

# COMPUTATIONAL ANNOTATION AND ANALYSIS OF WATER BUFFALO (*BUBALUS BUBALIS*) MITOCHONDRIAL GENOME

Patel Rajesh<sup>1</sup>, Himani Gandhi<sup>2</sup>

*Department of Life Sciences, Hem. North Gujarat University, Patan (India)*

## ABSTRACT

*The comparison of complete organelle genome sequences is becoming increasingly important for many tasks, including reconstructing the evolutionary relationships of organisms and understanding the inheritance. In the present work, an attempt has been carried out to annotated mitochondrial genome of Water Buffalo (*Bubalus bubalis*). Genome annotation compared with another bovine mitochondrial genome. The genome analyzed for whole genome alignment and codon usage and also explored for mapping the EST under lactating and non-lactating. The results of comparative analysis of the mitochondrial genome revealed activation of ATP8, ATP6, COX3 (subunit 3 of cytochrome) and COB (Apocytochrome B) during lactation. The presence of such genes indicated the mammary mitochondrial ATP synthesis activity during lactation. Genome comparison exhibits 98% similarity for genome sequences used in the study. The mtDNA gene pool was found to be more closely related to similarity in *Bubalis bubalis* and *Bostaurus*.*

**KEYWORDS:** *EST Analysis, Mitochondrial Genome Analysis*

## I. INTRODUCTION

Water buffalo (*Bubalus bubalis*) contributes very larger to the agricultural economy of the Indian subcontinent, South East Asian countries through milk, meat, hides, fertilizer, fuel and draught animal power. A large part of the human population depends on this species than any other livestock species in the world. There is 188.3 million buffalo population in the world which contributes around 55 – 60% of total milk production[1]. Asia has nearly 97% of buffaloes and is an integral part of agriculture in India, China, Pakistan, Nepal, Bangladesh, Thailand, Myanmar, and Malaysia. The productivity of buffaloes in these regions is higher as compared to cattle[2]. Several international organizations have emphasized the potential of buffaloes in the agricultural economy of many developing nations in Asia, but domestic buffalo continues to be neglected despite its superiority over cattle under the harsh environments of the tropics[3]. Genome analysis of water buffalo has advanced significantly in recent years. A total of 66,935 nucleotide sequences for the water buffalo have been deposited in the GenBank database and are mainly 64,212 whole genome shotgun sequences, while the rest includes 974 mitochondrial genomic sequences and 1,748 nuclear genes/genomic DNA sequences [4].The mitochondrial genome analysis is likely to be found significant importance in the ATP generation[5]. The water buffalo mt molecule is 16.355 base pair length and shows a genome organization similar to those reported for another mitochondrial genome. These new data provide a useful tool for much research area, i.e.

evolutionary study and identification of food origin[4]. Along with that, Gene expression-based approaches are useful research tools for several studies directed to elucidate the broad range of aspects of animal physiology and pathology the different types of water buffaloes[6]. Although, the resources like cytogenetic characterization, whole genome mapping and next-generation sequencing of *Bubalus bubalis* revealed significant findings, newer technologies and approach in computational biology help to optimize production potential, reproduction efficiency, product quality, nutritional value and resistance to diseases. The present study was carried out to find out the hidden truth of the mitochondrial genome of *Bubalis bubalis* using *in silico* approach. EGAssembler (<http://egassembler.hgc.jp/>) is a web server, which provides an automated as well as a user-customized analysis tool for cleaning, repeat masking, vector trimming, organelle masking, clustering and assembling ESTs and genomic fragments[7].

The Dual Organellar GenoMe Annotator (DOGMA) automates the annotation of organelle (plant chloroplast and animal mitochondrial) genomes. It is a Web-based package that allows the use of BLAST searches against a custom database and conservation of base pairing in the secondary structure of animal mitochondrial tRNAs to identify and annotate genes. DOGMA provides a graphical user interface for viewing and editing annotations. Annotations were stored on password protected server to enable repeated sessions of working on the same genome[8]. ACUA has got the potential to serve as a comprehensive platform, and also as a part of biologist's essential toolkit to extract all the prerequisite data required for statistical analysis of codon usage[9].

## II. MATERIALS AND METHODS

### 2.1 Collection of Mitochondrial Genome Data

Mitochondrial genome data for Genome browser for the Water Buffalo (*Bubalus bubalis*) Pure Murrah Buffalo available at <http://210.212.93.84/Bubalusbubalis> (water buffalo) and (<http://www.ncbi.nlm.nih.gov/nuccore/33334119>) India, GI 33334119 *Bubalus bubalis* (water buffalo) India, GI 347801295 *Bubalus bubalis* (water buffalo) France, GI 52220982 *Bubalus bubalis* (Swamp buffalo) China, GI 44894095 *Bubalus bubalis* (water buffalo), Italy and GI 51173180 *Bubalus bubalis* (Swamp buffalo) China collected from the NCBI.

### 2.2 Quality Trimming

To increase the accuracy of the data, the quality trimming was done using EGAssembler software using its default parameters[7].

### 2.3 Mitochondrial Genome Annotation and Whole Genome Alignment

Mitochondrial genome annotation was carried out by Dual organelle genome annotator (DOGMA) as described by Wyman, 2004. CLUSTALW did multiple sequence alignment. While, entire genome alignment performed by Mauve software using instruction provided by the software provider[10].

## 2.4 Condon Usage Analysis

Codon usage analysis was carried out using ACUA (Automated Codon Usage Tool) as per instruction provided by the software provider[9].

## 2.5 Mapping of Differentially Expressed EST During Lactation

Differentially expressed EST sequences *Bubalus bubalis* collected from EST database and mapping was carried out by using EST2GENOME available in EMBOSS suite for finding homology-based similarity for EST and Genome[11].

## III. RESULTS AND DISCUSSION

### 3.1 Collection of Mitochondrial Genome Data

In the present study, complete mitochondrial genome data of six different *Bubalus bubalis* breed was taken. Out of total six two sequences were of Indian origin. One was from Pure Murrah Buffalo (Water Buffalo scientific name *Bubalus bubalis*) was taken from <http://210.212.93.84> generated by National Bureau of Animal Genetic Resources (NBAGR) in Karnal and another collaborator. While another Indian sequence collected with GI 333334119 submitted from CCMB, Hyderabad India. Besides that two other sequences for *Bubalus bubalis* (water buffalo) with accession no. GI 347801295 and GI 44894095 was from France and Italy respectively. Remaining two other sequences with GI 52220982 and GI 51173180 were for *Bubalus bubalis* (Swamp buffalo) from China. (Table 1). The complete water buffalo mitochondrial genome was sequenced by three groups of scientists at Hainan Medical College, China (Acc. No. AY702168), Centre for Cellular and Molecular Biology, Hyderabad, India (Acc. No. AF547270) and Istituto Spallanzani, Italy (Acc. No. number. AY488491).[4]

### 3.2 Mitochondrial Genome Annotation

The comparison of complete organellar genome sequences is becoming increasingly important for many tasks, including reconstructing the evolutionary relationships of organisms and understanding the inheritance of certain human diseases. In the past, annotating organellar genomes has been time-consuming and error-fraught process and, with the input of high-throughput genome sequencing centers, has been the rate-limiting step in the production of complete chloroplast and mitochondrial genome sequences. DOGMA is the first tool that automates this process. Mitochondrial genomes submitted in FAST format to DOGMA. Annotation visualized in the annotation panel of DOGMA. For the entire *Bubalus bubalis* mitochondrial genome total, 41 genes were reported. Genes were also found to have similar locus also. In comparison to *Bubalus bubalis* variation was observed for *Bostaurus* trnS-gga and trnV-cac gene were absent and trnI-gau, trnS-gcu, trnN-auu, trnT-uguwere found in different loci only in *Bos taurus* only. (Table: 2) The essential gene for more exploration based on the presence of the gene for Lactoferrin enzyme pathway. This enzyme had found significant

importance in the industrial applications. In *Bubalis bubalis*, this gene was observed with the size of 2,127bp. While sequence can found in NCBI with Acc.No. AJ005203. [12]

### 3.3 Whole-Genome Annotation

Whole-genome annotation was carried out by eight different mitochondrial genomes from *Bubalus bubalis* NBAGR, INDIA, *Bosindicus* 33321647, *Bostaurus* 42521312, *Bubalus bubalis* 51173180 China, *Bubalus bubalis* 52220982 China, *Bubalus bubalis* 347801295 France, *Bubalus bubalis* 33334119 CCMB, INDIA and *Bubalus bubalis* 44894095 Italy. All the mitochondrial genomes were found to align significantly, and only gene rearrangement obtained with two scores only. For both the water buffalo from India, no gene rearrangement was recorded for the swamp buffalo from China gene rearrangement were evident. (Figure: 1) According to the study carried out by *Vanessa et al.*, 2010 13 coding mitochondrial genes in water buffalo, the length of the coding sequences of four genes – *COX1*, *ND4L*, *ND4* and *ND6* are identical to those in *Bison bison* (American bison, NC\_012346), *Bosgrunniens* (domestic yak, NC\_006380), *Bos indicus* (zebu, AF492350), *Bos taurus* (domestic cattle, NC\_006853), *Capra hircus* (goat, NC\_005044), *Equusasinus* (donkey, NC\_001788), *Equus caballus* (horse, NC\_001640), *Lama glama* (llama, NC\_012102), *Oryctolagusuniculus* (rabbit, NC\_001913), *Ovis aries* (sheep, NC\_001941) and *Sus scrofa* (swine, NC\_000845).[4] Along with that, the D-loop region (926 bp) of water buffalo was found somewhat small in size comparative to 888 bp of *B. bison* to 1800 bp in *O. cuniculus*[13]. While, the comparison of sixteen genes overlap in the water buffalo mitochondrial genome, including two three-gene-overlaps (*ND1/tRNA-Ile/tRNA-Gln ATP8/ATP6/COX3*), and five two-gene-overlaps *ND2/tRNA-Trp*, *tRNA-Tyr/COX1*, *ND4L/ND4*, *ND5/ND6* and *tRNA-Thr/tRNA-Pro*. [14]

### 3.4 Codon Usage Analysis

For comparative Codon usage analysis of all the *Bubalus bubalis* *rrnS* gene were compared with similar gene from the *Bos Taurus* (GI42521312), *Lama pacos* (GI 9558343), *Musmusculus* (GI 33115104), *Rattusnorvegicus* (GI 26983975) and *Bosindicus* (GI 33321647). AT Skewness was in the range of 0.178 to 0.245 and highest, and similar Codon usage pattern observed for two sequences from *Bubalus bubalis* (water buffalo) of Indian origin and *Bosindicus*. (Table 3)

### 3.5 EST 2Genome

Out of total 312 EST from lactating 29 mapped on Chromosome M during 209 from non-lactating 27. Comparison of mapped gene between lactating and non-lactation revealed many common gene mapping. For non-lactating condition unique mapping observed for *nad5*(NAD5) only While, in case of lactating unique condition mapping observed for *atp8*(ATP8), *atp6*(ATP6), *cox3*(subunit 3 of cytochrome) and *cob*(apocytochrome B).The mapping of these EST suggested the increases in

mammary mitochondria ATPsynthesis activity that is in agreement with the report by Hadsell *et al.*, 2011. [15]

## IV. CONCLUSION

In the present study, complete mitochondrial genome data of two different *Bubalus bubalis* breed was from Indian origin were annotated and compared with other available mitochondrial genome and another related animal. Genome annotation, codon usage pattern, and whole alignment revealed the high similarity between the two sequences from Indian origin. The EST2Genome mapping showed activation of Atp8 (ATP8), atp6 (ATP6), cox3 (subunit 3 of cytochrome) and cob (apocytochrome B) during lactation. The Mammary glands mitochondrial ATP synthesis activity observed in during lactation condition.

## REFERENCES

- [1] Food and Agriculture Organization of the United Nations, “FAO Statistical Yearbook 2013,” 2013.
- [2] M. S. Tantia, R. K. Vijn, V. Bhasin, P. Sikka, P. K. Vij, R. S. Kataria, B. P. Mishra, S. P. Yadav, A. K. Pandey, R. K. Sethi, B. K. Joshi, S. C. Gupta, and K. M. L. Pathak, “Whole-genome sequence assembly of the water buffalo (*Bubalus bubalis*),” *Indian J. Anim. Sci.*, vol. 81, pp. 465–473, 2011.
- [3] S. M. Totey, G. Singh, M. Taneja, C. H. Pawshe, and G. P. Talwar, “In vitro maturation, fertilization and development of follicular oocytes from buffalo (*Bubalus bubalis*),” *J. Reprod. Fertil.*, vol. 95, no. 2, pp. 597–607, Jul. 1992.
- [4] V. N. Michelizzi, M. V. Dodson, Z. Pan, M. E. J. Amaral, J. J. Michal, D. J. McLean, J. E. Womack, and Z. Jiang, “Water buffalo genome science comes of age,” *International Journal of Biological Sciences*, vol. 6, pp. 333–349, 2010.
- [5] P. Parma, M. Erra-Pujada, M. Feligini, G. Greppi, and G. Enne, “Water buffalo (*Bubalus bubalis*): Complete nucleotide mitochondrial genome sequence,” *DNA Seq.*, vol. 15, no. 5–6, pp. 369–73, 2004.
- [6] V. Terzi, C. Morcia, M. Spini, R. Tudisco, M. I. Cutrignelli, F. Infascelli, A. M. Stanca, and P. Faccioli, “Identification and validation of reference genes for gene expression studies in water buffalo,” *animal*, vol. 4, p. 853, 2010.
- [7] A. Masoudi-Nejad, K. Tonomura, S. Kawashima, Y. Moriya, M. Suzuki, M. Itoh, M. Kanehisa, T. Endo, and S. Goto, “EGassembler: online bioinformatics service for large-scale processing, clustering and assembling ESTs and genomic DNA fragments,” *Nucleic Acids Res.*, vol. 34, no. Web Server issue, pp. W459–62, Jul. 2006.

- [8] S. K. Wyman, R. K. Jansen, and J. L. Boore, “Automatic annotation of organellar genomes with DOGMA.,” *Bioinformatics*, vol. 20, no. 17, pp. 3252–5, Nov. 2004.
- [9] U. Vetrivel, V. Arun Kumar, and S. Dorairaj, “ACUA: a software tool for automated codon usage analysis.,” *Bioinformation*, vol. 2, pp. 62–63, 2007.
- [10] A. E. Darling, B. Mau, and N. T. Perna, “Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement,” *PLoS One*, vol. 5, no. 6, 2010.
- [11] R. Mott, “EST \_ GENOME : a program to align spliced DNA sequences to unspliced genomic DNA,” *Cabios Appl. Note*, vol. 13, no. 4, pp. 477–478, 1997.
- [12] R. Z. Jing-fen Kang, Xiang-long Li and X. Guo, Lan-hui Li , Fu-jun Feng, “Bioinformatics Analysis of Lactoferrin Gene for Several Species,” *Biochem Genet*, vol. 46, no. 2008, pp. 312–322, 2008.
- [13] T. Faraut, S. De Givry, C. Hitte, Y. Lahbib-Mansais, M. Morisson, D. Milan, T. Schiex, B. Servin, A. Vignal, F. Galibert, and M. Yerle, “Contribution of radiation hybrids to genome mapping in domestic animals,” *Cytogenetic and Genome Research*, vol. 126. pp. 21–33, 2009.
- [14] M. E. J. Amaral, J. R. Grant, P. K. Riggs, N. B. Stafuzza, E. A. R. Filho, T. Goldammer, R. Weikard, R. M. Brunner, K. J. Kochan, A. J. Greco, J. Jeong, Z. Cai, G. Lin, A. Prasad, S. Kumar, G. P. Saradhi, B. Mathew, M. A. Kumar, M. N. Miziara, P. Mariani, A. R. Caetano, S. R. Galvão, M. S. Tantia, R. K. Vijn, B. Mishra, S. T. B. Kumar, V. A. Pelai, A. M. Santana, L. C. Fornitano, B. C. Jones, H. Tonhati, S. Moore, P. Stothard, and J. E. Womack, “A first generation whole genome RH map of the river buffalo with comparison to domestic cattle.,” *BMC Genomics*, vol. 9, p. 631, Jan. 2008.
- [15] D. L. Hadsell, W. Olea, J. Wei, M. L. Fiorotto, R. K. Matsunami, D. A. Engler, and R. J. Collier, “,” *Physiol. Genomics*, vol. 43, no. 6, pp. 271–85, Mar. 2011.

## Tables and Figures

**Table 1: Bubalus Bubalis Mitochondrial Genome**

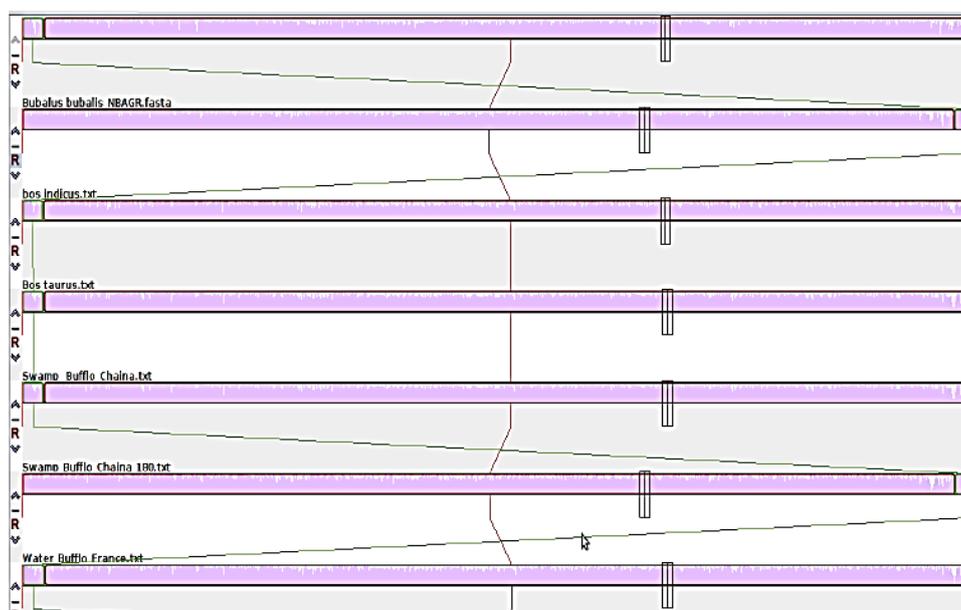
Availability/GI No	Source of Mitochondrial genome	Country
Genome browser for the Water Buffalo (Bubalus bubalis) Pure Murrah Buffalo	Bubalus bubalis (Water buffalo)	India
GI 33334119	Bubalus bubalis (Water buffalo)	India
GI 347801295	Bubalus bubalis (Water buffalo)	France
GI 52220982	Bubalus bubalis (Swamp buffalo)	China
GI 44894095	Bubalus bubalis (Water buffalo)	Italy
GI 51173180	Bubalus bubalis (Swamp buffalo)	China

**Table 2: Comparison of Different Mitochondrion Genome**

<i>Bubalus bubalis</i> NBAGR, INDIA				<i>Bubalus bubalis</i> 33334119 CCMB, INDIA				<i>Bubalus bubalis</i> 52220982 CHINA				<i>Bostaurus</i> 42521312			
Gene	Start	End	Str	Gene	start	end	Str	Gene	start	end	Str	Gene	start	end	Str
tmF-gaa	373	440	p	tmF-gaa	387	455	p	tmF-gaa	378	446	p	tmF-gaa	365	431	p
rmS	441	1394	p	rmS	456	1409	p	rmS	447	1400	p	rmS	432	1386	p
tmA-ggc	743	811	m	tmA-ggc	759	827	m	tmA-ggc	750	818	m	tmA-ggc	1386	1452	m
tmV-uac	1397	1463	p	tmV-uac	1413	1479	p	tmV-uac	1404	1470	p	tmV-uac	1387	1453	p
rmL	1464	3027	p	rmL	1480	3043	p	rmL	1471	3035	p	rmL	1454	3024	p
tmL-uaa	3036	3110	p	tmL-uaa	3049	3123	p	tmL-uaa	3040	3114	p	tmL-uaa	3025	3099	p
tmQ-uug	3056	3124	m	tmQ-uug	3069	3137	m	tmQ-uug				tmQ-uug	3045	3113	m
nad1	3113	4059	p	nad1	3126	4079	p	nad1	3117	4070	p	nad1	3102	4055	p
tmI-gau	4062	4130	p	tmI-gau	4082	4150	p	tmI-gau	4073	4141	p	tmI-gau	4058	4126	p
tmQ-uug	4128	4199	m	tmQ-uug	4148	4219	m	tmQ-uug	4139	4210	m	tmQ-uug	4124	4195	m
tmM-cau	4202	4270	p	tmM-cau	4222	4290	p	tmM-cau	4213	4281	p	tmM-cau	4198	4266	p
nad2	4271	5308	p	nad2	4291	5331	p	nad2	4282	5322	p	nad2	4265	5307	p
tmS-uga	5309	5375	m	tmS-uga	5332	5398	m	tmS-uga	5323	5389	m	tmS-uga	5308	5374	m
tmY-gua	5310	5376	p	tmW-uca	5333	5399	p	tmW-uca	5324	5390	p	tmW-uca	5309	5375	p
tmA-ugc	5378	5446	m	tmA-ugc	5401	5469	m	tmA-ugc	5392	5460	n	tmA-ugc	5377	5445	m
tmN-guu	5448	5521	m	tmN-guu	5471	5543	m	tmN-guu	5462	5534	n	tmN-guu	5447	5519	m
tmC-gca	5554	5620	m	tmC-gca	5576	5642	m	tmC-gca	5567	5633	n	tmC-gca	5552	5618	m
tmY-gua	5621	5688	m	tmY-gua	5643	5709	m	tmY-gua	5634	5700	n	tmY-gua	5619	5686	m
cox1	5690	7232	p	cox1	5711	7252	p	cox1	5702	7243	n	cox1	5688	7229	p
tmK-cuu	7233	7301	m	tmK-cuu	7253	7321	m	tmK-cuu	7244	7312	m	tmS-uga	7230	7298	m
tmD-guc	7309	7377	p	tmD-guc	7329	7397	p	tmD-guc	7320	7388	p	tmD-guc	7306	7373	p
cox2	7379	8054	p	cox2	7399	8079	p	cox2	7390	8070	p	cox2	7375	8055	p
tmL-aag	8061	8128	p	tmL-aag	8086	8153	p	tmL-aag	8077	8144	p	tmL-aag	8062	8128	p
atp8	8143	8322	p	atp8	8155	8352	p	atp8	8146	8343	p	atp8	8130	8327	p
atp6	8286	8961	p	atp6	8316	8993	p	atp6	8307	8984	p	atp6	8291	8968	p
cox3	8964	9742	p	cox3	8996	9793	p	cox3	8987	9766	p	cox3	8971	9771	p
tmS-gga	9744	9810	m	tmS-gga	9780	9846	m	tmS-gga	9771	9837	m	tmS-gga			
tmG-ucc	9744	9812	p	tmG-ucc	9780	9848	p	tmG-ucc	9771	9839	p	tmG-ucc	9755	9823	p
nad3	9813	10163	p	nad3	9849	10199	p	nad3	9840	10190	p	nad3	9824	10177	p
tmR-ucg	10160	10227	p	tmR-ucg	10196	10264	p	tmR-ucg	10187	10255	p	tmR-ucg	10171	10239	p
nad4l	10228	10521	p	nad4l	10265	10558	p	nad4l	10256	10549	p	nad4l	10240	10533	p
nad4	10518	11894	p	nad4	10555	11931	p	nad4	10546	11922	p	nad4	10530	11906	p
tmV-cac	11895	11965	m	tmV-cac	11932	12002	m	tmV-cac	11923	11993	m	tmV-cac			
tmH-gug	11896	11966	p	tmH-gug	11933	12033	p	tmH-gug	11924	11994	p	tmH-gug	11977	11908	p
tmL-uag	12028	12098	p	tmL-uag	12065	12134	p	tmL-uag	12056	12125	p	tmL-uag	12039	12109	p
nad5	12099	13881	p	nad5	12135	13952	p	nad5	12126	13943	p	nad5	12110	13927	p
nad6	13900	14424	m	nad6	13942	14466	m	nad6	13933	14457	m	nad6	13917	14441	m
tmE-uuc	14425	14496	m	tmE-uuc	14467	14535	m	tmE-uuc	14458	14526	m	tmE-uuc	14442	14510	m
Cob	14498	15629	p	cob	14540	15752	p	cob	15735	15667	p	cob	14515	15651	p
tmT-ugu	15639	15708	p	tmT-ugu	15683	15752	p	tmT-ugu	15143	15743	p	tmT-ugu	15658	15727	p
tmS-aga	15707	15776	m	tmS-aga	15752	15817	m	tmP-ugg	15743	15808	n	tmP-ugg	15727	15792	m
												tmI-gau	11907	11976	m
												tmS-gcu	11978	12037	p
												tmN-auu	10187	10258	m
												tmT-ugu	4437	4506	m

**Table 3 : Codon Usage Analysis**

Gene	A	T	G	C	Total bp	AT Percent	GC Percent	AT Skewness
<i>Bubalus bubalis</i> _NBAGR, INDIA	348	211	172	223	954	58.6	41.41	0.245
<i>Bubalus bubalis</i> 33334119 CCMB, INDIA	347	211	172	224	954	58.49	41.51	0.244
<i>Bos indicus</i> 33321647	349	214	174	218	955	58.95	41.05	0.24
<i>Bubalus bubalis</i> 347801295 France	345	212	174	223	954	58.39	41.61	0.239
<i>Bubalus bubalis</i> 44894095 Italy	345	212	174	223	954	58.39	41.61	0.239
<i>Bubalus bubalis</i> 51173180 china	345	212	174	223	954	58.39	41.61	0.239
<i>Bostaurus</i> 42521312	351	216	172	216	955	59.37	40.63	0.238
<i>Lama pacos</i> 9558343	338	212	187	231	968	56.82	43.18	0.229
<i>Rattusnorvegicus</i> 26983975	350	241	162	203	956	61.82	38.18	0.184
<i>Bubalus bubalis</i> 52220982 CHINA	469	324	234	327	1354	58.57	41.43	0.183
<i>Musmusculus</i> 33115104	361	252	154	188	955	64.19	35.81	0.178



**Figure 1 : Genome Comparison Using Mauve**