Animal Biotechnology

ISSN: 1049-5398 (Print) 1532-2378 (Online) Journal homepage: https://www.tandfonline.com/loi/labt20

Analysis and Functional Annotation of Expressed Sequence Tags of Water Buffalo

Garima Bajetha , Jyotika Bhati , Sarika , M. A. Iquebal , Anil Rai , Vasu Arora & Dinesh Kumar

To cite this article: Garima Bajetha , Jyotika Bhati , Sarika , M. A. Iquebal , Anil Rai , Vasu Arora & Dinesh Kumar (2013) Analysis and Functional Annotation of Expressed Sequence Tags of Water Buffalo, Animal Biotechnology, 24:1, 25-30, DOI: 10.1080/10495398.2012.737884

To link to this article: https://doi.org/10.1080/10495398.2012.737884

Animal Biotechnology, 24: 25–30, 2013

ISSN: 1049-5398 print/1532-2378 online
DOI: 10.1080/10495398.2012.737884

Taylor & Francis
Taylor & Francis Group

ANALYSIS AND FUNCTIONAL ANNOTATION OF EXPRESSED SEQUENCE TAGS OF WATER BUFFALO

Garima Bajetha[1], Jyotika Bhati[1], Sarika[1], M. A. Iquebal[2], Anil Rai[1], Vasu Arora[1], and Dinesh Kumar[1,3]

[1]*Center for Agricultural Bioinformatics, Indian Agricultural Statistics Research Institute, New Delhi, India*

[2]*Division of Biometrics & Statistical Modeling, Indian Agricultural Statistics Research Institute, New Delhi, India*

[3]*Genes & Genetic Resources Molecular Analysis Lab, National Bureau of Animal Genetic Resources, Karnal, Haryana, India*

An elucidated genome of domestic livestock river buffalo will contribute enormously to economy and better understanding of genome evolution as well. An attempt is made to obtain genomic information on buffalo, based on total Expressed Sequence Tags (ESTs) of* Bubalus bubalis *available in public domain. These ESTs were annotated and classified into 15 different functional categories based on their homology to the known proteins. Interestingly, 41.79% of the contigs were found to be buffalo specific novel ESTs with respect to other species used in analysis which needs further studies. Also, 224 pSNPs (putative Single Nucleotide Polymorphism) were detected. This study will provide a home base for further genomic studies of buffalo and comparative studies enabling a starting point for the genome annotation of the organism. Supplementary materials are available for this article online.

Keywords: Expressed sequence tags; Gene ontology; Markers; Water buffalo

Water buffalo (*Bubalus bubalis*) contributes immensely to the agricultural economy of Indian subcontinent, South East Asian countries through milk, meat, hides, fertilizer, fuel, and draught animal power. A large part of human population depends on this species than any other livestock species in the world (1). Buffaloes are in the family *Bovidae* with a diploid count of chromosomes being 48 (swamp) and 50 (riverine) (2). EST data are important resource for gene identification and genome annotation (3).

The sequencing and analysis of expressed sequence tags (ESTs) has been a primary tool for the discovery of novel genes in animals, especially in non-model

Grants from the National Agricultural Innovation Project (NAIP), the Indian Council of Agricultural Research, Ministry of Agriculture, Government of India, New Delhi entitled ''Establishment of National Agriculture Bioinformatics Grid in ICAR'' is gratefully acknowledged. We would like to express our gratitude to the two anonymous reviewers and Chief Editor for helpful comments and guidance.

Address correspondence to Sarika, Center for Agricultural Bioinformatics, Indian Agricultural Statistics Research Institute, Library Avenue, New Delhi-110012, India. E-mail: sarika@iasri.res.in or aijaiswal@gmail.com

animals for which full genome sequences are not currently available. The developments in biotechnological facilities led to a rise in ESTs, which provides genetic information to the scientific community for comparative and functional genomic studies as well as helps in annotation of genomic sequences. Detection of SNPs (Single Nucleotide Polymorphisms) in the ESTs of buffalo may help in designing a new set of genomic DNA-based markers and may provide more insight in the evolution of this species. ESTs available in public domain for *Bubalus bubalis* have been taken for *in-silico* studies, including detection of SNPs and functional annotation.

MATERIALS AND METHODS

A total of 1825 ESTs deposited until March 2012 in the public database NCBI for *Bubalus bubalis* were downloaded in FASTA format for EST clustering, functional annotation, SNP mining, and identification of protein domains. The available ESTs of Buffalo were clustered into a nonredundant set of contigs using EGAssembler (4). Contigs obtained from alignment of ESTs were analyzed for homology searches in existing *Bos taurus* sequence databases. The sieved hits were searched against a nonredundant database using BLASTX (5) for annotation. Genes obtained from the study were compared to those of *Homo sapiens, Mus musculus, Rattus norvegicus, Bos taurus,* and *Gallus gallus* using HomoloGene (http://www.ncbi.nlm.nih.gov/homologene). All raw EST sequences were analyzed for putative SNPs using SNPFinder (6) with reference to *Bos Taurus* unigene which was batch downloaded from NCBI. Further, the resulting protein sequence were analyzed to find protein domains. The contigs were compared in Batch Web CD-Search tool (7) and Pfam database.

RESULTS AND DISCUSSION

The available 1825 EST sequences for buffalo were downloaded from GenBank including dbESTs and clustered into 151 contigs, and less-abundant/expressed transcripts remained as 1060 singletons. Using *Bos taurus* gene indices, a comparative analysis of *Bos taurus* with *Bubalus bubalis* contigs exhibited high level of similarity. These annotated genes were classified into 15 different functional categories (Fig. 1) based on their homology to known proteins. Out of total contigs, 58.21% could be assigned with biological functions. The most predominant functional category of contigs was *metabolism and energy* consisting of 12.69% of the total contigs. Though such analysis has not been performed in case of *Bubalus bubalis*, but in *Camelus dromedarius,* assembled EST sequences with 16.83% transcripts were involved in *metabolism and energy* processes. Further, the second dominant functional category was cellular transport which consisted of 11.94% of the total contigs, which was also reported in camel (8). Other dominant functional categories are structural/catalytic protein, protein synthesis, transcription factor, and cell cycle and DNA processing. Interestingly, the remaining 41.79% were found to be buffalo specific novel ESTs with respect to other species used in analysis which needs further study. EST sequences of buffalo that showed homology with genes of *M. musculus, R.norvegicus, B.taurus, H. sapiens,* and *G.gallus* are represented as Venn diagram along with graphical delineation of shared genes in terms of functions (Fig. 2a and 2b).

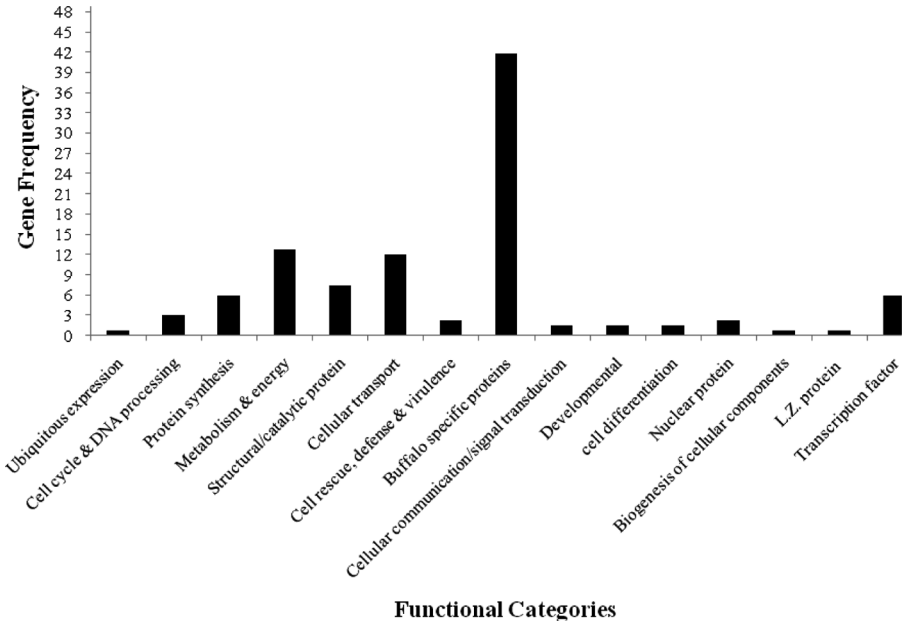

Figure 1 Frequency of genes in different functional categories based on EST analysis with five species (*M. musculus, R.norvegicus, B.taurus, H. sapiens, G.gallus*).

In the present study, 224 pSNP were detected from all 151 contigs (Table 1), which can be used as molecular markers for breeding. All predicted contigs were further analyzed for their domain structure using the Pfam database (9). Out of 151 contigs only 42 contigs with PSSM-ID of conserved domains were found. Most

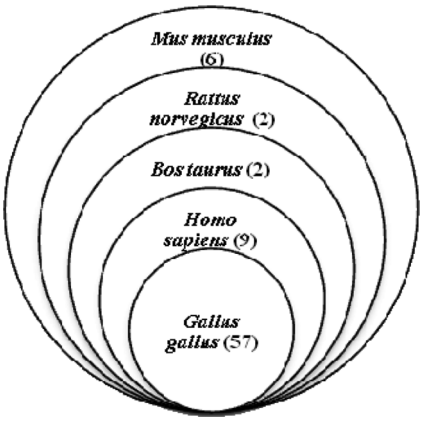

Figure 2 Venn diagram for shared genes in five species, viz., *M. musculus, R.norvegicus, B.taurus, H. sapiens,* and *G.gallus* (figures in Venn diagram (57, 9, 2, 2, 6) has been derived by taking relatively omni-present numbers in corresponding five species). These shared genes are distributed into the aforementioned 26 functions.

Table 1 Contigs with more than five SNPs and their Gene Ontology IDs along with major categories of functions

Contig No.	SNPs	Position	GO IDs assigned (NA = no GO categories assigned)	Major categories of functions		
				Molecular Function	Biological Process	Cellular Component
Contig 3	5	166, 261, 265, 272, 274	NA	NA	NA	NA
Contig 23	6	33, 127, 337, 373, 396, 438	NA	NA	NA	NA
Contig 27	12	174, 202, 223, 280, 295, 310, 313, 325, 328, 388, 397, 406	6796, 5737, 287, 16787, 4427, 71344, 46872, 16462	Hydrolase, Magnesium ion binding	Pyrophosphate metabolism	All of the contents of a cell excluding the plasma membrane and nucleus, but including other subcellular structures
Contig 30	5	182, 210, 332, 377, 460	5737, 5634, 3674, 8150	Elemental activities like catalysis or binding	Physiological process	Cell nucleus
Contig 32	5	20, 93, 122, 125, 150	45449, 3713, 5634, 4674	Transcription co-activator activity, Protein serine/threonine kinase activity	Regulation of cellular transcription	Cell nucleus
Contig 53	6	85, 111, 114, 156, 221, 323	NA	NA	NA	NA
Contig 61	10	67, 77, 105, 152, 164, 175, 196, 200, 226, 391	5524, 5737, 166, 6468, 9851, 5515, 4674, 4672, 16301	Protein serine/threonine kinase activity, Nucleotide and amino acid binding	Auxin biosynthetic process, Protein amino acid phosphorylation	Cell nucleus
Contig 82	9	246, 255, 368, 387, 433, 441, 448, 467, 537	NA	NA	NA	NA
Contig 93	6	461, 465, 521, 530, 594, 599	1824, 45944, 45449, 5515, 3713, 16592, 5634, 3702, 6357, 3899, 6350, 6366	Protein amino acid binding, Deoxyribonucleic acid-dependent ribonucleic acid polymerase activity	Regulation of transcription	Cell nucleus
Contig 95	13	50, 133, 380, 424, 549, 559, 580, 616, 700, 782, 969, 1045, 1053	5576, 6810, 5215, 8191, 4857, 16209, 8217	Small-molecule carrier or transporter, Metalloprotease inhibitor	Auxiliary transport protein activity	Extracellular
Contig 97	8	17, 66, 78, 153, 189, 215, 228, 285	NA	NA	NA	NA

Contig 99	9	36, 53, 76, 110, 186, 191, 236, 244, 375	5576, 6810, 5215, 42742	Small-molecule carrier or transporter	Auxiliary transport protein activity, Defence response to bacteria	Extracellular
Contig 101	10	28, 83, 98, 128, 136, 203, 218, 238, 267, 330	NA	NA	NA	NA
Contig 112	6	24, 117, 231, 264, 284, 302	NA	NA	NA	NA
Contig 121	6	358, 385, 400, 457, 496, 506	8083, 3779, 5622, 8047, 4860, 6468, 19207	Membrane associated actin binding, Protein kinase inhibitor activity	Protein amino acid phosphorylation	Internal to cell
Contig 126	7	27, 60, 72, 89, 116, 158, 314	48471, 5737, 16192, 6888, 6350, 5783, 5794, 5622, 6810, 1501, 8134, 6355, 5515, 30008, 8150	TF binding, Protein amino acid binding	Transport activity, Cellular transcription, Skeletal development	Golgi complex, Endoplasmic reticulum, Internal to cell
Contig 135	8	49, 123, 290, 300, 308, 334, 338, 341	45121, 43499, 7165, 42110, 51789, 5515, 9897, 6952, 45576, 43234, 3823, 46658, 5887, 5886	Protein amino acid binding	T lymphocyte activation	Outer surface of cytoplasmic membrane, Cell membrane, Cellular Component
Contig 137	7	39, 79, 137, 208, 283, 299, 317	NA	NA	NA	NA
Contig 138	8	38, 45, 65, 87, 134, 227, 234, 335	NA			
Contig 140	10	15, 31, 40, 119, 322, 421, 550, 607, 625, 649	42254, 5515, 5634, 5730, 3674, 8150, 5575	Protein amino acid binding	Ribosome biogenesis and assembly	Cell nucleus, Cellular Component
Contig 145	6	30, 41, 80, 250, 286, 442	NA	NA	NA	NA
Contig 151	13	9, 21, 34, 84, 90, 101, 143, 233, 237, 282, 289, 329, 381	NA	NA	NA	NA

of the detected domains matched perfectly with the obtained BLAST results against *Bos taurus* (10), *Gallus gallus* (11), *Mus musculus* (12), *Rattus norvegicus* (13), and *Homo sapiens*.

To conclude, exhaustive analysis of available ESTs will prove valuable for future research for better knowledge of buffalo and the whole lineage of *Bovidae*. A genome scale comparison of cattle with buffalo at the sequence level provides an excellent opportunity to find some economically important genes, to clone and use them in breeding programs of livestock. This is the first attempt to obtain more genomic and transcript information of buffalo. Also, the reported pSNPs would be useful for generating genetic and physical maps, study of genetic diversity, marker-assisted selection and even positional cloning of useful genes. Comparative genomics can facilitate the study of the evolution of sequences and functions of orthologous genes and also to understand diversification and adaptation.

REFERENCES

1. Scherf BD. *World Watch List for Domestic Animal Diversity*. 3rd ed. Rome, Italy: Food and Agriculture Organization of the United Nations, Communication Division, FAO; 2000.
2. Fischer H, Ulbrich F. Chromosome of the Murrah Buffalo and its crossbreds with the Asiatic Swamp Buffalo. (Bubalus bubalis) Z Tierzücht Züchtungsbiol 1967; 84:110–114.
3. Liang F, Holt I, Pertea G, Karamycheva S, Salzberg SL, Quackenbush J. Gene index analysis of the human genome estimates approximately 120,000 gene. Nature Genet 2000; 25:239–240.
4. Masoudi-Nejad A, Tonomura K, Kawashima S, et al. EGassembler: Online bioinformatics service for largescale processing, clustering and assembling ESTs and genomic DNA fragments. Nucleic Acids Res 2006; 34:W462–W469.
5. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Molecul Biol 1990; 215:403–410.
6. Song J, Xu Y, White S, Miller KWP, Wolinsky M. SNPsFinder - a web-based application for genome-wide discovery of single nucleotide polymorphisms in microbial genomes. Bioinformatics 2005; 21:2083–2084.
7. Marchler-Bauer A, Anderson JB, Chitsaz F, et al. CDD: Specific functional annotation with the Conserved Domain Database. Nucleic Acids Res 2009; 37:D205–D210.
8. Al-Swailem AM, Shehata MM, Abu-Duhier FM, et al. Sequencing, analysis, and annotation of expressed sequence tags for *Camelus dromedarius*. PLoS ONE 2010; 5(5):e10720.
9. Finn RD, Mistry J, Schuster-Böckler B, et al. Pfam: Clans, web tools and services. Nucleic Acids Res 2006; 34:D247–D251.
10. Abdel-Rahman S M. Evidences reveal that cattle and buffalo evolutionary derived from the same ancestor based on cytogenetic and molecular markers. Biotechnol Anim Husb. 2006; 22(3–4):1–10.
11. Berthouly C, Leroy G, Nhu Van T, et al. Genetic analysis of local Vietnamese chickens provides evidence of gene flow from wild to domestic populations. BMC Genet 2009; 10:1.
12. Snijders AM, Nowak NJ, Huey B, et al. Mapping segmental and sequence variations among laboratory mice using BAC array CGH. Genome Res 2005; 15:302–311.
13. Wang Y, Bruenn JA, Queener SF, Cody V. Isolation of rat dihydrofolate reductase gene and characterization of recombinant enzyme. Antimicrob Agents Chemother 2001; 45(9):2517–2523.