ORIGINAL RESEARCH PAPER

Sequencing and de novo analysis of the Chinese Sika deer antler-tip transcriptome during the ossification stage using Illumina RNA-Seq technology

Baojin Yao · Yu Zhao · Haishan Zhang · Mei Zhang · Meichen Liu · Hailong Liu · Juan Li

Abstract Deer antlers are the only mammalian appendages capable of repeated rounds of regeneration. Every year, deer antlers are shed and regrown from blastema into large branched structures of cartilage and bone. Little is known about the genes involved in antler development particularly during the later stages of ossification. We have produced more than 39 million sequencing reads in a single run using the Illumina sequencing platform. These were assembled into 138,642 unique sequences (mean size: 405 bp) representing 50 times the number of Sika deer sequences previously available in the NCBI database (as of Nov 2, 2011). Based on a similarity search of a database of known proteins, we identified 43,937 sequences with a cut-off E-value of 10^{-5}.

Assembled sequences were annotated using Gene Ontology terms, Clusters of Orthologous Groups classifications and Kyoto Encyclopedia of Genes and Genomes pathways. A number of highly expressed genes involved in the regulation of Sika deer antler ossification, including growth factors, transcription factors and extracellular matrix components were found. This is the most comprehensive sequence resource available for the deer antler and provides a basis for the molecular genetics and functional genomics of deer antler.

Keywords Antler tip · Chinese Sika deer · Gene expression · Ossification · Transcriptome

B. Yao · J. Li (✉)
School of Public Health, Jilin University,
Changchun 130021, China
e-mail: li_juan@jlu.edu.cn

B. Yao · Y. Zhao (✉) · M. Zhang · M. Liu · H. Liu
Center for New Medicine Research,
Changchun University of Chinese Medicine,
Changchun 130117, China
e-mail: cnzhaoyu@yahoo.com.cn

H. Zhang
Department of Colorectal Surgery, China-Japan Union
Hospital, Jilin University, Changchun 130033, China

Introduction

Deer antlers serve as useful models of bone growth and mineralization in mammals (Szuwart et al. 2002). In the spring, the nascent antler regenerates from a permanent cranial bony protuberance, which is known as a pedicle, following casting of the previous hard antler. Rapid growth and bifurcation occur in summer after which the regenerating antler attains full size and becomes completely calcified. Development culminates with the shedding of the velvet skin in autumn. In the winter, a bare bony antler is firmly attached to its living pedicle and is not cast off until the following spring, when its casting triggers another round of antler regeneration (Li et al. 2009).

The longitudinal growth of deer antler occurs by a process of modified endochondral ossification (Price and Allen 2004). During endochondral ossification, chondrocytes proliferate, undergo hypertrophy and die; the cartilage extracellular matrix they construct is subsequently invaded by blood vessels, osteoclasts, bone marrow cells and osteoblasts, the last of which deposit bone on the remnants of the cartilage matrix (Mackie et al. 2008). Comprehensive analyses of in vivo gene expression profiles that underlie these processes are needed in order to obtain a complete understanding of the mechanisms that convert deer antlers from cartilage into bone.

Next-generation sequencing technology has emerged as a powerful tool for high-throughput sequence determination and has dramatically improved the efficiency and speed of gene discovery (Metzker 2009). The development of novel high-throughput DNA sequencing methods has provided a new means of both mapping and quantifying transcriptomes. This method, known as RNA-Seq (RNA sequencing), has clear advantages over existing approaches and is expected to revolutionize the manner in which eukaryotic transcriptomes are analyzed (Wang et al. 2009). Despite their obvious benefits, next-generation sequencing methods have not yet been applied to Sika deer antler research.

In this study, we generated over three billion bases of high-quality cDNA sequence using Illumina technology, and we demonstrated the suitability of short-read sequencing for the de novo assembly and annotation of genes expressed in the growing tip of the ossifying Sika deer antler. After assembling these reads into 138,642 unique transcripts, we searched the annotated sequences for genes associated with GO classifications, COG classifications and KEGG pathways. We also analyzed the gene expression patterns that are involved in endochondral ossification. These sequencing results and analyses provide an invaluable resource for further research on the mechanisms of deer antler development.

Materials and methods

Sample collection and preparation

Antler tips were collected from three anaesthetized 4-year-old Chinese Sika deers during he ossification stage (after 90 days of growth) at the Shuangyang deer farm in Changchun, China. All experimental procedures were approved by the Animal Ethics Committee of Jilin University. The velvet skin was removed and the rest of the tissues were cut into small pieces and immediately stored in liquid nitrogen until further processing.

RNA isolation and library preparation for transcriptome analysis

Total RNA was isolated using TRIzol reagent (Invitrogen) according to the manufacturer's instructions. RNA integrity (Schroeder et al. 2006) was confirmed using the 2100 Bioanalyzer (Agilent Technologies) with a minimum RNA integrity number (RIN) of 8. The samples for transcriptome analysis were prepared using the Illumina kit following the manufacturer's recommendations. Samples of mRNA were purified from 10 μg total RNA using oligo-(dT) magnetic beads. Following purification, the mRNA was fragmented into small pieces using divalent cations at an elevated temperature. The cleaved RNA fragments were used for first-strand cDNA synthesis using reverse transcriptase and random primers. This reaction was followed by second-strand cDNA synthesis using DNA polymerase I and RNaseH after which end repair and adapter ligation were conducted on the cDNA fragments. These products were purified and enriched with PCR to create the final cDNA library.

Sequencing, assembly and gene expression analysis

The cDNA library was sequenced on the Illumina sequencing platform (HiSeq 2000). The average size of inserts in the library was 200 bp, and both ends of each fragment were sequenced. Image deconvolution and quality value calculations were performed using the Illumina GA pipeline 1.3. The raw reads were cleaned by removing adaptor sequences, empty reads low-quality sequences (reads with unknown 'N' sequences), and reads with more than 10% $Q < 20$ bases. The reads obtained were randomly clipped into different K-mers for assembly using de Bruijn graph and SOAPdenovo software (Li et al. 2010). By testing different K-mer sizes (21, 25, 29, 33, 37, 41, 45, 49, 53, and 57), we found that the 29-mer provided the best results for transcriptome assembly (Fig. S1). After

sequence assembly, the resulting contigs were joined into scaffolds using the read mate pairs. To obtain distinct gene sequences, the scaffolds were clustered using TGI Clustering tools (Lee et al. 2005).

To further verify the sequencing accuracy and completeness, the BLASTN program was used to compare our sequences with the ESTs in the Sika deer EST database at NCBI (http://www.ncbi.nlm.nih.gov/nucest?term=%20Cervus%20nippon), and results with E-values equal to or less than 10^{-10} were treated as "significant matches", whereas ESTs with no hits, or matches with E-values greater than 10^{-10} to nucleotides in NCBI were classified as "no significant matches".

Gene expression levels were expressed as RPKM (reads per kilobase of exon model per million mapped reads) (Mortazavi et al. 2008) using the formula RPKM $= 10^9 C/NL$, where C is the number of mappable reads that uniquely align to a unigene, N is the total number of mappable reads that uniquely align to all unigenes, and L is the length of a unigene in base pairs.

Functional annotation by sequence comparison with public databases

BLASTX alignment with an E-value cut-off of 10^{-5} between unigenes and protein databases (non-redundant) nr, Swiss-Prot, Kyoto Encyclopedia of Genes and Genomes (KEGG) and Clusters of Orthologous Groups of proteins (COG) is performed, and the best aligning results were used to decide sequence direction of unigenes. If results of different databases conflict with each other, a priority order of nr, Swiss-Prot, KEGG and COG was followed when deciding sequence direction of unigenes. When a unigene happens to be unaligned to none of the above databases, the software ESTScan (Iseli et al. 1999) was introduced to predict its coding regions as well as to decide its sequence direction. Functional annotation with Gene Ontology (GO) terms was performed using Blast2go (Conesa et al. 2005) and the GO database (Harris et al. 2004) with an E-value cut-off of 10^{-5}. After obtaining the GO annotation for every unigene, we use WEGO software (Ye et al. 2006) to do GO functional classification for all unigenes and to understand the distribution of gene functions of the species from the macro level. Annotation with COGs, and KEGG pathways was performed by searching with BLASTX against the COG database (Tatusov et al. 2001) and the KEGG database (Kanehisa et al. 2004) with an E-value cut-off of 10^{-5}.

Real-time PCR analysis

In order to confirm the transcriptome data, a qPCR assay was performed using the One Step SYBR PrimeScript PLUS RT-PCR Kit (Takara, Japan) and the BioRad Chromo4 real-time PCR system. Results were analyzed with Opticon Monitor software version 3.1. Total RNA were treated by DNase I (Takara, Japan). Reaction was performed in 25 µl containing 2 µl total RNA (50 ng), 1 µl primer mix containing 10 µM of each forward and reverse gene specific primers, 1.5 µl *Ex Taq* HS Mix, 0.5 µl PrimeScript PLUS RTase Mix, 12.5 µl 2× One Step SYBR RT-PCR Buffer 4 and 6.5 µl RNAase free H_2O. PCR amplification used: 42°C for 5 min and 95°C for 10 s, followed by 40 cycles of 95°C for 5 s and 60°C for 30 min. The specificity of the individual PCR amplification was checked using a heat dissociation protocol from 60 to 95°C following the final cycle of the PCR and agarose gel electrophoresis. All reactions were performed in triplicate, the gene expressions of target genes were normalized against an internal reference gene, 60S ribosomal protein L40. The relative gene expression was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001).

Results

Illumina sequencing, assembly and sequence analysis

To obtain an overview of the antler tip gene expression profile during ossification, a cDNA sample was prepared from the antler tip and was sequenced using the Illumina sequencing platform. After cleaning and quality checks, we obtained ~ 39 million clean reads, each of 90 bp. Over 96% of the clean reads had quality scores at the Q20 level (a base quality greater than 20 and an error probability of 0.01). The data sets were deposited in the NCBI ArrayExpress repository with the accession number E-MTAB-853.

De novo assembly of the clean reads was performed using SOAPdenovo (http://soap.genomics.org.cn/soapdenovo.html) with the default settings, except for the

Table 1 Overview of the sequencing and assembly

Feature	Statistic
Total number of reads	39,400,004
Total base pairs (bp)	3,546,000,360
Average read length (bp)	90
Total number of contigs	1,129,708
Mean length of contigs (bp)	127
N50 of contigs (bp)	90
Total number of scaffolds	187,866
Mean length of scaffolds	331
N50 of scaffolds (bp)	423
Total number of unigenes	138,642
Mean length of unigenes (bp)	405
N50 of unigenes (bp)	487

K-mer value, which was set at a specific value (Li et al. 2010). By testing different K-mer sizes (21, 25, 29, 33, 37, 41, 45, 49, 53, and 57), we found that the 29-mer provided the best results for transcriptome assembly. As shown in Supplementary Fig. 1, K29 can produce much longer unigenes compared to any others. So the best assembly was achieved with K = 29, which was chosen for de Bruijn graph construction. These short 29-mers were assembled, resulting in 1,129,708 contigs with an N50 of 90 bp (50% of the assembled bases were incorporated into contigs 90 bp or longer) (Table 1). The mean contig size was 127 bp, with lengths ranging from as short as 50 bp to as long as 7,187 bp. Using paired-end joining and gap-filling, the contigs were further assembled into 187,866 scaffolds with a mean size of 331 bp and an N50 of 423 bp, including 9,003 scaffolds larger than 1,000 bp (Table 1). After clustering using TGICL software (Lee et al. 2005), the 187,866 scaffolds generated 138,642 distinct sequences (unigenes) with a mean size of 405 bp and an N50 of 487 bp (Table 1).

To evaluate the quality of the dataset, we analyzed the ratio of the gap's length to the length of assembled unigenes (Supplementary Fig. 2). The majority of the unigenes showed gap lengths that were less than 5% of the total length, which accounted for 92.6% of total unigenes numbers (128,359 unigenes). In addition, sequencing bias was analyzed by detecting random distribution of reads in assembled unigenes (Supplementary Fig. 3). Although the 3′-ends of all assembled unigenes contained relatively fewer numbers of reads, other positions of all assembled unigenes showed greater and more even distribution. Comparisons with Sika deer EST database using BLASTN showed that 2,018 out of 2,749 ESTs in the Sika deer EST database had significant matches (E-value $\leq 10^{-10}$) to sequences produced by our transcriptome sequencing.

Annotation of predicted proteins

Distinct gene sequences were first searched using BLASTX against the NCBI nr database using a cut-off E-value of 10^{-5}. Using this approach, 43,937 genes (31.7% of all distinct sequences) returned a BLAST result above the cut-off value. Due to lack of genome and EST information in Sika deer, 68.3% of unigenes could not be matched to known genes. Similarly, up to 98,778 unigenes (71.3% of all unigenes) had no Swissprot annotation.

Functional annotation of the transcriptome

Based on GO classifications, 22,743 sequences were categorized into 48 functional groups (Fig. 1). In the three GO ontologies (biological process, cellular component and molecular function), "cell part" (19,245 members), "binding" (17,770 members) and "cellular process" (15,930 members) were, respectively, the most frequently used terms. We also noticed a high percentage of genes from the "organelle" (13,472 members) and "metabolic process" (12,463 members) categories. Few genes were found in the categories of "electron carrier activity" (4 members) and "metallochaperone activity" (3 members) (Fig. 1). Moreover, 177 unigenes were annotated with the category of "ossification", including positive regulation of ossification (170 members) and negative regulation of ossification (7 members). 56 unigenes were annotated with the category of "chondrocyte differentiation".

We also searched the annotated sequences for genes that are part of COG groups. In total, out of 43,937 nr hits, 9,908 sequences had a COG classification (Fig. 2). Among the 25 COG categories, "general function prediction only" represents the largest group (3,418 members) followed by "translation, ribosomal structure and biogenesis" (2,057 members), "replication, recombination and repair" (1,769 members) and "transcription" (1,518 members). The "extracellular structure" (8 members) and "nuclear structure" (7 members) categories were found to be the smallest
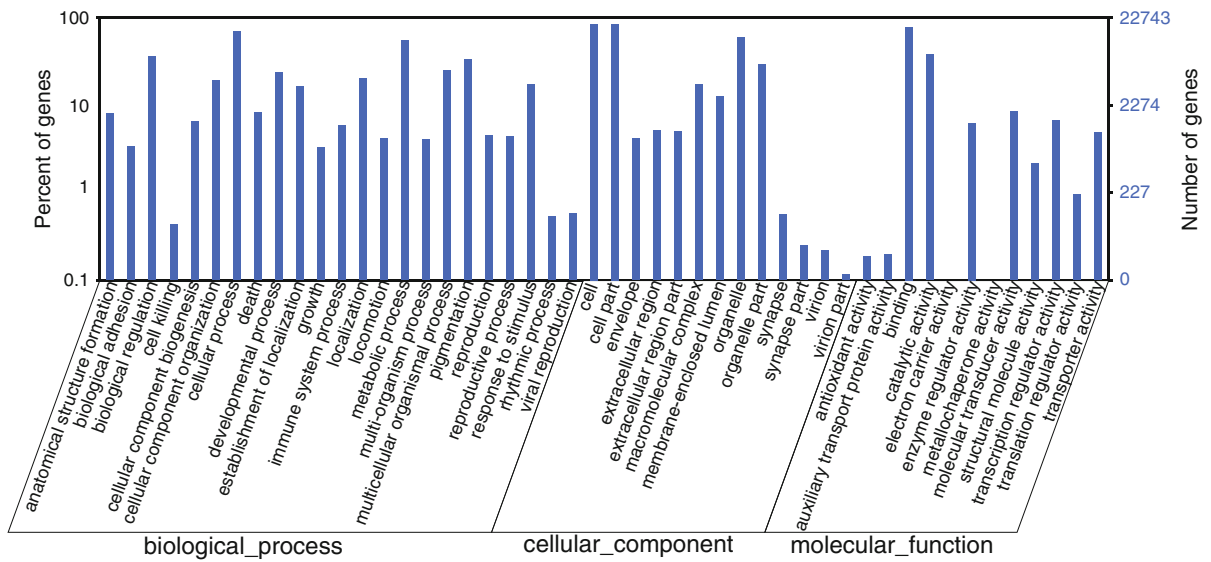
Fig. 1 Histogram presentation of Gene Ontology classification. The results are summarized in three main categories: biological process, cellular component and molecular function. The right y-axis indicates the number of genes in a category. The left y-axis indicates the percentage of a specific category of genes in that main category

groups (Fig. 2). We also mapped the 43,937 annotated sequences to the reference canonical pathways in the KEGG database. In total, we assigned 27,356 sequences to 218 KEGG pathways (Supplementary Table 1). The pathways with the most representation among the unique sequences were the metabolic pathways (2,804 members) followed by those related to focal adhesion (1,601 members), regulation of the actin cytoskeleton (1,397 members) and pathways in cancer (1,313 members).

Highly expressed genes involved in endochondral ossification

We identified a number of highly expressed genes related to endochondral ossification, including growth factors, transcription factors and extracellular matrix components. The results are summarized in Table 2. The most highly expressed genes were those encoding members of the collagen family, including collagen types I, V, XII, IX, X and XI, followed by other extracellular matrix proteins, adhesion molecules and matrix metalloproteinases, such as matrix Gla protein, cartilage link protein 1, integrin beta-1, cadherin-11 and matrix metalloproteinases 13 and 14. Signaling molecules, as well as osteoblast and osteclast differentiation markers, were found to be enriched in the

transcriptome of the ossifying antler. We randomly selected 9 genes and designed 9 pairs of primers (Supplementary Table 2) for qPCR validation. The results showed consistent expression patterns with the observed findings in transcriptome analysis (Fig. S4). Their significant agreement indicated that the abundance of the Illumina sequences from the Sika deer transcriptom closely mirrors the actual expression level.

Discussion

Illumina RNA-Seq technology is an efficient and cost-effective method for discovering novel genes and investigating gene expression patterns, especially in non-model organisms that do not have sequenced genomes (Shi et al. 2011). We used high-throughput sequencing data to characterize the antler-tip transcriptome of Chinese Sika deer during the process of ossification. In this study, we used the Illumina sequencing platform to produce 138,642 high-quality sequences with a mean length of 405 bp and an N50 fo 487 bp (Table 1). To evaluate the quality of the dataset, we analyzed sequence quality, sequencing bias and gap lengths. This results is consistent with previous publications (Shi et al. 2011; Wang et al. 2010),
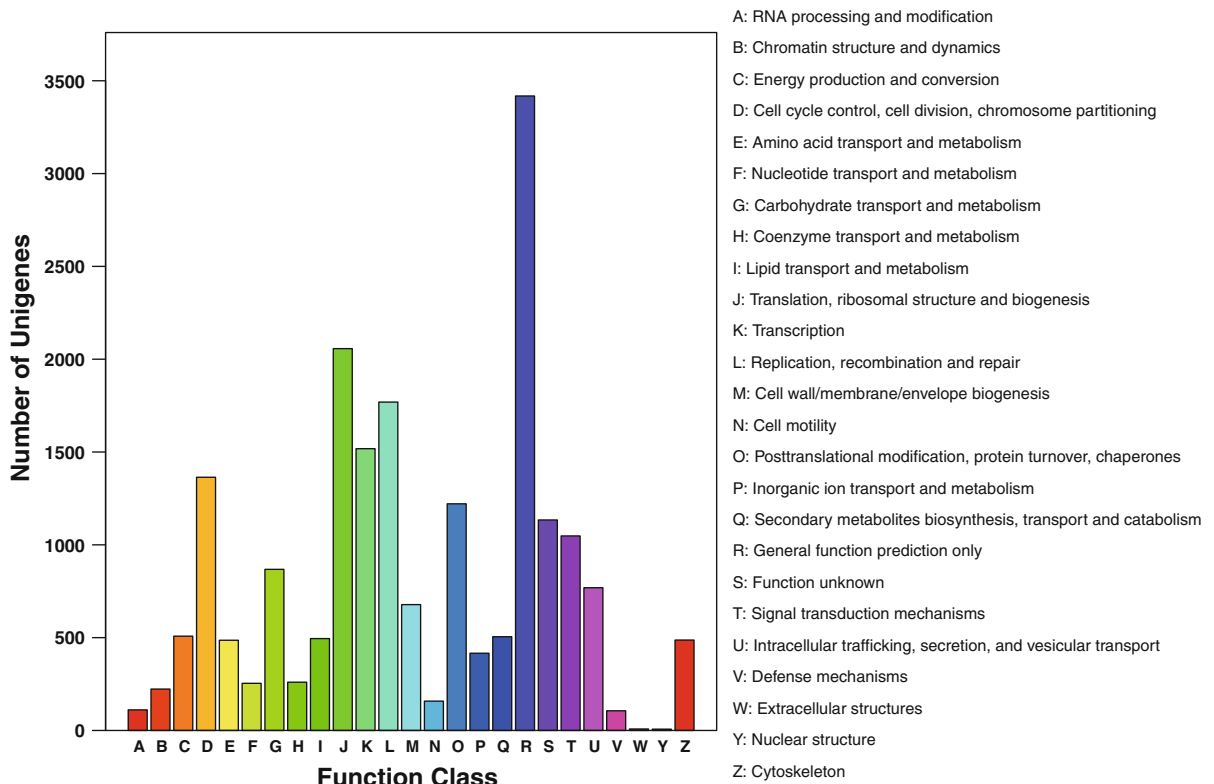
A: RNA processing and modification
B: Chromatin structure and dynamics
C: Energy production and conversion
D: Cell cycle control, cell division, chromosome partitioning
E: Amino acid transport and metabolism
F: Nucleotide transport and metabolism
G: Carbohydrate transport and metabolism
H: Coenzyme transport and metabolism
I: Lipid transport and metabolism
J: Translation, ribosomal structure and biogenesis
K: Transcription
L: Replication, recombination and repair
M: Cell wall/membrane/envelope biogenesis
N: Cell motility
O: Posttranslational modification, protein turnover, chaperones
P: Inorganic ion transport and metabolism
Q: Secondary metabolites biosynthesis, transport and catabolism
R: General function prediction only
S: Function unknown
T: Signal transduction mechanisms
U: Intracellular trafficking, secretion, and vesicular transport
V: Defense mechanisms
W: Extracellular structures
Y: Nuclear structure
Z: Cytoskeleton

Fig. 2 Histogram presentation of clusters of orthologous groups (COG) classification. The histogram shows the distribution of sequences among different COG categories: Out of 43,937 nr hits, 9,908 sequences have a COG classification among the 25 categories

suggesting that the quality of our dataset was comparable to similar reports in other non-model species. Comparisons with Sika deer EST database using BLASTN revealed that our transcriptome dataset had both a high degree of consistency (73.40%) with previous EST data and an approximate 50 times increase in coverage. However, because of the relatively short length of distinct gene sequences (mean size of 405 bp) and the lack of genome information for Sika deer, most of the 138,642 assembled sequences (68.3%) could not be matched to known genes.

To further evaluate the completeness of our transcriptome library and the effectiveness of our annotation process, we searched the annotated sequences for genes associated with GO terms, COG classifications and KEGG pathways. In total, we assigned 22,743 sequences to 48 GO terms, 9,908 sequences to 25 COG classifications and 27,356 sequences to 218 KEGG pathways. These annotations provide a valuable resource for investigating specific processes,

functions and pathways involved in Sika deer antler development.

Deer antlers grow by a process of modified endochondral ossification. During endochondral ossification, bones initially develop as cartilage templates in which chondrocytes proliferate, differentiate into hypertrophic chondrocytes and are gradually replaced by bone (Li and Suttie 1994; Hartmann 2009). Based on de novo sequencing and an in-depth analysis of the transcriptome, we found 31 transcripts related to endochondral ossification that were highly expressed, including growth factors, transcription factors and extracellular matrix components (Table 2). Collagen I, which was the most highly expressed transcript, is an essential protein in endochondral bone. It was detected only in the perichondrium, in vascular cavities, and in osteoid and bone, and it was absent from the hypertrophic cartilage matrix (von der Mark 1977; Kirsch and von der Mark 1992). Matrix Gla protein was the next most highly expressed gene followed by collagen

Table 2 List of highly expressed genes involved in endochondral ossification

Gene name	RPKM[a]
Collagen I	2,061.24
Matrix Gla protein	1,066.72
Collagen V	885.34
Collagen XII	832.96
Integrin beta-1	791.60
Cartilage link protein 1	422.01
Collagen IX	402.45
Matrix metalloproteinase-13	332.05
Cadherin-11	327.70
Collagen X	296.64
Aggrecan core protein	256.35
Alkaline phosphatase, tissue-nonspecific isozyme	249.26
Osteopontin	247.33
Matrix metalloproteinase-14	231.10
Osteoclast-stimulating factor 1	230.84
Cartilage oligomeric matrix protein	228.72
Osteoglycin	224.97
Cartilage-associated protein	220.07
Collagen XI	213.67
Osterix	168.13
Cathepsin K	162.65
Transforming growth factor beta-1	158.96
Osteomodulin	156.51
Indian hedgehog	153.37
Bone morphogenetic protein-1	153.28
Osteocalcin	151.48
Fibroblast growth factor receptor 3	146.70
Connective tissue growth factor	106.72
Runt-related transcription factor 2	105.86
Matrix metalloproteinase-9	103.91
Osteoprotegerin	102.33

[a] Refers to reads per kilobase of exon model per million mapped reads

V and collagen XII. Matrix Gla protein (MGP) is a 14-kD extracellular matrix protein of the mineral-binding Gla protein family (Price and Williamson 1985). The MGP gene is widely expressed, but the protein only accumulates significantly in bone, cartilage, and dentin (Hale et al. 1988). Coordinated expression of matrix Gla protein is required for chondrocyte survival during endochondral ossification (Newman et al. 2001). Collagen V, a member of the fibril subclass of collagens, plays important roles in the regulation of collagen fibrillogenesis and matrix assembly (Wenstrup et al. 2004). In normal bone, only two types of collagen are consistently present, types I and V. Type V collagen is copolymerized with type I collagen in the extracellular matrix of bone (Niyibizi and Eyre 1994). Collagen XII is a member of the subfamily of fibril-associated collagens with interrupted triple helices (FACITs). FACITs are associated with striated fibrils assembled from fibrillar collagens, such as type I collagen. Collagen XII also plays important regulatory roles in osteoblast differentiation and bone matrix formation. Genetic deletion of type XII collagen results in abnormal osteoblast differentiation, decreased bone matrix deposition, and decreased bone quality. This phenotype results from impaired terminal differentiation of osteoblasts (Izu et al. 2011). Like the aforementioned collagens, several other collagens, including collagen types IX, X and XI, were also highly expressed. Collagens IX and XI belongs to a tissue-specific set of collagens unique to cartilage (Eyre 2002). Collagen X is the only known hypertrophic chondrocyte–specific molecular marker. Collagen X is a direct transcriptional target of Runx2, which is an essential transcription factor for endochondral ossification (Zheng et al. 2003).

Integrins and cadherins are two major classes of cell adhesion molecules that mediate cell–matrix interactions. Integrins are composed of α and β subunits; they bind extracellular matrix (ECM) proteins and counter-receptors, and they regulate various biological processes, including the supramolecular assembly of ECM proteins, adhesion and migration of cells, cell cycle progression, cell survival, and differentiation. $\beta1$ integrins are highly expressed in chondrocytes, where they mediate adhesion to cartilage matrix proteins. $\beta1$-integrin-dependent motility and proliferation of chondrocytes are necessary events for the formation of endochondral bone (Aszodi et al. 2003). Cadherin-11, a type II classical cadherin from the cadherin superfamily, is specifically involved in osteoblast commitment and differentiation. Cadherin-11 is also an essential factor involved in the process of chondrocyte calcification (Benedetto et al. 2010; Matsusaki et al. 2006). Several metalloproteinases, including matrix metalloproteinase-9 (MMP9), matrix metalloproteinase-13 (MMP13), matrix metalloproteinase-14 (MMP14) and bone morphogenetic protein-1 (BMP-1), were also highly expressed in the transcriptome. MMP9, MMP13 and MMP14, three

members of the MMP family, are able to cleave a variety of substrates, including ECM proteins, extracellular non-ECM proteins and cell-surface proteins. Protease activity is present in developing bone and all three of the MMPs listed above are essential for normal endochondral ossification (Ortega et al. 2004). BMP-1, the most highly expressed signaling molecule, belongs to the peptidase M12A family of bone morphogenetic proteins (BMPs); unlike other BMPs, BMP-1 does not belong to the TGF-β superfamily. BMP-1 was initially discovered to work similarly to other BMPs by inducing bone and cartilage development. BMP-1, however, is a metalloproteinase that processes a wide variety of extracellular matrix pro-proteins, most notably collagen types I, II, III, V, and VII, lysyl oxidase, laminin 5, probiglycan and dentin matrix protein (Trackman 2005). A variety of genetic markers are predominantly expressed in osteoblasts and osteoclasts, including alkaline phosphatase, osteopontin, osteoclast-stimulating factor 1, osteoglycin, osterix, cathepsin K, osteomodulin, osteocalcin and osteoprotegerin. These genes were also found to be highly expressed, which makes sense because osteoblasts and osteoclasts play important roles in bone formation through cell–cell contact (Matsuo and Irie 2008).

Several important signaling molecules were enriched in the transcriptome, including transforming growth factor beta-1 (TGFβ1), indian hedgehog (Ihh), fibroblast growth factor receptor 3 (Fgfr3), runt-related transcription factor 2 (Runx2) and connective tissue growth factor (CTGF). The TGF-βs, including TGF-β1, TGF-β2 and TGF-β3, regulate a large variety of cellular activities. TGF-β1 was found in the proliferative and hypertrophic chondrocytes; it enhances chondrocyte proliferation but inhibits the terminal differentiation of chondrocytes, and it helps chondrocytes remain in the prehypertrophic stage (Li et al. 2005). Ihh is a member of the hedgehog (Hh) family that plays crucial roles in regulating many developmental processes. Apart from controlling chondrocyte hypertrophy indirectly through regulating PTHrP expression, Ihh signaling also directly regulates chondrocyte hypertrophy in the absence of PTHrP (Mak et al. 2008). Fgfr3 is one of four fibroblast growth factor receptors (Fgfrs), and it is part of a family of tyrosine kinase receptors. FGF signaling has a major role in a variety of developmental processes, especially bone morphogenesis. Fgfr3-mediated STAT1-p21 signal transduction induces

apoptosis in hypertrophic chondrocytes and modulates the replacement of cartilage with bone. These mechanisms may regulate chondrogenesis and promote endochondral ossification (Nakajima et al. 2003). Runx2 is a member of a family of transcription factors that share a DNA-binding domain with the *Drosophila* pair rule gene *runt*. Runx2 plays an important role in endochondral ossification: it drives proliferative chondrocytes to differentiate further into hypertrophic chondrocytes (Hinoi et al. 2006). Runx2 also activates the Ihh promoter and stimulates Ihh expression (Yoshida et al. 2004). CTGF is a secreted protein containing several domains that mediate interactions with growth factors, integrins and extracellular matrix components. CTGF is important for cell proliferation and matrix remodeling during chondrogenesis, and it is a key regulator linking the remodeling of the extracellular matrix to angiogenesis at the growth plate. A role for CTGF in extracellular matrix production is suggested by its ability to mediate collagen deposition during wound healing (Ivkovic 2003).

Many cartilage matrix proteins were also highly expressed, including cartilage link protein 1 (CRTL1), aggrecan core protein (ACAN), cartilage oligomeric matrix protein (COMP) and cartilage-associated protein (CRTAP). CRTL1 is a member of the hyaluronan and proteoglycan binding link protein (HAPLN) gene family, and it is also known as HAPLN1. It plays an essential role in endochondral bone formation by stabilizing supramolecular assemblies of aggrecan and hyaluronan (Deák et al. 1999). ACAN and COMP are two major components of the ECM of the musculoskeletal system. COMP in its native conformation can bind ACAN, and this interaction is regulated by the calcium-sensitive conformation of COMP (Chen et al. 2007). COMP also neutralizes ECM1-mediated inhibition of chondrocyte hypertrophy, matrix mineralization, and endochondral bone formation. COMP and ECM1 play opposite roles in regulating endochondral ossification, and that balanced regulation depends on their interaction (Kong et al. 2010). CRTAP is an essential cofactor for proper post-translational chain modification and collagen folding. CRTAP is required for prolyl 3-hydroxylation, and mutations in CRTAP cause recessive osteogenesis imperfecta, a connective-tissue disorder characterized by bone fragility and low bone mass (Morello et al. 2006).

According to our results and analysis, when Sika deer antlers grow for 90 days, they show a significant

increase in mineralization throughout the antler tip, as indicated by a visible sign of chondrocyte hypertrophy and large populations of invading osteoblasts, endothelial cells and osteoclasts. This finding was confirmed by the expression of a large number of putative molecular markers and signaling molecules related to these physiological processes. Endochondral ossification of the Sika deer antler is a complex process that requires the coordinated activity of transcription factors, signaling molecules, extracellular matrix proteins and other factors. In summary, the RNA-Seq approach was proven to be effective for transcriptome-profiling analyses. These results described here will provide a foundation for other functional genomic investigations concerning Sika deer and closely related species.

The sensitivity of RNA-Seq is a function of both molar concentration and transcript length. Therefore, we quantified transcript levels in terms of RPKM. The RPKM measure of read density reflects the molar concentration of a transcript in the starting sample by normalizing for RNA length, as well as the total read number in the measurement. This facilitates transparent comparisons of transcript levels both within and between samples (Mortazavi et al. 2008). We have selected a set of highly expressed genes related to endochondral ossification to validate our gene annotations. Each one of them generated the expected band size and showed consistent expression patterns. We are confident that our transcriptome dataset is a valuable addition to the publicly available Sika deer genomic information.

Conclusions

By analyzing transcript sequences from the ossifying antler tips of Chinese Sika deer using Illumina RNA-Seq technology, we have generated more than 130,000 high-quality sequences with a mean size of 405 bp, of which 43,937 sequences had a significant BLAST hit. Most of the transcript sequences obtained in this study were the first representatives of these transcripts for Sika deer. We assigned 22,743 sequences to 48 GO terms, 9,908 sequences to 25 COG classifications and 25,829 sequences to 203 KEGG pathways. We also identified many known genes encoding growth factors, transcription factors and extracellular matrix proteins related to endochondral ossification. These findings contribute substantially to the existing sequence resources for Sika deer, and they are certain to accelerate research on deer antler growth.

References

Aszodi A, Hunziker EB, Brakebusch C et al (2003) β1 integrins regulate chondrocyte rotation, G1 progression, and cytokinesis. Gene Dev 17:2465–2479

Benedetto AD, Watkins M, Grimston S et al (2010) N-cadherin and cadherin 11 modulate postnatal bone growth and osteoblast differentiation by distinct mechanisms. J Cell Sci 123:2640–2648

Chen FH, Herndon ME, Patel N et al (2007) Interaction of cartilage oligomeric matrix protein/thrombospondin 5 with aggrecan. J Biol Chem 282:24591–24598

Conesa A, Götz S, García-Gómez JM et al (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676

Deák F, Mátés L, Krysan K et al (1999) Characterization and chromosome location of the mouse link protein gene (Crtl1). Cytogenet Cell Genet 87:75–79

Eyre D (2002) Collagen of articular cartilage. Arthritis Res 4:30–35

Hale JE, Fraser JD, Price PA (1988) The identification of matrix Gla protein in cartilage. J Biol Chem 263:5820–5824

Harris MA, Clark J, Ireland A et al (2004) The gene ontology (GO) database and informatics resource. Nucleic Acids Res 32:D258–D261

Hartmann C (2009) Transcriptional networks controlling skeletal development. Curr Opin Genet Dev 19:437–443

Hinoi E, Bialek P, Chen YT et al (2006) Runx2 inhibits chondrocyte proliferation and hypertrophy through its expression in the perichondrium. Gene Dev 20:2937–2942

Iseli C, Jongeneel CV, Bucher P (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. Proc Int Conf Intell Syst Mol Biol 138–148

Ivkovic S (2003) Connective tissue growth factor coordinates chondrogenesis and angiogenesis during skeletal development. Development 130:2779–2791

Izu Y, Sun M, Zwolanek D et al (2011) Type XII collagen regulates osteoblast polarity and communication during bone formation. J Cell Biol 193:1115–1130

Kanehisa M, Goto S, Kawashima S et al (2004) The KEGG resource for deciphering the genome. Nucleic Acids Res 32:D277–D280

Kirsch T, Von der Mark K (1992) Remodelling of collagen types I, II and X and calcification of human fetal cartilage. Bone Miner 18:107–117

Kong L, Tian Q, Guo F et al (2010) Interaction between cartilage oligomeric matrix protein and extracellular matrix protein 1 mediates endochondral bone growth. Matrix Biol 29:276–286

Lee Y, Tsai J, Sunkara S et al (2005) The TIGR gene indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. Nucleic Acids Res 33:D71–D74

Li C, Suttie JM (1994) Light microscopic studies of pedicle and early first antler development in red deer (*Cervus elaphus*). Anat Rec 239:198–215

Li TF, O'Keefe RJ, Chen D (2005) TGF-β signaling in chondrocytes. Front Biosci 10:681–688

Li C, Yang F, Sheppard A (2009) Adult stem cells and mammalian epimorphic regeneration-insights from studying annual renewal of deer antlers. Curr Stem Cell Res Ther 4:237–251

Li R, Zhu H, Ruan J et al (2010) De novo assembly of human genomes with massively parallel short read sequencing. Genome Res 20:265–272

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta ct}$ Method. Methods 25(4):402–408

Mackie EJ, Ahmed YA, Tatarczuch L et al (2008) Endochondral ossification: how cartilage is converted into bone in the developing skeleton. Int J Biochem Cell B 40:46–62

Mak KK, Kronenberg HM, Chuang PT et al (2008) Indian hedgehog signals independently of PTHrP to promote chondrocyte hypertrophy. Development 135:1947–1956

Matsuo K, Irie N (2008) Osteoclast–osteoblast communication. Arch Biochem Biophys 473:201–209

Matsusaki T, Aoyama T, Nishijo K et al (2006) Expression of the cadherin-11 gene is a discriminative factor between articular and growth plate chondrocytes. Osteoarthr Cartil 14:353–366

Metzker ML (2009) Sequencing technologies–the next generation. Nat Rev Genet 11:31–46

Morello R, Bertin TK, Chen Y et al (2006) CRTAP is required for prolyl 3- hydroxylation and mutations cause recessive osteogenesis imperfecta. Cell 127:291–304

Mortazavi A, Williams BA, Mccue K et al (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5:1–8

Nakajima A, Shimizu S, Moriya H et al (2003) Expression of fibroblast growth factor receptor-3 (FGFR3), signal transducer and activator of transcription-1, and cyclin-dependent kinase inhibitor p21 during endochondral ossification: differential role of FGFR3 in skeletal development and fracture repair. Endocrinology 144:4659–4668

Newman B, Gigout LI, Sudre L (2001) Coordinated expression of matrix Gla protein is required during endochondral ossification for chondrocyte survival. J Cell Biol 154:659–666

Niyibizi C, Eyre DR (1994) Structural characteristics of cross-linking sites in type V collagen of bone. Chain specificities

and heterotypic links to type I collagen. Eur J Biochem 224:943–950

Ortega N, Behonick DJ, Werb Z (2004) Matrix remodeling during endochondral ossification. Trends Cell Biol 14:86–93

Price J, Allen S (2004) Exploring the mechanisms regulating regeneration of deer antlers. Phil Trans R Soc Lond B Biol Sci 359:809–822

Price PA, Williamson MK (1985) Primary structure of bovine matrix Gla protein, a new vitamin K-dependent bone protein. J Biol Chem 260:14971–14975

Schroeder A, Mueller O, Stocker S et al (2006) The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol 7:3

Shi CY, Yang H, Wei CL et al (2011) Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. BMC Genomics 12:131

Szuwart T, Kierdorf H, Kierdorf U et al (2002) Histochemical and ultrastructural studies of cartilage resorption and acid phosphatase activity during antler growth in fallow deer (*Dama dama*). Anat Rec 268:66–72

Tatusov RL, Natale DA, Garkavtsev IV et al (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res 29:22–28

Trackman PC (2005) Diverse biological functions of extracellular collagen processing enzymes. J Cell Biochem 96:927–937

von der Mark K, von der Mark H (1977) The role of three genetically distinct collagen types in endochondral ossification and calcification of cartilage. J Bone Joint Surg Br 59B:458–464

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63

Wang XW, Luan JB, Li JM et al (2010) De novo characterization of a whitefly transcriptome and analysis of its gene expression during development. BMC Genomics 11:400

Wenstrup RJ, Florer JB, Brunskill EW et al (2004) Type V collagen controls the initiation of collagen fibril assembly. J Biol Chem 279:53331–53337

Ye J, Fang L, Zheng H et al (2006) WEGO: a web tool for plotting GO annotations. Nucleic Acids Res 34(Web Server issue):W293–W297

Yoshida CA, Yamamoto H, Fujita T et al (2004) Runx2 and Runx3 are essential for chondrocyte maturation, and Runx2 regulates limb growth through induction of Indian hedgehog. Gene Dev 18:952–963

Zheng Q, Zhou G, Morello R et al (2003) Type X collagen gene regulation by Runx2 contributes directly to its hypertrophic chondrocyte-specific expression in vivo. J Cell Biol 162:833–842