Genomics 112 (2020) 252–262

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Original Article

Muscle transcriptome signature and gene regulatory network analysis in two divergent lines of a hilly bovine species Mithun (*Bos frontalis*)

Sabyasachi Mukherjee[a,1], Anupama Mukherjee[a,b,1], Rahul Singh Jasrotia[c], Sarika Jaiswal[c], M.A. Iquebal[c], Imsusosang Longkumer[a], Moonmoon Mech[a], Kezhavituo Vüpru[a], Kobu Khate[a], Chandan Rajkhowa[a], Anil Rai[c], Dinesh Kumar[c,*]

[a] ICAR-National Research Centre on Mithun, Medziphema, Dimapur, Nagaland 797106, India
[b] Division of Dairy Cattle Breeding, ICAR-National Dairy Research Institute, Karnal, Haryana 132001, India
[c] Centre for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, New Delhi 110012, India

ARTICLE INFO

Keywords:
Bos frontalis
de novo assembly
Marker discovery
Gene regulatory network
Transcriptome

ABSTRACT

A massive bovine, *Bos frontalis*, also known as Mithun or Gayal, found at higher altitude is very promising meat and milk animal. For candidate gene and marker discovery, RNA-seq data was generated from *longissimus dorsi* muscle tissues with Illumina-HiSeq. Such markers can be used in future for genetic gain of traits like feed conversion efficiency (FCE) and average daily gain (ADG). Analysis revealed 297differentially expressed genes (DEGs) having 173 up and 124 down-regulated unigenes. Extensive conservation was found in genic region while comparing with *Bos taurus*. Analysis revealed 57 pathways having 112 enzymes, 72 transcriptional factors and cofactors, 212 miRNAs regulating 71 DEGs, 25,855 SSRs, mithun-specific 104,822 variants and 7288 indels, gene regulatory network (GRN) having 24 hub-genes and transcriptional factors regulating cell proliferation, immune tolerance and myogenesis. This is first report of muscle transcriptome depicting candidate genes with GRN controlling FCE and ADG. Reported putative molecular markers, candidate genes and hub proteins can be valuable genomic resources for association studies in genetic improvement programme.

1. Introduction

Mithun or Gayal (*Bos frontalis*), also known as 'Ceremonial Ox' or 'Cattle of mountain' is an important bovine species roaming freely inside difficult terrains of the North Eastern Hilly forest region of India, Bangladesh, Bhutan, China and Myanmar and provide low-cost highly nutritious meat to tribal people of the region [1]. The world's largest cattle species, *Bos frontalis* Lambert, 1804 is domesticated form whereas *Bos gaurus* Hamilton-Smith, 1827 is the wild form of gaur population [2].

Studies on mithun offers an interesting evolutionary events and window to see how speciation mechanism in divergence of genus has 'ecological signature' along with adaptive 'behavioral signature'. The mitochondrial DNA based phylogenetic study has revealed close genetic relationship among various species of genus *Bos*, like *B. frontalis*, *B. indicus, B. taurus, B. primigenius, B. mutus, B. javanicus, B gaurus, B. grunniens* [3,4].

Genus *Bos* has > 10 genetically close living species as on today. For example, *B. grunniens* (Yak) has adaptation for higher altitude (> 3000 m above mean sea level). Whereas *B. frontalis* has adaption for hilly forest (1000–3000 m above mean sea level) having mineral deficiency due to leaching and abundance of coarse fodder. The signature of this unique ecological niche can be seen in mithun's behaviour in form of salt licking to compensate mineral deficiency as well as feeding behaviour on coarse fodder unlike cattle (*B. taurus*) [5].

India has the highest mithun population in the world, 0.30 million which represents 97.57% of the world population [6]. Mithun plays an important role in the socio-economic and cultural life of the local tribal population. Since is used as meat animal as a source of protein having projected demand increase of 73% by 2050, thus it can be an alternative source [7]. It also gives 2–4 l milk per day having high nutritional values: fat (8–13%), solid-not-fat (SNF) (15–18%) and protein (5–7%). Its leather is superior due to its toughness, tensile strength and longevity [8,9].

Muscle tissue has immense biological value for the human diet, thus, global demand for a quality product has increased. Mithun has the

* Corresponding author.
E-mail address: dinesh.kumar@icar.gov.in (D. Kumar).
[1] Both authors equally contributed.

https://doi.org/10.1016/j.ygeno.2019.02.004
Received 1 September 2018; Received in revised form 30 January 2019; Accepted 6 February 2019
Available online 27 February 2019
0888-7543/

very unique attribute of being an organic meat of choice having high tenderness, having less muscle fiber diameter [10] and it also fetches almost double market price over other bovine species [11]. Further, cattle meat producers and consumers have no threat perception of mad cow disease and associated market havoc due to its reported resistance [12]. Mithun is also known for adaptive climatic resilience and resistance to the disease toxoplasmosis [13]. The total animal productivity of this unique bovine has been hitherto unexploited.

There is huge variation in growth rate of mithun. In order to increase the meat productivity, there is enough scope for improvement. With optimal feeding, the mithun growth rate varies very widely from 266 to 733 g per day with an average of 400 g per day [14]. It is projected that it can be increased upto 900 g per day [15]. The estimated heritability of body weight at birth and maturity are 0.23 ± 0.11 and 0.11 ± 0.03, respectively [16]. Improvement of such low heritable traits requires discovery of key candidate genes and its markers for improvement of feed conversion efficiency (FCE) [17]. It is imperative to use a transcriptomic approach to delineate causative genetic variants underlying phenotypes of interest [18].

In livestock meat, the higher relative mass of muscles and associated quality trait makes it commercially important. Though very recently, draft assembly of mithun genome [19] is available but there is no tissue specific transcriptomic data for annotation, covering muscle tissue. Transcriptomic studies on longest muscle of bovine, *longissimus dorsi* (*LD),* is widely used to study the growth and phenotypic variation in cattle [20]. Tissues from this longest muscle of largest bovine can be used for knowledge discovery by transcriptomic approach deciphering key pathways associated with FCE. Such studies can discover key candidate genes and their markers, which can be used for Marker Assisted Selection (MAS) in the genetic improvement programme of this species [21]. Though similar studies have been done in other domestic animals leading to industrious use but there is no report on muscle transcriptome analysis of this unique promising bovine species. Such data can also be used to predict miRNAs and transcription factors (TFs) involved in muscle development which is also yet to be reported.

Therefore, the present investigation was aimed at the discovery of differentially expressed genes (DEGs), novel transcripts and pathways operating in muscles using two divergent growth lines of mithun population. The study also aims to delineate the number of DEGs controlled by miRNA, transcription factors, and the gene regulatory network depicting hub protein genes, along with the discovery of SSR, SNP and Indel molecular markers.

2. Results

2.1. Pre-processing and de novo assembly

Using two divergent growth lines of mithun classified as high growth (HG) and low growth (LG), a total of 82 GB of transcriptome data (RNAseq) was generated from *longissimus dorsi muscle* (LDM) tissues of four animals (two each from high and low growth lines) with Illumina HiSeq (2x100bp paired-end) chemistry. The sequencing data are deposited at NCBI (BioProject accessions: PRJNA307305; BioSample accessions: SAMN04384021, SAMN04384020, SAMN04384019 and SAMN04384018). A total of 253,477,291 paired-end raw reads with a length of 101 bp were generated from which 11,739,269 low quality reads were removed. Finally 241,738,022 high quality cleaned reads with a phred score ≥ Q20 were obtained.

Two biological replicates of HG and LG reads were pooled together for a *de novo* transcriptome assembly. Among the three different methods used *namely*, *SOAPdenovo-trans*, *Spades* and *Trinity*. Assembly generated by *trinity* was further used for downstream analysis based on the N50 value (Table 1). The generated *trinity* assembly had 186,086 transcripts with 47.79% of GC content, average contig length 770.95 and N50 value of 1635. The minimum and maximum transcript lengths were 201 and 31,699 bp, respectively (Fig. 1 and Table 2). It was found that the majority of transcript lengths ranged between 201 and 299 and between 300 and 399. In the *de novo* transcriptome assembly, a total of 10,129 transcripts were found to be involved in the formation of the isoforms, and overall 20,532 isoforms were present. The number of isoforms exhibited in a single unigene varied from 2 to 23.

Table 1
Construction of transcriptome assembly by three different tools.

Tools	Number of assembled transcripts	N50 (bp)	GC%
SOAPdenovo-trans	309,936	1280	47.30
Spades 3.9.0	465,147	1191	47.13
Trinity	186,086	1635	47.79

2.2. Identification of differential expressed genes

HG and LG reads were mapped to a *de novo* transcriptome assembly separately to compute abundance estimate values in the form of FPKM (fragments mapped per kilo base of exon per million reads mapped). On the basis of these expression values, 297 differentially expressed genes (DEGs) were identified among HG and LG using the *edgeR* package of Bioconductor. Stringent parameters with a FDR < 0.05 and $\log_2$ fold change value > 4 were used to obtain the DEGs. The obtained range of fold change for identified DEGs varied from 4.0 to 15.78 Out of 297 DEGs, 173 and 124 unigenes were up and down regulated, respectively. For graphical representation of the DEGs, a MA and volcano plot were generated. DEGs are shown as red dots (Fig. 2). Hierarchical clustering and a heat map were also generated to represent the DEGs (Fig. 2B).

2.3. Functional categorization and pathway analysis of differential expressed genes

A similarity search of DEGs was performed using a NCBI standalone tool using a NCBI non-redundant (NR) database (ftp://ftp.ncbi.nlm.nih.gov/blast/db). Out of 297 DEGs, a significant match with a known sequence was found for 269 (90.5%) unigenes. Among these, only 222 unigenes were involved in the annotation process. *Bos taurus* showed a maximum similarity of 114 unigenes, followed by 40 and 36 with *Bison bison* and *Bos mutus*, respectively. Blast results of 15 top hits species distribution of DEGs of mithun are shown in figure (Fig. 3) (Supplementary Table S1).

Gene ontology analysis revealed a cellular process (biological process), binding (molecular functions) and cell (cellular components) in 171, 174 and 179 transcripts, respectively. It was followed by a single-organism process (biological process), catalytic activity (molecular functions) and organelle (cellular components) in 149, 90 and 158 sequences, respectively (Fig. 4).

A KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway using a Blast2Go Pro revealed a total of 57 pathways having 112 enzymes. Classification of these enzymes showed a maximum (seven times) of 'Biosynthesis of antibiotics' type followed by 'Purine metabolism' and 'Sphingolipid metabolism' type with a frequency of six and five times, respectively (Supplementary Table S2).

Transcription factors and cofactors play very important roles in the expression of genes. Using AnimalTFDB2.0, 60 TFs and cofactors were predicted when searched against *Bos taurus* (Supplementary Table S3). Further, 12 more TFs and cofactors were found when searched for against all species. Out of all 72 TFs and cofactors, *CENPF* and *CDK2* were found to control the maximum number of DEGs, *i.e.*, 9 and 5, respectively. Five TFs *viz. USP22*, *NEO1, EHMT1, CDK7, TAF1D* were also found controlling two DEGs.

2.4. Prediction miRNA regulated DEGs

We obtained hits of 297 DEGs with all 793 miRNAs with a total

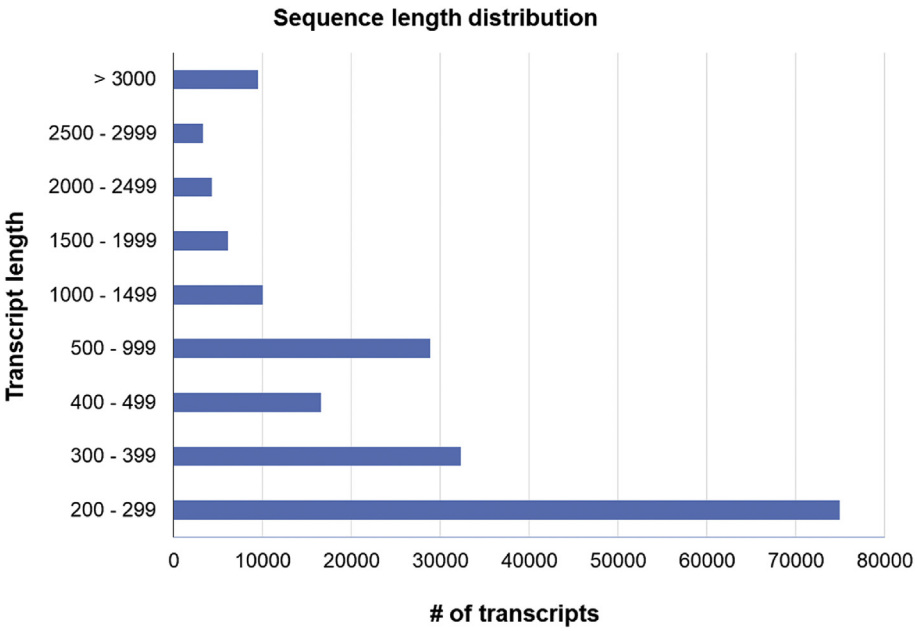

Fig. 1. Sequence length distribution of *de novo* transcriptome assembly.

Table 2
De novo transcriptome assembly statistics.

Total *trinity* transcripts	186,086
Total *trinity*'genes'	165,558
Percent GC	47.79
Contig N50	1635
Median contig length	346
Average contig	770.95
Total assembled bases	143,463,330

score of 140 to 3132 and total energy of −1.37 to −503.9 kcal/mol. After putting a cut-off threshold of the total score at 1000, 212 miRNA regulating 71 DEGs were obtained (Supplementary Table S4).

2.5. Detection of simple sequence repeats (SSR) markers

A total of 25,855 and 136 simple sequence repeats (SSRs) were found from the transcriptome assembly and DEGs, respectively. We could successfully design primers for 18,113 and 103 from *de novo* assembly and DEGs loci, respectively (Table 3 and Supplementary Table S5).

2.6. Variant detection

Aligning with the *Bos taurus* version of the UMD3.1 assembly, a total of 267,096 and 252,300 variants were obtained from HG and LG, respectively. Among these 195,867 variants were common. In these variants, 23,681 and 21,508 indels were found in HG and LG, respectively (Table 4). The maximum number of variants, *i.e.*, 14,889 and 14,182 were found in chromosome number 19 against *Bos taurus* in both HG and LG, respectively. While aligning against the *de novo* transcriptome, mithun-specific 70,794 SNPs and 3771 indels in HG and 64,525 SNPs and 3517 indels were found in LG. Among these, 34,014 SNPs and indels were found common in the same unigene ID and same position (Fig. 5A). While mining these variants, a common filtering criterion of minimum 15× read depth and quality of 30 were used [13,14]. Unigene ID c41909_g1_i1 (interferon-induced very large GTPase 1-like protein) contained the maximum number (158 and 117) of SNPs in HG and LG, respectively, against the *de novo* transcriptome assembly (Supplementary Table S6). Annotations of identified variants were performed against the *Bos taurus* version UMD3.1 and detected the region where SNPs were found a maximum number of times (top seven categories of the genic region) were shown in Fig. 5B.

2.7. Gene regulatory network

Out of 297 DEGs, 26 were duplicates, so only 271 genes were selected for constructing a network in which 160 were up and 111 were down regulated. We used the edge betweenness algorithm to determine the functional components in the network [22]. A total 8987 edges were found in the network. We used the edge betweenness values 35 to 535.37 to obtain the network with 1414 edges. The GRN was constructed in which genes were interconnected to other genes on the basis of degree with maximum 120 and minimum 14 (Fig. 6). In the GRN of transcription factors and cofactors construction, out of 72 only 52 were used after removing similar types of TFs and cofactors. In this GRN, we found two clusters, *i.e.*, one with 44 genes and other with 8 genes, which were having 2 to 23 degrees (Fig. 7).

3. Discussion

Transcriptome analysis of two divergent growth lines of mithun elucidated a set of 297 differentially expressed genes (DEGs) which has been used to construct GRN along with variant analysis. We also found differences in signaling pathways of muscle differentiation between these two lines of mithun.

The differentiation of myogenic cells is mainly controlled by two signaling pathways, *namely*, the phosphatidylinositol 3-kinase/Akt pathway and the NOTCH/Hes pathway. In the NOTCH signaling pathway, which is highly conserved [23], we found differential gene expression in its receptor gene *NOTCH4* with up-regulation in HG. Contrary to this, in the other pathway, the phosphatidylinositol phosphatase (*PTPRQ*) gene was found up-regulated in LG. Since microtubule-associated protein 4 (*MAP4*) protein is known to promote microtubule assembly and counteract the destabilization of microtubule catastrophe promotion, thus it plays a critical role in the formation and maintenance of muscle. This gene was found to be up-regulated in HG. This gene is reported to vary in isoform in different muscles during myogenesis. Its isoform is also associated with different level of differentiation in myogenesis [24].

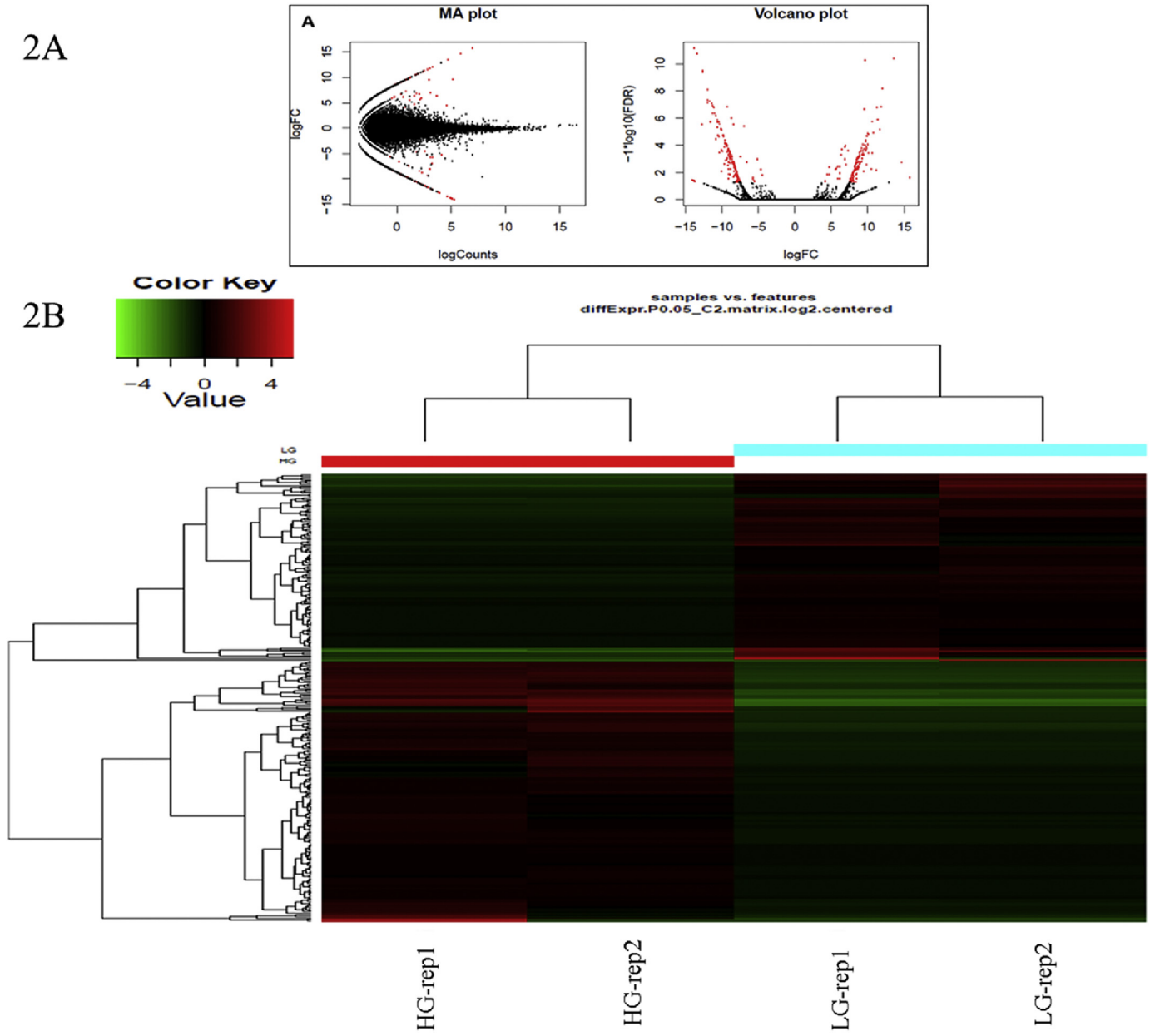

Fig. 2. A: MA plot and volcano plot represents the differential expressed genes, red dot showed the expressed genes and black color dots were non-expressed genes. **B:** Heat map and hierarchical clustering represents the differentially expressed genes of two replicate samples of high growth (HG-rep1 and HG-rep2) and low growth (LG-rep1 and LG-rep2) of mithun using *p* value 0.05 and log2 fold change 2. Green color showed the downregulated genes and red showed the upregulated genes.

The Janus kinase (JAK)/signal transducer and activator of transcription (STAT) signaling pathway plays a role in myogenic differentiation. We found JAK2 down-regulated and STAT2 up-regulated in HG. This is because STAT2 operates downstream to JAK2 [25]. Collagen, which regulates fibrillogenesis, is an abundant connective tissue protein and is a contributing factor to variation in meat tenderness and texture. We found genes of this pathway, COL14A1 down-regulated in HG. Marbling is an important trait regarding the quality of meat. SLC27A6, long-chain fatty acid transport protein 6, which is associated with higher marbling, was found to be down-regulated in the HG class of mithun [26]. This may be because of the free-ranging nature of this species where there is no intensive selection like cattle. The *PPP6R3* gene in the Wnt signaling pathway is associated with lean body mass in humans, and its variation has been associated with bone strength. It also exerts pleiotropic effects on muscles mass [27]. We found this gene is up-regulated in HG. The Coiled-coil domain-containing protein 8 (*CCDC8*) gene is already reported with differential expression (up-regulated) in skeletal muscle of beef cattle breeds with respect to dairy breeds [28]. We found a similar pattern of expression in HG. The centrosomal protein kiaa1731 (*CEP295* or *KIAA1731*) gene, which is associated with adipose fat distribution along with insulin biology in humans, was found over-expressed in HG [29]. The ubiquitin-specific peptidase 9, Y-linked (*USP9Y*) gene, which are known to be associated with muscle development and differentiation, were found to be up-regulated in HG [30].

Interestingly, the mitogen-activated protein kinase (MAPK) signaling pathway, which is highly relevant to lipid metabolism in muscle and adipose tissue differentiation [26], in this pathway, the *Sema4C* gene was found to be over-expressed in LG. The *SYNJ1* gene of the phosphatidylinositol signaling system is down-regulated in HG mithun, which has been reported to be over-expressed in pigs [31]. Interestingly, we found two isoforms of gene sphingomyelin phosphodiesterase

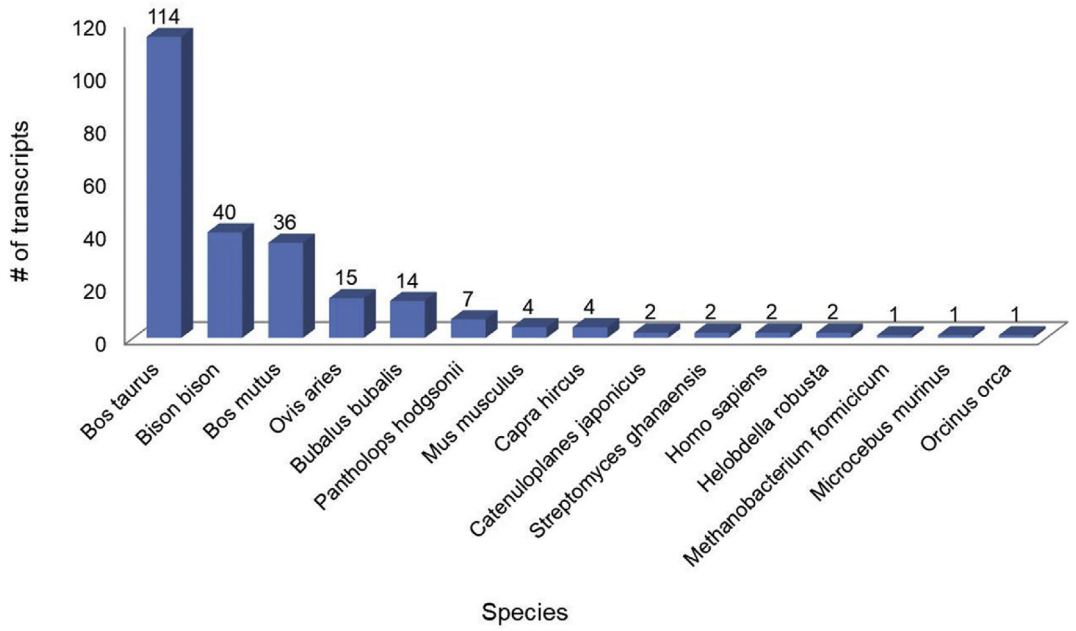

Fig. 3. Top 15 species distribution of mithun differential expressed genes.

4 (*SMPD4*) being up and down-regulated in HG. This gene has been reported with alternative miRNA splicing of exon 11 and exhibits differential biological activity affecting sphingolipid and oxidant signaling, which is related to glucose uptake and atrophy [32]. The glycolipid transfer protein (*GLTP*) gene related to lipid transport/metabolism was found to be down-regulated in HG, which is similar to dogs [33]. The Bestrophin 3 (*best-3*) gene is known to be involved in the regulation of cell proliferation, apoptosis and differentiation. Out of three isoforms of this gene, two of them are over-expressed in HG. This gene is reported with three different isoforms having differential activity in mice and human [34]. The *LMO7* gene was found with three isoforms (two up- and one down-regulated). This gene has been reported with two isoforms having an association with cell proliferation in humans and rats [35]. The interferon-induced 44 (*IFI44*) gene is well reported with the trait of feed conversion efficiency, and it is differentially expressed in *Bos taurus* in low and high fat heifers. In high fat

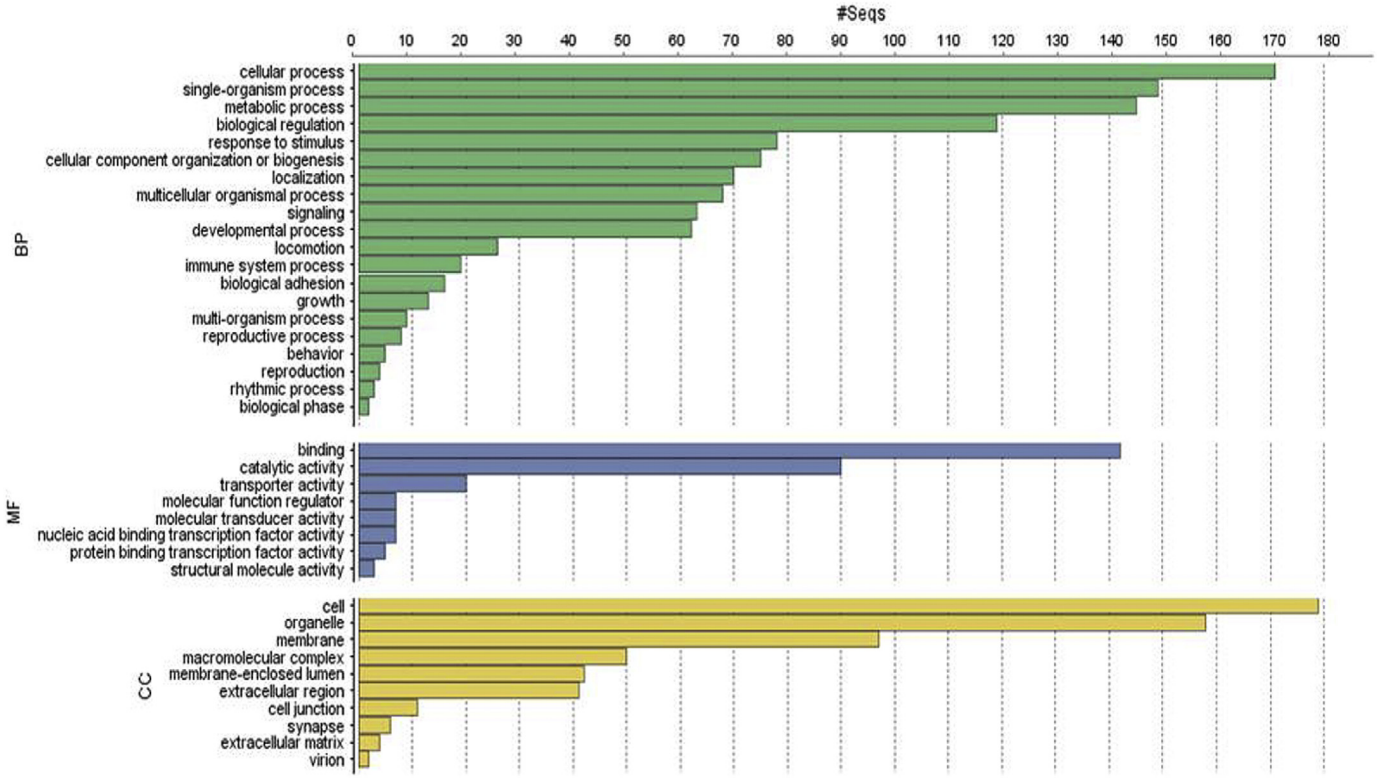

Fig. 4. Functional categorization of annotated differentially expressed genes of mithun. Green color bars represent the biological process, blue color bars represent the molecular functions and yellow bars represent the cellular components. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

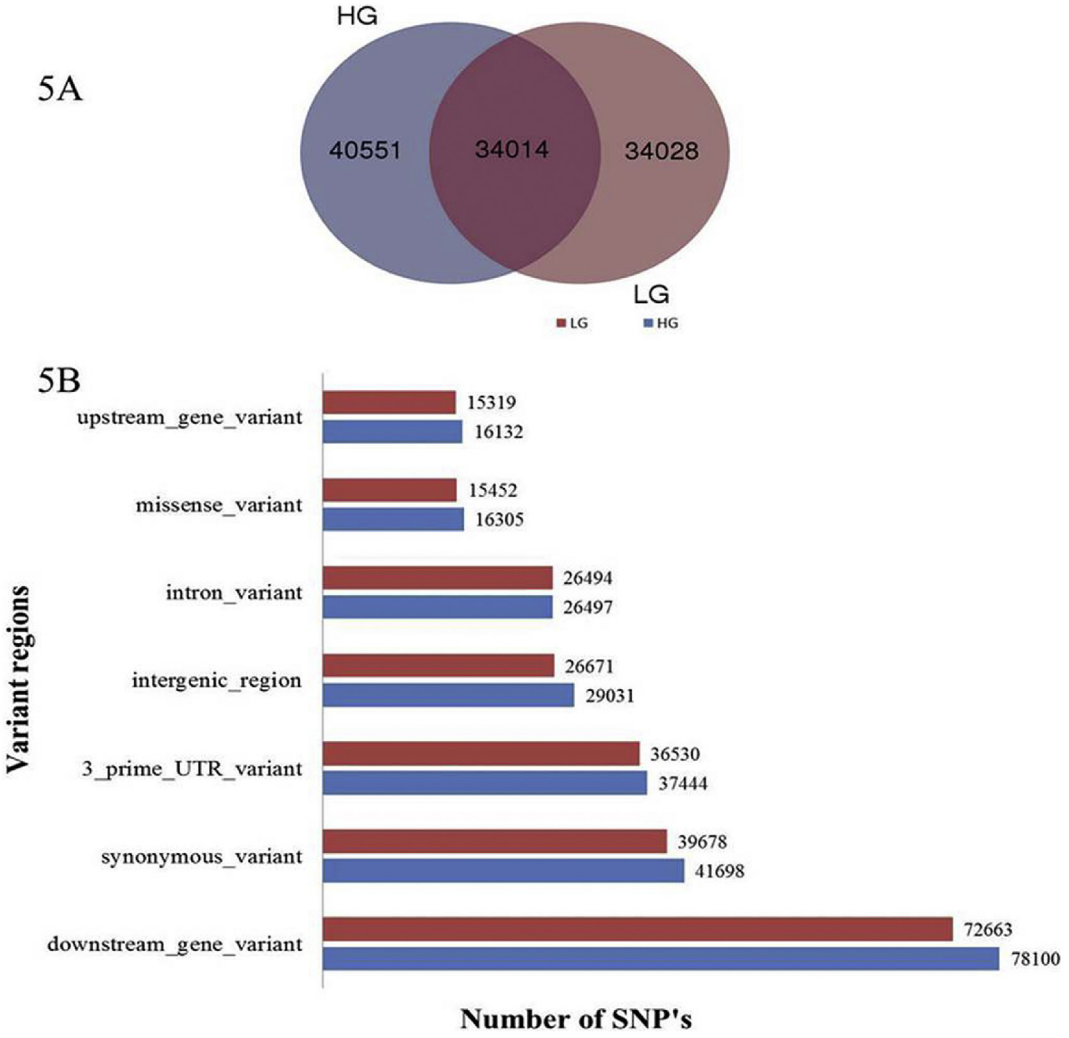

Fig. 5. A: Venn diagram of common SNP and Indels were found in HG (high growth) and LG (low growth) against *de novo* transcriptome reference. **B:** Annotation graph of high growth and low growth SNPs against *Bos taurus* genome.

Table 3
Putative SSR markers identified from *de novo* transcriptome assembly and DEG's.

	Transcriptome assembly	DEGs
Sequences examined	186,087	297
Identified SSRs	25,855	136
SSR containing sequences	19,820	97
Sequences containing more than 1 SSR	3935	30
SSRs present in compound formation	1181	5
Mono-nucleotide	15,173	77
Di- nucleotide	6256	32
Tri- nucleotide	4180	26
Tetra- nucleotide	231	1
Penta- nucleotide	9	Nil
Hexa- nucleotide	6	Nil

Table 4
Table represent the numbers of SNP's were identified in high growth and low growth against the *Bos taurus* reference as well as *de novo* transcriptome assembly.

Genotype	Variant detection			
	Against *Bos Taurus* version UMD3.1		Against *de novo* transcriptome assembly	
	SNP	Indel	SNP	Indel
High Growth	243,415	23,681	70,794	3771
Low Growth	230,792	21,508	64,525	3517

heifers, it was down-regulated. A similar observation was found in HG mithun as this gene is related with interferon signaling affecting immunity and thus directing energy balance towards FCE [36]. The *SGMS1* gene has been reported in association study of FCE using PorcineSNP 60 BeadChip in pigs [37]. In mithun, we found this gene is up-regulated.

All 793 miRNAs of *Bos taurus* showed a very high hit over all 297 DEGs with variable score and energy. Such computational prediction of miRNA targets is a critical initial step in identifying miRNA:mRNA target interactions for future experimental validation studies [38]. This is because each miRNA may have targets of approximately 200 genes [37]. We found extensive conservation of *Bos taurus* miRNA in mithun, which is because of the genetic propinquity being a common genus of *Bos*. Both the species of *Bos* had a common progenitor *Bos primigenius* [39]. MiRNAs have important roles in muscle development and hypertrophy, adipose tissue growth and various other physiological roles in farm animals [40]. It also regulates different stages of myogenesis [41]. This is the first report showing the role of miRNA controlling DEGs in mithun muscle transcriptome. In the case of cattle, the miRNA and its polymorphism in the seed region have been found with a trade-

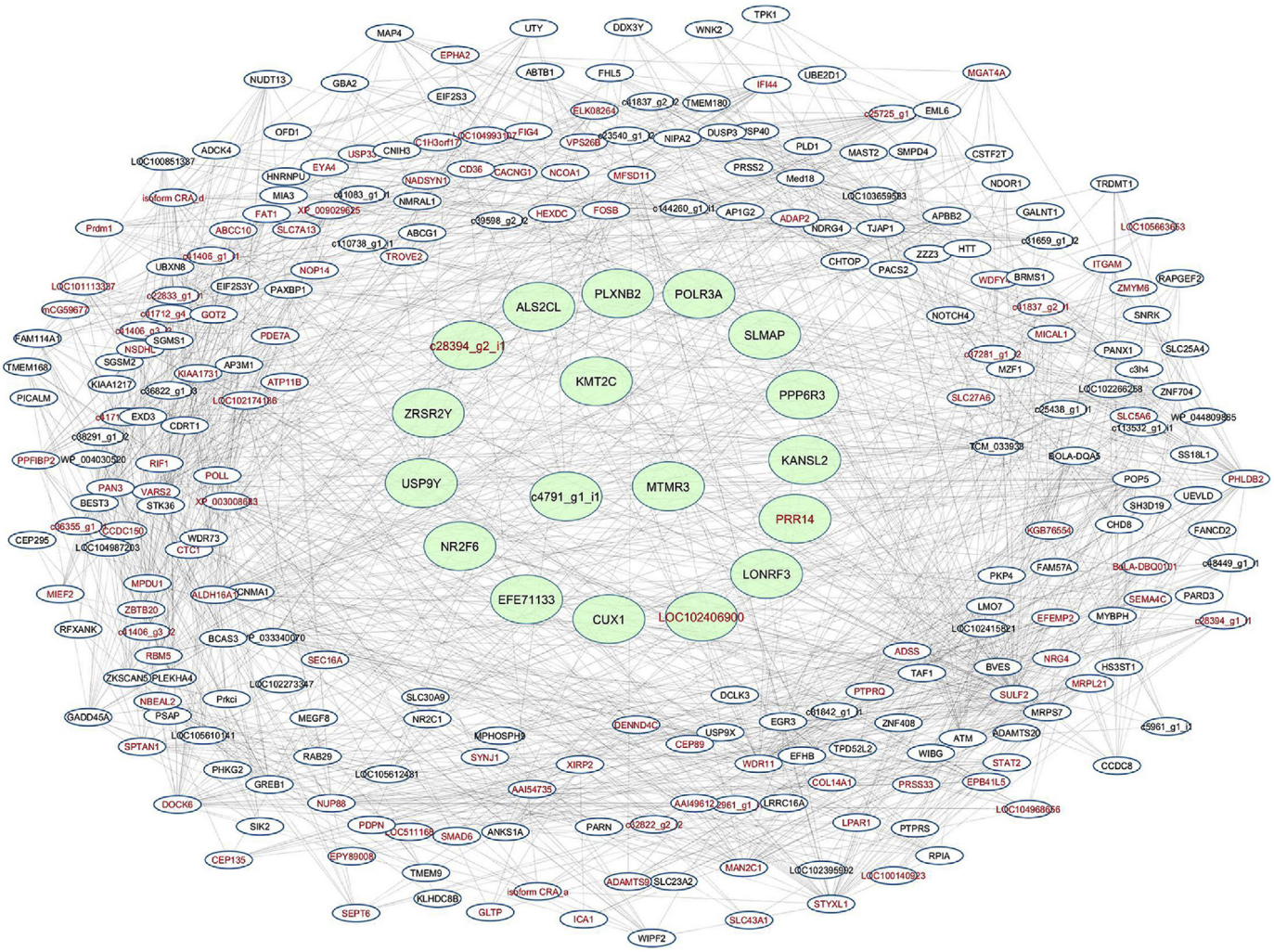

Fig. 6. Gene regulatory networks of differentially expressed genes (black color represent upregulated genes, red color represent downregulated genes, hub genes at the center are represented with light green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

off role between FCE and the tissue construction controlling phenotype. MicroRNA motifs have also been identified near SNP associated with residual feed intake in cattle, suggesting that miRNA has a functional role in the regulation of genes that affect feed efficiency [42]. Such polymorphisms in miRNA genes, especially within the seed region of 3'UTR, can be a putative biomarker for phenotypic traits of interest in mithun [37]. Our enlisted miRNA controlled DEGs can be mapped to individual chromosomes over the future *de novo* mithun genome assembly as and when it gets improved from the present draft assembly [19]. Our findings would be pivotal in further predicting the extended region of 3' UTR of transcripts. Recent studies indicate that SNP within the 3'UTR region may affect traits of interest by affecting the function of miRNA. Such 3'UTR SNPs are functional candidates and thus of interest to subsequent functional research like eQTL, GWAS [43] and genomic selection studies [44].

Out of 18 hub genes, 14 genes are well described in the literature with their respective role. Among these, six, *namely*, *MTMR3* [45], *CUX1* [46], *LONRF3* [47], *PLXNB2* [48], *KMT2C* [49] and *ZRSR2Y* [50] genes were related to overall cell growth, proliferation, differentiation and cell cycle regulation. Further, four genes *namely NR2F6* [51], *ALS2CL* [52], *PPP6R3* [53] and *POLR3A* [54] were found to be related to tolerance and regulation of immunity. The rest of the four genes *namely*, *PRR14* [55], *USP9Y* [30], *SLMAP* [56] and *KANSL2* [57], were found to be associated with skeletal muscle, myogenesis, development, differentiation and multiprotein assembly.

The *POLR2A* transcription factor interacts with various proteins involved in the coordination of skeletal muscle development or myogenesis and repair [58]. The *KMT2C* transcription factor is known for its role in cell proliferation [49], *POLM* (Polymerase DNA Directed Mu) is known to act as DNA damage repair genes [59], the *TRIM25* (Tripartite motif containing 25) hub transcription factor is known to modulate cell proliferation and apoptosis [60] and the *SLC2A4RG* transcription factor is known to act as a myocyte enhancer gene in the *AMPK* signaling pathway [61].

Candidate genes of our transcriptome analysis results, *i.e.*, up- and down-regulated pathway genes, transcriptional factors, and hub genes can be taken up for future SNP discovery work to get alleles associated with different phenotypes/traits of interest. Such candidate gene-based SNP discovery has been found to be successful in various association studies of other domestic animal species, for example, the bovine *SPP1* C > T SNP marker with higher body weight [62], FCE/ADG genes on chromosome $4^{10}$and 4 Mbp long region on *BTA6* was found to be a potential part to harbour candidate genes influencing growth [63]. Forty-eight SNPs were located on *BTA6*, leptin gene SNP *A59V* and *BM1500* microsatellite for rump fat thickness and slaughter muscle color [64], the *UCP3* gene 3076 A > G with marbling score of cattle [65], the *DGATE1* gene with AA genotype and sirloin weight/fat depth [66],gene markers for *HAL, ESR, MC4R, PRKAG3 (RN)* and *IGF2* in the pig; *DGAT1, GHR, MC1R, MSUD, ASS* and *CAST* in cattle; *CLPG1, FGFR3*, and *DMP1* in sheep; the B complex and *FM03* in chickens [67]

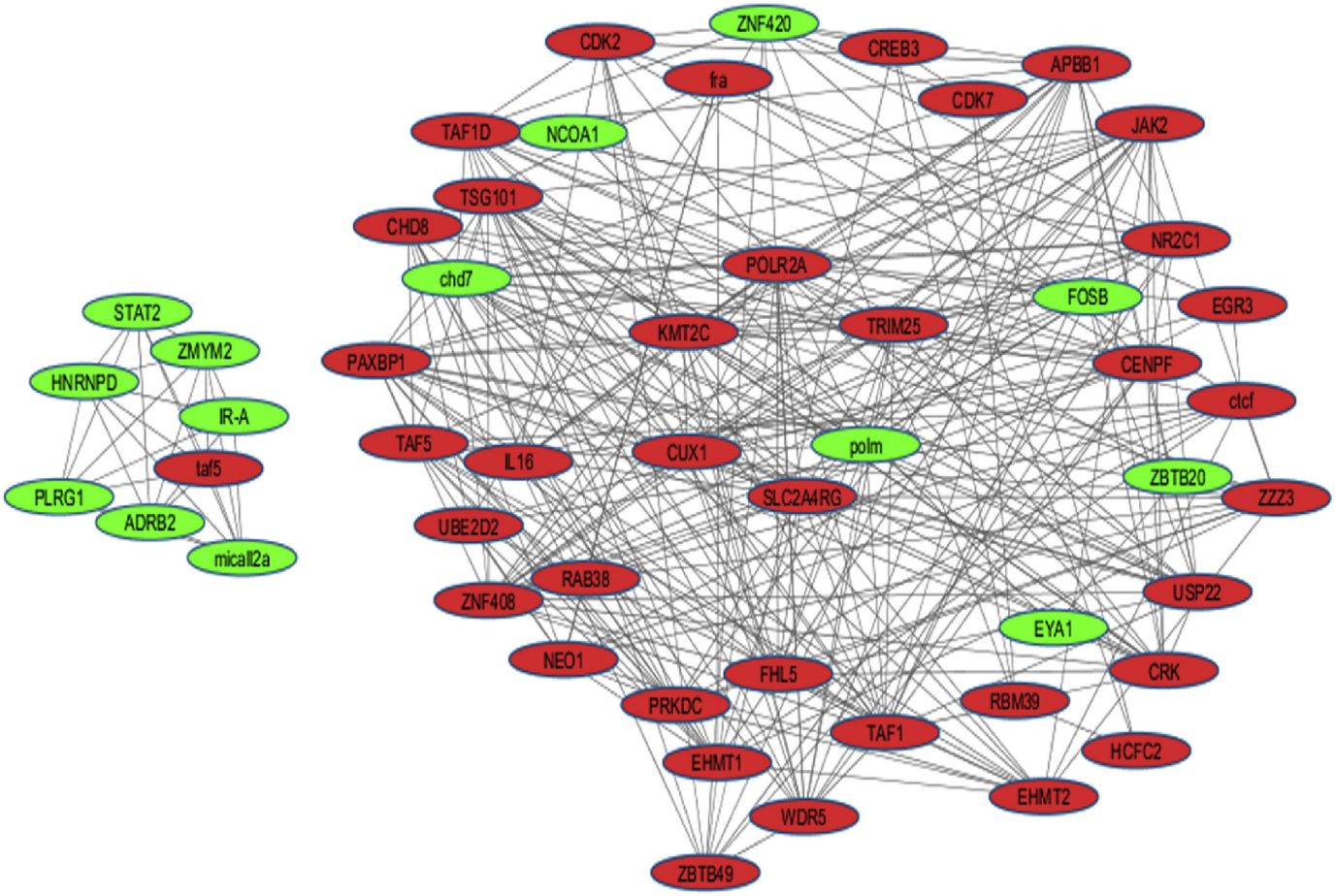

Fig. 7. Gene regulatory network of transcriptional factors and cofactors (red color represents upregulated and green represents downregulated). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

have been reported.

Present finding has immense value in terms of knowledge discovery as well as generation of mithun genomic resources. Recently sequenced mithun genome reports 41,227 predicted genes having 26,667 final Glean models. Due to scarcity of transcriptomic data in mithun, these predictions are yet to be annotated by evidence of transcriptomic data as done in other domestic animals like buffalo [68]. Interestingly, mithun genome has 19,666 highly conserved genes of cattle (*Bos taurus*) due to similar genera having high genetic propinquity. This high extent of conservation has also been observed in our transcriptomic findings where highest similarity was found with *Bos taurus* (out of 269, 114 DEGs) (Fig. 3). Mapping of total transcripts over *Bos taurus* assembly also revealed > 90% of conservation. Further, *in silico* investigated miRNAs, transcription factors and putative marker has advantage in terms of reducing cost and time of *in vitro* discovery but requires further validation. These putative markers may be valuable genomic resource for future association studies.

4. Conclusion

This is the first report of the muscle transcriptome analysis of the hilly bovine species, Indian mithun. We have identified a total of 297 differentially expressed genes (DEGs), out of which 173 and 124 unigenes were up- and down-regulated, respectively and many of these genes are responsible for muscle growth. Extensive conservation was found in the genic region of mithun with other *Bos* genus reflecting common ancestry. Analysis revealed 57 pathways having 112 enzymes, 72 TFs and cofactors, 212 miRNA regulating 71 DEGs, 25,855 SSRs, 104,822 SNPs, 7288 indels, a gene regulatory network having 24 hub genes and TFs controlling three major process, *namely*, cell proliferation, immune tolerance and myogenesis in two divergent growth lines of mithun. The present finding generated many significant baseline information which may be of immense use for a future genomic approach to increase the rate of genetic gains of important economic growth traits, *namely*, FCE and ADG compared to traditional breeding of mithun. Such an approach is imperative in the endeavour of fostering higher productivity of mithun in terms of better growth traits.

5. Material and methods

5.1. Animal selection and sample collection

Entire experiment were carried out in accordance with CPCSEA (Committee for the Purpose of Control and Supervision on Experiments on Animals) guidelines, prescribed by ICAR, Ministry of Agriculture and Farmers Welfare, Government of India. Accordingly, Animal Ethics Committee approved the collection of muscle biopsy samples from the experimental mithun. This permission has been documented (NRCM (RPM) 4/2010 (Vol. I)) in the proceeding with approval in the 14th meeting dated 8th October 2012. Subsequently, it was also approved by apex research body, *i.e.*, IRC (Institute Research Committee) vide document number ICAR-NRCM(G)135/99, dated 26th February 2013.

Average daily gain (ADG) of body weight is considered to be most important trait to evaluate production ability or FCE with meat quality [69]. Two divergent lines, low growth (LG) having average ADG 100.0 ± 5.0 g per day ($P \leq 0.5$) and high growth (HG) lines having 400.0 ± 30.0 g per day ($P \leq 0.5$) being maintained at ICAR-National Research Centre on Mithun, Medziphema, Nagaland, India were used

for this study. Two male animals of 2 year age from each line were selected for transcriptomic data generation. Meat quality and carcass were obtained from 12 animals of each lines. Marbling was scored from maximum to minimum in scale of 1–10 which were slightly and moderately abundant in HG and LG lines, respectively. Among carcass quality trait dressing percentage, adjusted fat thickness (inch) and rib eye area (sq. inch) were found to be 58.71 ± 0.11, 55.96 ± 0.34, 1.26 ± 0.12, 1.02 ± 0.09, 13.07 ± 0.34 and 11.64 ± 0.56 in HG and LG lines, respectively. Among meat quality traits, marbling score, texture of marbling, color of lean, firmness of lean and texture of lean were recorded. All the traits were found to be significantly different at 5% level of significance. *LD* muscle tissue samples (approximately 5 g) were collected aseptically through open muscle biopsy method under conscious anaesthesia. Fresh tissues were immediately frozen on liquid nitrogen after collection and then stored at −80 °C for preservation until use.

5.2. RNA extraction, cDNA synthesis, library preparation and sequencing

RNA was extracted from each of the four muscle tissues following standard guidelines of Illumina Low Sample Protocol (TruSeq® RNA Sample Preparation v2 Guide). In brief, total RNA integrity following isolation was checked using an Agilent Technologies 2100 Bioanalyzer for each sample with an RNA Integrity Number (RIN) value greater than or equal to 8.The first step in the workflow involved purifying the poly-A containing mRNA molecules using poly-T oligo-attached magnetic beads. Following purification, the mRNA was fragmented into small pieces using divalent cations under elevated temperature. The cleaved RNA fragments were copied into first strand cDNA using reverse transcriptase and random primers. This was followed by second strand cDNA synthesis using DNA Polymerase I and RNase H. These cDNA fragments then went through an end repair process, the addition of a single 'A' base, and then ligation of the adapters. The products were then purified and enriched with PCR to create the final cDNA library.

This protocol for transcriptome analysis was performed on RNA after mRNA purification using elevated temperatures, resulting in libraries with insert size ranging from 120 to 200 bp with a median size of 150 bp. Transcriptome sequencing was carried out using Illumina Hiseq 2000 platform to generate paired-end reads.

5.3. Pre-processing and de novo assembly

Quality assessment of paired end *Bos frontalis* (mithun) RNA-seq data was performed by using FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The raw reads were filtered using Trimmomatic tool version 0.33 [70] to remove the low quality reads (phred-like Q-value ≤20 and reads with ambiguous bases 'N'), trimming of bases from 3' end and adaptor sequences. The high quality filtered reads of high growth and low growth (two biological replicates of both growth conditions) data of mithun were pooled together and *de novo* assembly was performed by three different methods using tools *namely*, *SOAPdenovo-trans* [71], *Spades version 3.9.0* [72] and *Trinity* v2.0.6 [73].

5.4. Transcription abundance and differential expressed genes analysis

Reads of HG and LG were separately mapped onto *de novo* assembled transcripts by using Bowtie [74] to calculate the read density, which provides the gap free alignment. The filtered reads were mapped to assembled transcripts with maximum two mismatches allowed. Abundance estimation of each transcript was performed using 'RNA-Seq by Expectation-Maximization (RSEM)' tool [75]. A fragment per kilobase of exon per million mapped reads (FPKM) method was used to calculate expression value of each transcript. Identification of differentially expressed genes from two different biological replicate samples of *Bos frontalis* (HG and LG) was carried out using *edgeR* package of Bioconductor [76]. In order to minimize the inaccuracy, significant DEGs were obtained with stringent parameters by taking four fold change and FDR < 0.05 [77].

5.5. Functional categorization and pathway analysis of differential expressed genes

Standalone local ncbi-blast-2.2.31+ was used for the homology search of differential expressed genes against NCBI non-redundant (version nr.36) database (ftp://ftp.ncbi.nlm.nih.gov/blast/db/) using BLASTX algorithm with default expected value to retrieve maximum possible information of genes. Gene ontology (GO), interproscan, functional categorization and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways analysis of DEG's of mithun was performed by using Blast2Go Pro version 3.1 software by importing the local blast results which contain the information of accession numbers, gene ID, gene description. Gene ontology (GO) further categorized DEGs into three sub categories *i.e.* biological process, molecular function and cellular component [78]. It also classified unigene into different pathway and provide enzyme classification number (EC).

5.6. Prediction of miRNA and TFs regulating DEGs

Prediction of transcriptional factor and cofactor from DEGs was performed using AnimalTFDB 2.0 (http://www.bioguo.org/AnimalTFDB/). Since most of the bovine miRNAs are highly conserved [79] thus all 793 mature miRNA sequences of *Bos taurus* were retrieved from miRBase database release 21 (http://www.mirbase.org/). Moreover, attempt for mithun specific miRNA discovery approach was not feasible in present transcriptomic data because of size selection during library preparation [80]. Target predictions of these miRNAs against DEGs of mithun were performed using miRanda-3.3a tool (http://www.microrna.org/microrna/getDownloads.do). Miranda aligned the *Bos taurus* miRNA sequences and DEGs to detect the potential miRNA targets with threshold of total score at 1000.

5.7. Detection of simple sequence repeats (SSR) markers

Mining of simple sequence repeat markers from mithun combined *de novo* transcriptome assembly as well as differential expressed genes was performed separately using MISA-MIcroSAtellite identification tool [81]. Ten repeating units for mono-, 6 repeating units for di- and five repeating units for tri-, tetra-, penta- and hexa nucleotides criteria were taken for detecting SSR markers. For validation of markers, the forward and reverse primers from the flanking regions of markers were generated using PRIMER3 software [82].

5.8. Variant detection

Variants were detected from reference as well as *de novo* transcriptome assembly. Since mithun genome assembly is in the form of draft assembly without having individual chromosomal molecule, thus reference genome of very same genus, *Bos* species cattle, *Bos taurus* version UMD3.1 was retrieved from ensembl database (http://asia.ensembl.org/*Bos*_taurus/Info/Index?redirect=no). In order to identify SNPs and InDels in these two divergent lines transcriptomic dataset, separately alignment of HG and LG was done over *de novo* transcriptome assembly. Burrows-Wheeler Aligner (BWA) tool [83] was used for alignment, followed by SNPs and Indels calling by SAMtools package [84]. Transitions and transversions in SNPs were obtained using BCFtools module of SAMtools. Annotation of SNPs and Indels were performed using *SnpEff* tool [85].

5.9. Gene regulatory network

Visualization of complex network and analysis of co-expression

interaction of 297 DEGs was performed using the open-source tool Cytoscape 3.2.1 [86]. It provides network in form of graph representing nodes and edges. Each node represents a gene and edges represent the interaction between genes. The gene co-expression matrix was constructed using Expression Correlation Plugin which is a robust program and based on Pearson correlation statistics. Highly interconnected clustered sub graphs were obtained using MCODE plugin [87]. GRN were constructed using DEGs and TFs. Network Analyzer program was used to compute and display the complete set of topological parameters containing; network diameter, clustering coefficient, heterogeneity, radius, density, centralization, average clustering coefficients, neighborhood connectivity, edges and sharing of nodes [88].

6. Author contribution

SM, AM and CR conceived theme of the study and generated data. SM, AM, IL, MM, KV, and KK collected samples and performed the wet lab work. RSJ, SJ, MAI and AR did the computational analysis of generated data. DK, MAI and SJ, drafted the manuscript. DK, MAI, SJ, AM, SM, and AR edited the manuscript. All co-authors read and approved the final manuscript.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2019.02.004.

Competing interests

The authors declare that they have no conflict of interests.

Acknowledgement

Authors are thankful to Indian Council of Agricultural Research, Ministry of Agriculture and Farmers Welfare, Govt. of India for funding of the research project and providing all facilities. Authors are also thankful to National Agricultural Innovation Project funded by World Bank for creating computational facility, ASHOKA at ICAR-IASRI, New Delhi, India where all analyses were performed. We also acknowledge the help provided by DBT Biotech Hub of the Institute to carry out some of the experiments.

References

[1] K.C. Nath, B.C. Deka, B.N. Borgohain, C.K. Rajkonwar, Semen characteristics in a captive Mithun (*Bos frontalis*), J. Zoo An. Med. 16 (1985) 62–64, https://doi.org/10.2307/20094743.
[2] F.S. Ahrestani, *Bos frontalis* and *Bos gaurus* (Artiodactyla: Bovidae), Mamm. Species 50 (959) (2018) 34–50.
[3] Q. Ren, et al., Complete mitochondrial genome of bovine species Gayal (*Bos frontalis*), Conserv. Genet. Resour. 10 (4) (2018) 889–891.
[4] M. Baig, et al., Mitochondrial DNA diversity and origin of *Bos frontalis*, Curr. Sci. (2013) 115–120.
[5] L. Sangpuii, M.A. Ali, L.I. Devi, Serum biochemical profile of Mizoram strain female Mithun (*Bos frontalis*) at different age and season, Biol. Rhythm. Res. (2018) 1–9.
[6] A. Mukherjee, et al., High-density genotyping reveals genomic characterization, population structure and genetic diversity of Indian Mithun (*Bos frontalis*), Sci. Rep. 8 (1) (2018) 10316.
[7] J.M. Romao, W. Jin, M. He, T. McAllister, L.L. Guan, MicroRNAs in bovine adipogenesis: genomic context expression and function, BMC Genomics 15 (2014) 137.
[8] M. Mondal, K.K. Baruah, C. Mithun Rajkhowa, An animal of Indian pride, Livest. Res. Rural. Dev. 26 (2014) 6.
[9] M.G. Shisode, et al., M.S. Mithun:the pride animal of north-eastern hilly region of India, Vet. World 2 (2009) 480–481.
[10] Y. Sun, et al., Isolation and cDNA characteristics of MHC-DRA genes from gayal (*Bos frontalis*) and gaytle (*Bos frontalis* × *Bos taurus*), Biotechnol. Biotechnol. Equip. 29 (2015) 33–39.
[11] O. Moyong, Commoditisation and sustainable Management of Mithun (*Bos frontalis*) in Arunachal Pradesh, north-East India, IOSR J. Agri. Vet. Sci. 1 (2012) 39–43.
[12] D. Xi, et al., Genetic variability of the coding region for the prion protein gene (PRNP) in gayal (*Bos frontalis*), Mol. Biol. Rep. 39 (2012) 2011–2020.
[13] J.K. Chamuah, et al., Seroprevalence of toxoplasma gondii in mithun (*Bos frontalis*) from north eastern hilly region of India, J. Parasit. Dis. 39 (2015) 560–562.
[14] M. Mondal, K.K. Baruah, C. Rajkhowa, Application of Biotechnology for Augmentation of Productivity in Mithun (Bos frontalis), Glob. J. Animal Sci. Res. 2 (2014) 4.
[15] National Research Centre on Mithun, Vision 2050, http://www.nrcmithun.res.in/publication/vision%202050.pdf, (2013) , Accessed date: 8 January 2019.
[16] A. Mukherjee, S. Mukherjee, S. Longkumer, C. Rajkhowa, Genetic parameters of growth and reproductive traits in mithun (*Bos frontalis*), Indian Vet. J. 90 (2013) 34–36.
[17] M.K. Abo-Ismail, et al., Single nucleotide polymorphisms for feed efficiency and performance in crossbred beef cattle, BMC Genet. 15 (2014) 14.
[18] A. Djari, et al., Gene-based single nucleotide polymorphism discovery in bovine muscle using next-generation transcriptomic sequencing, BMC Genomics 14 (2013) 307.
[19] M.-S. Wang, et al., Draft genome of the gayal, *Bos frontalis*, GigaScience 6 (2017) 1–7.
[20] A. Litwińczuk, M. Florek, P. Skałecki, Z. Litwińczuk, Chemical composition and physicochemical properties of horse meat from the longissimus lumborum and semitendinosus muscle, J. Muscle Foods 19 (3) (2008) 223–236.
[21] H.-M. Zhang, et al., Longissimus dorsi muscle transcriptomic analysis of Yunling and Chinese simmental cattle differing in intramuscular fat content and fatty acid composition, Genome 61 (8) (2018) 549–558.
[22] M. Girvan, M.E.J. Newman, Community structure in social and biological networks, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 7821–7826.
[23] S. Artavanis-Tsakonas, M.D. Rand, R.J. Lake, Notch signaling: cell fate control and signal integration in development, Science 284 (1999) 770–776.
[24] M.E. Mangan, J.B. Olmsted, A muscle-specific variant of microtubule-associated protein 4 (MAP4) is required in myogenesis, Development 122 (1996) 771–781.
[25] K. Wang, C. Wang, F. Xiao, H. Wang, Z. Wu, JAK2/STAT2/STAT3 are required for myogenic differentiation, J. Biol. Chem. 283 (2008) 34029–34036.
[26] D. Chen, W. Li, M. Du, M. Wu, B. Cao, Sequencing and characterization of divergent marbling levels in the beef cattle (Longissimus dorsi Muscle) Transcriptome, Asian-Australasian J. Anim. Sci. 28 (2015) 158–165.
[27] M.G. Carolina, et al., The PPP6R3/LRP5 locus influences lean mass in children of different ethnic background and highlights pleiotropic effects and muscle-bone interactions, Bone Abstr. 1 (S2) (2012) 1.
[28] T. Sadkowski, M. Jank, L. Zwierzchowski, J. Oprzadek, T. Motyl, Comparison of skeletal muscle transcriptional profiles in dairy and beef breeds bulls, J. Appl. Genet. 50 (2009) 109–123.
[29] D. Shungin, et al., New genetic loci link adipose and insulin biology to body fat distribution, Nature 518 (2015) 187–196.
[30] C. Wei, et al., Identification of the crucial molecular events during the large-scale myoblast fusion in sheep, Physiol. Genomics 46 (2014) 429–440.
[31] A. Canovas, R. Quintanilla, M. Amill, R.N. Pena, Muscle transcriptomic profiles in pigs with divergent phenotypes for fatness traits, BMC Genomics 11 (2010) 372.
[32] J.S. Moylan, et al., Neutral sphingomyelinase-3 mediates TNF-stimulated oxidant activity in skeletal muscle, Redox Biol. 2 (2014) 910–920.
[33] I.S. MiddelBos, B.M. Vester, L.K. Karr-Lilienthal, L.B. Schook, K.S. Swanson, Age and diet affect gene expression profile in canine skeletal muscle, PLoS ONE 4 (2009) e4481.
[34] W. Song, Z. Yang, B. He, Bestrophin 3 ameliorates TNFα-induced inflammation by inhibiting NF-κB activation in endothelial cells, PLoS ONE 9 (2014) e111093.
[35] A. Mull, G. Kim, J.M. Holaska, LMO7-null mice exhibit phenotypes consistent with emery-dreifuss muscular dystrophy, Muscle Nerve 51 (2015) 222–228.
[36] F. Paradis, et al., Transcriptomic analysis by RNA sequencing reveals that hepatic interferon-induced genes may be associated with feed efficiency in beef heifers, J. Anim. Sci. 93 (2015) 3331–3341.
[37] M. Zorc, et al., Catalog of genetic variants within mature microRNA seed regions in chicken, Poult. Sci. 94 (2015) 2037–2040.
[38] S.M. Peterson, J.A. Thompson, M.L. Ufkin, P. Sathyanarayana, L. Liaw, C.B. Congdon, Common features of microRNA target prediction tools, Front. Genet. 5 (2014) 23.
[39] S.D.E. Park, et al., Genome sequencing of the extinct Eurasian wild aurochs, *Bos primigenius*, illuminates the phylogeography and evolution of cattle, Genome Biol. 16 (2015) 234.
[40] X. Wang, Z. Gu, H. Jiang, MicroRNAs in farm animals, Animal 7 (2013) 1567–1575.
[41] C. Anderson, H. Catoe, R. Werner, MIR-206 regulates connexin43 expression during skeletal muscle development, Nucleic Acids Res. 34 (2006) 5863–5871.
[42] W. Barendse, et al., A validated whole-genome association study of efficient food conversion in cattle, Genetics 176 (2007) 1893–1905.
[43] C. Liu, et al., MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs, BMC Genomics 13 (2012) 661.
[44] Y. Wang, et al., Identification and profiling of microRNAs and their target genes from developing caprine skeletal muscle, PLoS ONE 9 (2014) e96857.
[45] B. Zheng, X. Yu, R. Chai, Myotubularin-related phosphatase 3 promotes growth of colorectal cancer cells, Sci. World J. 703804 (2014) 2014.
[46] L. Sansregret, A. Nepveu, The multiple roles of CUX1: insights from mouse models and cell-based assays, Gene 412 (2008) 84–94.
[47] L. Goyvaerts, et al., Prolactin receptors and placental lactogen drive male mouse pancreatic islets to pregnancy-related mRNA changes, PLoS ONE 10 (2015) e0121868.
[48] B. Saha, A.R. Ypsilanti, C. Boutin, H. Cremer, A. Chédotal, Plexin-B2 regulates the proliferation and migration of neuroblasts in the postnatal and adult subventricular zone, J. Neurosci. 32 (2012) 16892–16905.
[49] C. Guo, et al., KMT2D maintains neoplastic cell proliferation and global histone H3 lysine 4 monomethylation, Oncotarget 4 (2013) 2144–2153.
[50] J.L. Chitwood, G. Rincon, G.G. Kaiser, J.F. Medrano, P.J. Ross, RNA-seq analysis of single bovine blastocysts, BMC Genomics 14 (2013) 350.
[51] N. Hermann-Kleiter, et al., The nuclear orphan receptor NR2F6 is a central

checkpoint for cancer immune surveillance, Cell Rep. 12 (2015) 2072–2085.
[52] K. Suzuki-Utsunomiya, et al., ALS2CL, a novel ALS2-interactor, modulates ALS2-mediated endosome dynamics, Biochem. Biophys. Res. Commun. 354 (2007) 491–497.
[53] A.Y.O.M. Carvalho, et al., The role of vitamin D level and related single nucleotide polymorphisms in Crohn's disease, Nutrients 5 (2013) 3898–3909.
[54] C.G. Joseph, et al., Association of the autoimmune disease scleroderma with an immunologic response to cancer, Science 343 (2014) 152–157.
[55] M. Yang, Z.-M. Yuan, A novel role of PRR14 in the regulation of skeletal myogenesis, Cell Death Dis. 6 (2015) e1734.
[56] T. Endo, T. Masaki, Differential expression and distribution of chicken skeletal- and smooth-muscle-type alpha-actinins during myogenesis in culture, J. Cell Biol. 99 (1984) 2322–2332.
[57] J. Dias, et al., Structural analysis of the KANSL1/WDR5/KANSL2 complex reveals that WDR5 is required for efficient assembly and chromatin targeting of the NSL complex, Genes Dev. 28 (2014) 929–942.
[58] R. Groisman, et al., Physical interaction between the mitogen-responsive serum response factor and myogenic basic-helix-loop-helix proteins, J. Biol. Chem. 271 (1996) 5258–5264.
[59] T.G. Edwards, T.J. Vidmar, K. Koeller, J.K. Bashkin, Fisher, C. DNA damage repair genes controlling human papillomavirus (HPV) episome levels under conditions of stability and extreme instability, PLoS ONE 8 (2013) e75406.
[60] P. Zhang, et al., TRIM25 has a dual function in the p53/Mdm2 circuit, Oncogene 34 (2015) 5729–5738.
[61] J.B. Knight, C.A. Eyster, B.A. Griesel, A.L. Olson, Regulation of the human GLUT4 gene promoter: interaction between a transcriptional activator and myocyte enhancer factor 2A, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 14725–14730.
[62] C.S. Pareek, U. Czarnik, M. Pierzchała, L. Zwierzchowski, An association between the C > T single nucleotide polymorphism within intron IV of osteopontin encoding gene (SPP1) and body weight of growing polish Holstein-Friesian cattle, Anim. Sci. Paper Rep. 26 (2008) 251–257.
[63] D. Lu, et al., Genome-wide association analyses for carcass quality in crossbred beef cattle, BMC Genet. 14 (2013) 80.
[64] D.B.S. Silva, et al., Genetic variations in the leptin gene associated with growth and carcass traits in Nellore cattle, Genet. Mol. Res. 13 (2014) 3002–3012.
[65] E.R. Chung, S.C. Shin, J.P. Heo, Association between SNP marker of uncoupling protein 3 gene and meat yield and marbling score traits in Korean cattle, Korean J. Food Sci. Anim. Resour. 31 (2011) 530–536.
[66] J.L. Gill, S.C. Bishop, C. McCorquodale, J.L. Williams, P. Wiener, Association of selected SNP with carcass and taste panel assessed meat quality traits in a commercial population of Aberdeen Angus-sired beef cattle, Genet. Sel. Evol. 41 (2009) 36.
[67] M.F. Rothschild, G.S. Plastow, Applications of genomics to improve livestock in the developing world, Livest. Sci. 166 (2014) 76–83.
[68] J.L. Williams, et al., Genome assembly and transcriptome resource for river buffalo, Bubalus bubalis (2 n = 50), GigaScience 6 (10) (2017) 1–6.
[69] M.R.F. Lee, P.R. Evans, G.R. Nute, R.I. Richardson, N.D. Scollan, A comparison between red clover silage and grass silage feeding on fatty acid composition, meat stability and sensory quality of the M. Longissimus muscle of dairy cull cows, Meat Sci. 81 (2009) 738–744.
[70] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (15) (2014) 2114–2120.
[71] Y. Xie, et al., SOAPdenovo-trans: de novo transcriptome assembly with short RNA-Seq reads, Bioinformatics 30 (12) (2014) 1660–1666.
[72] A. Bankevich, et al., SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, J. Comput. Biol. 19 (5) (2012) 455–477.
[73] B.J. Haas, et al., De novo transcript sequence reconstruction from RNA-seq using the *Trinity* platform for reference generation and analysis, Nat. Protoc. 8 (8) (2013) 1494.
[74] B. Langmead, C. Trapnell, M. Pop, S.L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, Genome Biol. 10 (3) (2009) R25.
[75] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, BMC Bioinformatics 12 (1) (2011) 323.
[76] M.D. Robinson, D.J. McCarthy, G.K. Smyth, *edgeR*: a bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (1) (2010) 139–140.
[77] C.R. Yendrek, E.A. Ainsworth, J. Thimmapuram, The bench scientist's guide to statistical analysis of RNA-Seq data, BMC Res Notes. 5 (2012) 506.
[78] A. Conesa, et al., Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, Bioinformatics 21 (2005) 3674–3676.
[79] W. Jin, J.R. Grant, P. Stothard, S.S. Moore, L.L. Guan, Characterization of bovine miRNAs by sequencing and bioinformatics analysis, BMC Mol. Biol. 10 (2009) 90.
[80] Z. Miao, et al., De novo transcriptome analysis of Medicago falcata reveals novel insights about the mechanisms underlying abiotic stress-responsive pathway, BMC Genomics 16 (2015) 818.
[81] T. Thiel, W. Michalek, R. Varshney, A. Graner, Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.), Theor. Appl. Genet. 106 (3) (2003) 411–422.
[82] A. Untergasser, I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, S.G. Rozen, Primer3—new capabilities and interfaces, Nucleic Acids Res. 40 (15) (2012) e115-e115.
[83] H. Li, R. Durbin, Fast and accurate short read alignment with burrows–wheeler transform, Bioinformatics 25 (14) (2009) 1754–1760.
[84] H. Li, et al., The sequence alignment/map format and SAMtools, Bioinformatics 25 (16) (2009) 2078–2079.
[85] P. Cingolani, et al., A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3, Fly (Austin). 6 (2012) 80–92.
[86] P. Shannon, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (2003) 2498–2504.
[87] G.D. Bader, C.W.V. Hogue, An automated method for finding molecular complexes in large protein interaction networks, BMC Bioinformatics 4 (2003) 2.
[88] Y. Assenov, F. Ramírez, S.-E. Schelhorn, T. Lengauer, M. Albrecht, Computing topological parameters of biological networks, Bioinformatics 24 (2008) 282–284.