that a total of 36 pairs of collagen genes were orthologous between the two species. For them, a total of 23 orthologous collagens were detected in the milk samples of buffalo and cattle. Notably, they had a more similar expression level in the same species than that of different species, reflecting the hereditary discrepancy between buffalo and cattle concerning mammary gland development and lactation, respectively. Moreover, the different expression trend of orthologous collagens was also found within the same species during lactation, suggesting that these genes had a spatiotemporal expression. Interestingly, the duplicated gene pair of *COL4A1-COL4A2* during lactation was at a higher expressed level than that of cattle, whereas a higher expression level of *COL6A1-COL6A2* pair was found in cattle compared with that of buffalo. Compared with that of buffalo, moreover, a total of 12 collagen genes in cattle had a higher expression level. For them, three (*COL5A3*, *COL11A2*, and *COL18A1*) collagen genes were upregulated in mid-lactation. Two (*COL6A2* and *COL6A3*) collagens were upregulated in late lactation, and two (*COL25A1* and *COL6A1*) were upregulated in early lactation. Dai et al. [45] found that *COL8A1* and *COL1A2* both were upregulated in the bovine mammary gland during lactation compared to the dry period. The reason for the difference could be the limitation of available expression data for cattle collagen and the comparison method. Moreover, a total of 11 buffalo collagens had a higher expression level compared to that of cattle, three (*COL12A1*, *COL17A1*, and *COL5A2*) of which were upregulated in late lactation, as well as *COL16A1* and *COL4A4*, which were respectively upregulated in early and mid-lactation. These results were also supported by the qRT-PCR. Our findings suggested that these collagen genes displayed the specific biology function at different lactation stages. However, these findings have yet to be confirmed.

5. Conclusions

In this work, we performed genomic identification of the buffalo collagen family, with 128 collagen protein sequences confirmed. Next, we found that collagen sequences in buffalo were unequally broadcast on 16 chromosomes. Three tandems and one segmental duplicated gene pair were determined in buffalo, and 36 pairs of collagen genes were orthologous between buffalo and cattle genomes. Moreover, comparative transcription analyses revealed that the highly expressed orthologous collagen genes were different between the buffalo and cattle in the milk samples. The study provides valuable information on the collagen gene family in buffalo affecting mammary gland development and lactation and will assist in determining the collagen gene functions.

References

1. Gelse, K. Collagens—structure, function, and biosynthesis. *Adv. Drug Deliv. Rev.* **2003**, *55*, 1531–1546. [CrossRef] [PubMed]
2. Novaro, V.; Roskelley, C.D.; Bissell, M.J. Collagen-IV and laminin-1 regulate estrogen receptor alpha expression and function in mouse mammary epithelial cells. *J. Cell Sci.* **2003**, *116*, 2975–2986. [CrossRef] [PubMed]
3. Yano, T.; Aso, H.; Sakamoto, K.; Kobayashi, Y.; Hagino, A.; Katoh, K.; Obara, Y. Laminin and collagen IV enhanced casein synthesis in bovine mammary epithelial cells. *J. Anim. Feed. Sci.* **2004**, *13*, 579–582. [CrossRef]
4. Chen, Q.; He, G.; Zhang, W.; Xu, T.; Qi, H.; Li, J.; Zhang, Y.; Gao, M.-Q. Stromal fibroblasts derived from mammary gland of bovine with mastitis display inflammation-specific changes. *Sci. Rep.* **2016**, *6*, 27462. [CrossRef]
5. Crisà, A.; Ferre', F.; Chillemi, G.; Moioli, B. RNA-Sequencing for profiling goat milk transcriptome in colostrum and mature milk. *BMC Veter- Res.* **2016**, *12*, 264. [CrossRef]
6. Anderson, B.M.; MacLennan, M.B.; Hillyer, L.M.; Ma, D.W. Lifelong exposure to n-3 PUFA affects pubertal mammary gland development. *Appl. Physiol. Nutr. Metab.* **2014**, *39*, 699–706. [CrossRef]
7. Inman, J.L.; Robertson, C.; Mott, J.D.; Bissell, M.J. Mammary gland development: Cell fate specification, stem cells and the microenvironment. *Development* **2015**, *142*, 1028–1042. [CrossRef]
8. Mintoo, A.A.; Zhang, H.; Chen, C.; Moniruzzaman, M.; Deng, T.; Anam, M.; Huque, Q.M.E.; Guang, X.; Wang, P.; Zhong, Z.; et al. Draft genome of the river water buffalo. *Ecol. Evol.* **2019**, *9*, 3378–3388. [CrossRef]
9. Hazra, T.; Sharma, V.; Sharma, R.; De, S.; Arora, S.; Lal, D. Detection of cow milk paneer in mixed/buffalo milk paneer through conventional species specific polymerase chain reaction. *Indian J. Anim. Res.* **2016**, 522–528. [CrossRef]
10. Iamartino, D.; Nicolazzi, E.L.; Van Tassell, C.P.; Reecy, J.M.; Fritz-Waters, E.R.; Koltes, J.E.; Biffani, S.; Sonstegard, T.S.; Schroeder, S.G.; Ajmone-Marsan, P.; et al. Design and validation of a 90K SNP genotyping assay for the water buffalo (*Bubalus bubalis*). *PLoS ONE* **2017**, *12*, e0185220. [CrossRef]
11. De Camargo, G.M.F.; Aspilcueta-Borquis, R.R.; Fortes, M.R.; Porto-Neto, L.R.; Cardoso, D.; Santos, D.J.D.A.; Lehnert, S.; Reverter, A.; Moore, S.; Tonhati, H. Prospecting major genes in dairy buffaloes. *BMC Genom.* **2015**, *16*, 872. [CrossRef] [PubMed]
12. Strazzullo, M.; Gasparrini, B.; Neglia, G.; Balestrieri, M.L.; Francioso, R.; Rossetti, C.; Nassa, G.; De Filippo, M.R.; Weisz, A.; Di Francesco, S.; et al. Global transcriptome profiles of Italian Mediterranean buffalo embryos with normal and retarded growth. *PLoS ONE* **2014**, *9*, e90027. [CrossRef] [PubMed]
13. Chen, F.; Fu, Q.; Pu, L.; Zhang, P.; Huang, D.; Hou, Z.; Xu, Z.; Chen, N.; Huang, F.; Deng, T.; et al. Integrated analysis of quantitative proteome and transcriptional profiles reveals the dynamic function of maternally expressed proteins after parthenogenetic activation of buffalo oocyte. *Mol. Cell. Proteom.* **2018**, *17*, 1875–1891. [CrossRef] [PubMed]
14. Low, W.Y.; Tearle, R.; Bickhart, D.; Rosen, B.D.; Kingan, S.B.; Swale, T.; Thibaud-Nissen, F.; Murphy, T.D.; Young, R.; Lefevre, L.; et al. Chromosome-level assembly of the water buffalo genome surpasses human and goat genomes in sequence contiguity. *Nat. Commun.* **2019**, *10*, 260. [CrossRef] [PubMed]
15. National Center for Biotechnology Information. Available online: https://www.ncbi.nlm.nih.gov/genome (accessed on 6 April 2020).
16. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A.; et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2019**, *47*, D427–D432. [CrossRef]
17. Finn, R.D.; Clements, J.; Arndt, W.; Miller, B.L.; Wheeler, T.J.; Schreiber, F.; Bateman, A.; Eddy, S.R. HMMER web server: 2015 update. *Nucleic Acids Res.* **2015**, *43*, W30–W38. [CrossRef] [PubMed]
18. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **2011**, *39*, W29–W37. [CrossRef]
19. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870. [CrossRef]
20. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R.D.; Bairoch, A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **2003**, *31*, 3784–3788. [CrossRef]

21. Chen, C.; Xia, R.; Chen, H.; He, Y. TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. *BioRxiv* **2018**. [CrossRef]

22. Bailey, T.L.; Williams, N.; Misleh, C.; Li, W.W. MEME: Discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* **2006**, *34*, 369–373. [CrossRef] [PubMed]

23. Wang, Y.; Tang, H.; DeBarry, J.; Tan, X.; Li, J.; Wang, X.; Lee, T.-H.; Jin, H.; Marler, B.; Guo, H.; et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [CrossRef] [PubMed]

24. Rozas, J.; Ferrer-Mata, A.; Sã nchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.E.; Sã, N.-G.A. DnaSP 6: DNA sequence polymorphism analysis of large datasets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302. [CrossRef] [PubMed]

25. FastQC. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 6 April 2020).

26. Kim, D.; Langmead, B.; Salzberg, S. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **2015**, *12*, 357–360. [CrossRef] [PubMed]

27. Pertea, M.; Pertea, G.M.; Antonescu, C.M.; Chang, T.-C.; Mendell, J.T.; Salzberg, S. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **2015**, *33*, 290–295. [CrossRef]

28. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Boil.* **2014**, *15*, 31. [CrossRef]

29. Yu, L.; Wang, G.-D.; Ruan, J.; Chen, Y.-B.; Yang, C.-P.; Cao, X.; Wu, H.; Liu, Y.-H.; Du, Z.-L.; Wang, X.-P.; et al. Genomic analysis of snub-nosed monkeys (*Rhinopithecus*) identifies genes and processes related to high-altitude adaptation. *Nat. Genet.* **2016**, *48*, 947–952. [CrossRef]

30. Schmitz, S.; Pfaffl, M.; Meyer, H.H.D.; Bruckmaier, R.M. Short-term changes of mRNA expression of various inflammatory factors and milk proteins in mammary tissue during LPS-induced mastitis. *Domest. Anim. Endocrinol.* **2004**, *26*, 111–126. [CrossRef]

31. Livak, K.J.; Schmittgen, T.D.J.M. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **2001**, *25*, 402–408. [CrossRef]

32. Gordon, M.K.; Hahn, R.A. Collagens. *Cell Tissue Res.* **2010**, *339*, 247. [CrossRef]

33. Hu, G.; Li, L.; Xu, W. Extracellular matrix in mammary gland development and breast cancer progression. *Front. Lab. Med.* **2017**, *1*, 36–39. [CrossRef]

34. Ricard-Blum, S. The collagen family. *Cold Spring Harb. Perspect. Biol.* **2011**, *3*. [CrossRef] [PubMed]

35. Exposito, J.-Y.; Valcourt, U.; Cluzel, C.; Lethias, C. The fibrillar collagen family. *Int. J. Mol. Sci.* **2010**, *11*, 407–426. [CrossRef] [PubMed]

36. Sharma, U.; Carrique, L.; Goff, S.V.-L.; Mariano, N.; Georges, R.-N.; Delolme, F.; Koivunen, P.; Myllyharju, J.; Moali, C.; Aghajari, N.; et al. Structural basis of homo- and heterotrimerization of collagen I. *Nat. Commun.* **2017**, *8*, 14671. [CrossRef]

37. Hughes, A.L. The evolution of functionally novel proteins after gene duplication. *Proc. Biol. Sci.* **1994**, *256*, 119–124.

38. Freeling, M. Bias in plant gene content following different sorts of duplication: Tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Boil.* **2009**, *60*, 433–453. [CrossRef]

39. Liu, G.; Ventura, M.; Cellamare, A.; Chen, L.; Cheng, Z.; Zhu, B.; Li, C.-J.; Song, J.; Eichler, E.E. Analysis of recent segmental duplications in the bovine genome. *BMC Genom.* **2009**, *10*, 571. [CrossRef]

40. Feng, X.; Jiang, J.; Padhi, A.; Ning, C.; Fu, J.; Wang, A.; Mrode, R.; Liu, J.-F. Characterization of genome-wide segmental duplications reveals a common genomic feature of association with immunity among domestic animals. *BMC Genom.* **2017**, *18*, 293. [CrossRef]

41. Zhao, P.; Wang, D.; Wang, R.; Kong, N.; Zhang, C.; Yang, C.; Wu, W.; Ma, H.; Chen, Q. Genome-wide analysis of the potato Hsp20 gene family: Identification, genomic organization and expression profiles in response to heat stress. *BMC Genom.* **2018**, *19*, 61. [CrossRef]

42. Ghajar, C.M.; Bissell, M.J. Extracellular matrix control of mammary gland morphogenesis and tumorigenesis: Insights from imaging. *Histochem. Cell Boil.* **2008**, *130*, 1105–1118. [CrossRef]

43. Zhu, J.; Xiong, G.; Trinkle, C.; Xu, R. Integrated extracellular matrix signaling in mammary gland development and breast cancer progression. *Histol. Histopathol.* **2014**, *29*, 1083–1092. [PubMed]

44. Liu, J.; Wang, Z.; Li, J.; Li, H.; Yang, L. Genome-wide identification of diacylglycerol acyltransferases (DGAT) family genes influencing milk production in buffalo. *BMC Genet.* **2020**, *21*, 26. [CrossRef] [PubMed]

45. Dai, W.; Zou, Y.-X.; Liu, J.; Liu, H.; White, R.R. Transcriptomic profiles of the bovine mammary gland during lactation and the dry period. *Funct. Integr. Genom.* **2017**, *18*, 125–140. [CrossRef] [PubMed]