

DATABASE

Open Access

# *In silico* mining of putative microsatellite markers from whole genome sequence of water buffalo (*Bubalus bubalis*) and development of first BuffSatDB

Sarika<sup>1</sup>, Vasu Arora<sup>1</sup>, Mir Asif Iquebal<sup>2</sup>, Anil Rai<sup>1</sup> and Dinesh Kumar<sup>1,3\*</sup>

## Abstract

**Background:** Though India has sequenced water buffalo genome but its draft assembly is based on cattle genome *BTau 4.0*, thus *de novo* chromosome wise assembly is a major pending issue for global community. The existing radiation hybrid of buffalo and these reported STR can be used further in final gap plugging and “finishing” expected in *de novo* genome assembly. QTL and gene mapping needs mining of putative STR from buffalo genome at equal interval on each and every chromosome. Such markers have potential role in improvement of desirable characteristics, such as high milk yields, resistance to diseases, high growth rate. The STR mining from whole genome and development of user friendly database is yet to be done to reap the benefit of whole genome sequence.

**Description:** By *in silico* microsatellite mining of whole genome, we have developed first STR database of water buffalo, *BuffSatDb* (Buffalo MicroSatellite Database (<http://cabindb.iasri.res.in/buffsatdb/>)) which is a web based relational database of 910529 microsatellite markers, developed using PHP and MySQL database. Microsatellite markers have been generated using MlcrSATellite tool. It is simple and systematic web based search for customised retrieval of chromosome wise and genome-wide microsatellites. Search has been enabled based on chromosomes, motif type (mono-hexa), repeat motif and repeat kind (simple and composite). The search may be customised by limiting location of STR on chromosome as well as number of markers in that range. This is a novel approach and not been implemented in any of the existing marker database. This database has been further appended with Primer3 for primer designing of the selected markers enabling researcher to select markers of choice at desired interval over the chromosome. The unique add-on of degenerate bases further helps in resolving presence of degenerate bases in current buffalo assembly.

**Conclusion:** Being first buffalo STR database in the world, this would not only pave the way in resolving current assembly problem but shall be of immense use for global community in QTL/gene mapping critically required to increase knowledge in the endeavour to increase buffalo productivity, especially for third world country where rural economy is significantly dependent on buffalo productivity.

**Keywords:** *de novo*, Microsatellites, Primers, Radiation hybrid, Water buffalo

\* Correspondence: [dineshkumarbhu@gmail.com](mailto:dineshkumarbhu@gmail.com)

<sup>1</sup>Centre for Agricultural Bioinformatics, Indian Agricultural Statistics Research Institute, Library Avenue, New Delhi 110012, India

<sup>3</sup>Genes & Genetic Resources Molecular Analysis Lab, National Bureau of Animal Genetic Resources, Karnal, Haryana 132001, India

Full list of author information is available at the end of the article

## Background

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]. There is 188.3 million buffalo population in the world which contributes around 55 – 60% of total milk production [2]. 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 [3].

Molecular markers can play a significant role for livestock improvement through conventional breeding strategies. Scientific resources are limited in many of the countries where buffaloes are economically important livestock and as a consequence, genome research has not been supported at the level of some of the other species [4]. Limited number of researches has been conducted globally exploring the genetic diversity on molecular genetic basis in buffalo in comparison with other farm animal genetic resources. This depends, in part on the knowledge of their genetic structure based on molecular markers like microsatellites [5].

Microsatellites are sequences made up of a simple sequence motif, not more than six bases long, that is tandemly repeated and arranged head to tail without interruption by any other base or motif. Simple, tandemly repeated di- and tri- nucleotide sequences have been demonstrated to be polymorphic in length in a number of eukaryotic genome [6]. The frequency with which they occur (once every 50,000–60,000 bp), the high degree of polymorphism displayed, and their random distribution across the genome [7] make them potentially very useful as DNA markers in gene mapping studies. Furthermore, two or more microsatellites may be analyzed simultaneously [8,9], opening new opportunity for genetic analysis of large number of samples.

To cater the need of microsatellite especially for biodiversity analysis, cattle microsatellite markers have been used in heterologous mode in buffalo and up to 56% of them have been found polymorphic [10]. Cattle microsatellite markers have many disadvantages in such diversity analysis like low polymorphism and loss of amplification due to null alleles, size biasness, hitch hiking and potential exclusion of abundant STR in gene pool [11]. Even there is limited work of STR mining using partial enriched genomic library [12].

There is no thorough *in silico* STR marker mining from buffalo genome to represent more holistic and cumulative variability of genome to be used in gene pool or biodiversity analysis and gene/QTL mapping. Though

India has sequenced water buffalo genome but its draft assembly is based on cattle genome *BTau 4.0*, thus *de novo* chromosome wise assembly is a major pending issue for global community [3]. The existing radiation hybrid of buffalo by Amaral *et al.* [13] and these reported STR can be used further in final gap plugging and “finishing” expected in *de novo* genome assembly. Such work needs extensive STR mining from buffalo genome at equal interval on each and every chromosome. In order to cater this urgent need in resolving assembly, mapping issues and biodiversity analysis, we have developed first STR database of water buffalo, *BuffSatDb* (Buffalo MicroSatellite Database) which is a web based relational database of microsatellites.

## Construction and content

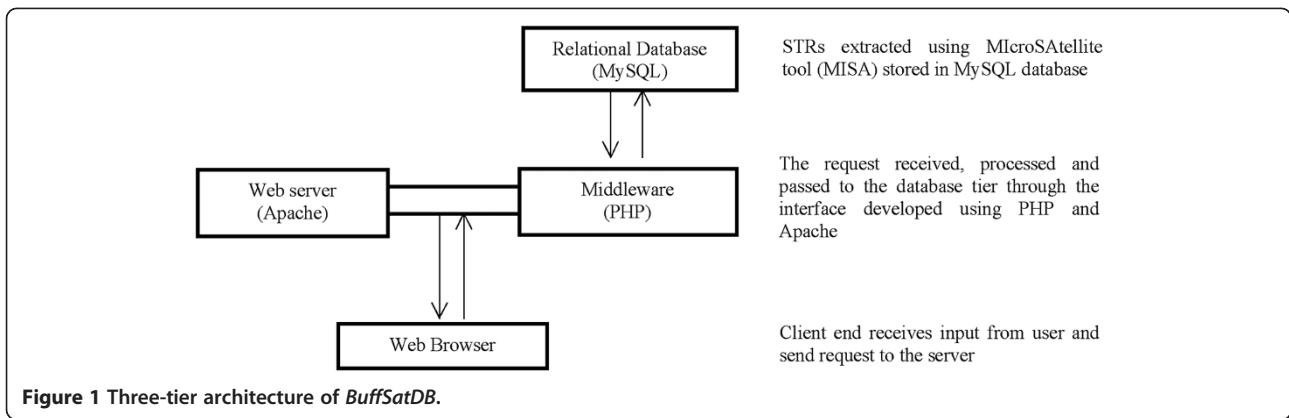
### Data collection and architecture

The *BuffSatDb* is an online relational database that catalogues information about the microsatellite repeats of the recently sequenced water buffalo. All the microsatellite markers extracted from buffalo genome have been generated using MicroSatellite tool (MISA) [14]. The database architecture is a “Three-tier architecture” (Figure 1) with a client tier, middle tier and database tier. This user-friendly interface for the database has been developed using PHP (Hypertext Preprocessor) which is an open-source server-side scripting language. In first tier of the architecture, the *in silico* mined STRs through MISA were stored in MySQL database. In the middleware, user need based customized query provisions have been made. For primer designing, Primer3 standalone code computes primers on user request. The information generated at the client end, i.e. third tier of the architecture are list of multiple primers along with their respective melting temperature, GC content, start position and product size (amplicon size).

*BuffSatDb* has eight tabs (Home, About, Database, Analysis, Tutorial, Links, Contact, Team). General information of the developed microsatellite database, information about Water buffalo, microsatellite markers, comparative analysis of the buffalo genome has been discussed. The tutorial of this database contains the guidelines for users and terminologies used in the database contents. *BuffSatDb* is appended with other useful links, the team and contact persons.

### In silico mining of microsatellite from whole genome of water buffalo

The *Bubalus bubalis* genome draft assembly version Bbu\_2.0-alpha which is with 17X-19X depth and 91%–95% coverage published by research group from India [3] and available in public domain at [http://210.212.93.84/bbu\\_2.0alpha/](http://210.212.93.84/bbu_2.0alpha/) was used for STR mining. All the 27 available chromosomes (Chromosome 1–24, M, U and X) were chopped into manageable range using PERL script.



These were fed to MicroSatellite identification tool, MISA to identify and find the location of perfect and compound microsatellites. The STR numbers, motifs, repeat number, length of the repeat, size of the repeat, repeat type, GC content, start and end position of the repeat and STR sequence were compiled. A total of 910529 STRs were generated from water buffalo genome, of which 830058

were simple and 80471 were compound STRs. BuffSatDb is the comprehensive and integrated resource for retrieval of information from water buffalo. Figure 2 shows the database search in BuffSatDb. The user can query for microsatellites, chromosome wise (1–24, X, M and U), where more than one chromosome may be selected at a time from water buffalo

**[1] Homepage of BuffSatDb**

**[2] Query page of BuffSatDb to select chromosome number, motif type, repeat motif, repeat kind, and customize advance search**

**[3] Display of mined STRs along with their details including start & end position & GC content**

S.No	chromosome	repeat type	motif type	motif	copy no	size range	start	end	GC Content	Run Primer
1	1	Simple	di	CA	9	18	216	793	56	Run Primer
2	1	Simple	di	TA	11	22	48708	48819	6	Run Primer
3	1	Simple	di	CA	6	12	31449	31456	56	Run Primer
4	1	Simple	di	AT	10	20	32896	32915	6	Run Primer
5	1	Simple	di	CA	14	28	93384	93411	56	Run Primer
6	1	Simple	di	AC	6	12	338269	338279	56	Run Primer
7	1	Simple	di	AT	16	32	133169	133149	6	Run Primer
8	1	Simple	di	CA	16	32	133393	133426	56	Run Primer

**[4] Size selection of locus amplicon along with option to select degenerate nucleotides**

**[5] Output display: Designed primer pair on desired STR locus**

Primer Name	Sequence	Melting Temperature (°C)	GC Content	Strand	Product Size
1 left	CATTGGCAAGGCGCAACATCA	59.502	49.0000	194	274
1 right	TATCTCGAGGGGCGCAAGAA	60.123	49.0000	199	269
2 left	CATTGGCAAGGCGCAACATCA	59.502	49.0000	194	269
2 right	TATCTCGAGGGGCGCAAGAA	59.829	49.0000	194	271

Figure 2 The flow of database search in BuffSatDb.

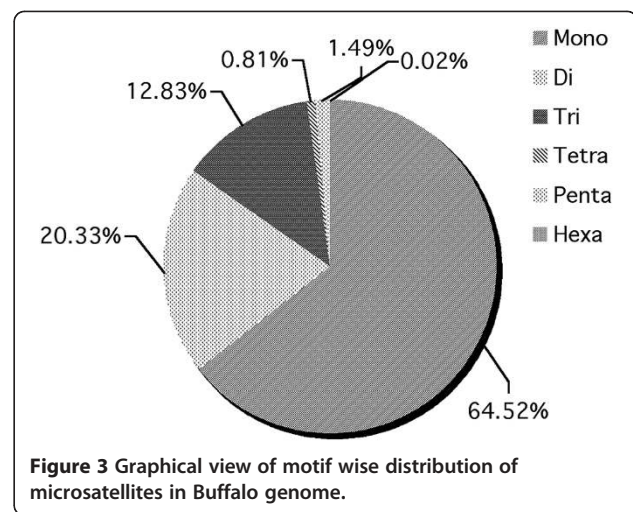


genome. These searches may further be customised based on microsatellite characteristics like motif type (mono, di, tri, tetra, penta, hexa), repeat motif and repeat kind (simple and composite). The user may further go for advance search like limiting the location on chromosome as well as the number of markers in that range. This is a novel approach and to the best of our knowledge, it has not been implemented in any of the existing marker database which may be useful for the researchers. Identification of QTL and fine mapping of economically important genes based on LOD (Logarithm of the Odds) score also needs STR preferably at equal interval. Also other parameters like GC content, range of STR location and copy number may be customised for the above selection according to the requirement of researchers. The results are then displayed in tabular format, giving chromosome number, motif type, motif, copy number, basepair, start and end position along with the GC content.

*BuffSatDb* is further appended with Primer3 tool [15]. The STRs traced by the query, may be selected with the help of radiobutton for generation of primers. Primer for selected STR locus may be designed with a template of approximately 1000 base pairs by selecting upto 500 base pairs of both flanking regions. These flexibilities would enable researchers to select markers of choice at desired interval over the chromosomes. Further one can use each individual STR of a targeted region over chromosome to narrow down location of gene of interest or linked QTL. A novel add-on for degenerate bases has been incorporated in this database search, where the users are given flexibility to replace degenerate bases with any of the alternative bases (A,T,G,C). This feature has been added to resolve the issue of some of the degenerate bases present in current buffalo genome assembly making the primer designing very difficult otherwise.

### Genome analysis

The chromosome wise distribution of STRs along with its respective motif frequencies in buffalo genome were analysed. It was observed that simple STRs constituted most abundantly with 91.16% of the total STRs. Various motif types like, mono, di, tri, tetra, penta and hexa type of microsatellites have been plotted to show the respective abundance of the type in chromosomes. Mono type (64.52%) was seen to have abundance than any other types while the hexa (0.02) was the one with least occurrence (Figure 3). It was found that the proportion of GC content in STRs in the range 0–10 was maximum (68.75%) followed by the range 41–50 (15.32%) while the minimum was in the range 81–90 (0.002%) (Figure 4). Figure 5 shows the distribution of length of microsatellites in context to GC percentage. No correlation was



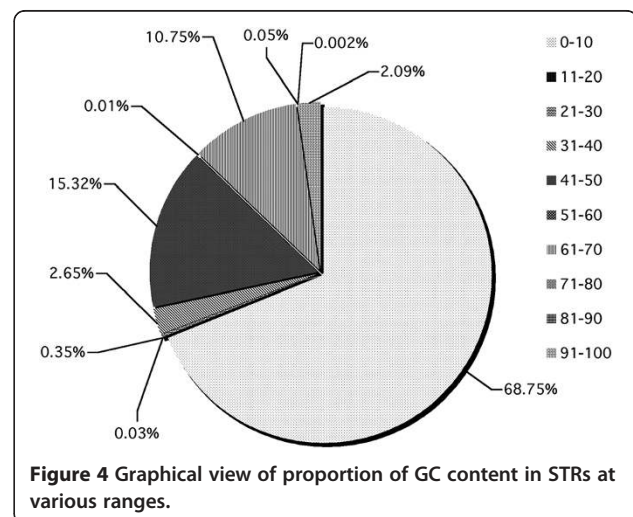
found between size and GC content. Table 1 depicts the frequency of STRs based on their sizes. Maximum numbers were reported for the size ranging between 11–13 followed by the size 14–16. A comprehensive chromosome wise STR profile with its repeat type is depicted in Table 2.

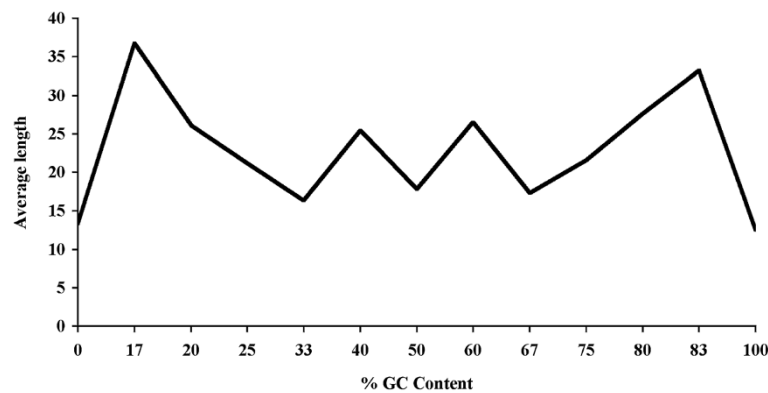
### STR validation

The previously published two sets of STR markers viz., heterologous [16] and homologous [17] were evaluated in the database using PERL script. The validated STRs are presented as positive primers in Table 3.

### Discussion and utility

A total of 910529 microsatellite markers have been searched by *in silico* mining. Simple STR were found to be most abundant (91.16%). Microsatellite density has been found positively correlated with genome size [18-20]. Among fully sequenced eukaryotic genomes, microsatellite





**Figure 5** Distribution of length of microsatellites in context to GC percentage.

density is highest in mammals. However in case of plant, microsatellite frequency is negatively correlated with genome size [21].

In the present study of water buffalo, mono- motif was found to be most abundant. Relative distributions of different microsatellite motif length classes in genomes differ considerably from species to species [22].

In case of water buffalo, it was found that longer repeats are less in abundance which is expected as reported and described in various studies [23,24]. It was also observed that microsatellite size range is increasing from 10 up to 14–16, however beyond this size range, it again starts decreasing. This is due to cyclical nature of microsatellite marker per say in its course of evolution. The birth of microsatellite starts with, out of register loop in event of DNA replication with a threshold size of 8 repeat unit or more, in the form of simple repeat. Gradually due to background mutation simple repeat gets converted in compound repeat. At the stage of simple repeat, the rate of mutation is high and predominantly it is addition of repeat unit and hence size increases. But once background mutation converts simple repeat into compound interrupted repeat, the smaller size simple repeat of less than 8 unit gets pinched off in subsequent replications. This maintains the size of microsatellite as evolutionary constraints otherwise microsatellite marker would have been always increasing in length during course of evolution. Thus individual microsatellites arrays have a “life cycle” of sorts, they are born, they grow and ultimately they perish. These events may stretch over tens or even hundreds of millions of years [25,26].

Water buffalo microsatellite profile exhibits the similar pattern. The relative abundance of repeat motif were in order of mono, di, tri, penta, tetra and hexa (Table 2). Though di-nucleotide repeats are most abundant in eukaryotic genome [27,28] but we found most abundance of mononucleotide repeats across all chromosome. This relatively higher abundance of mono over di nucleotide repeat type might be due to inherent limitation of the NGS

technology which adds more mono nucleotide causing sequencing error [29]. The longer the chromosome proportionately higher the total repeat content as expected in ubiquitously distributed STR markers [30].

In order to validate the previously reported STR markers, two sets viz. heterologous (cattle original species and buffalo focal species), homologous STR (developed from buffalo and validated in buffalo) were considered. The heterologous markers recommended by FAO-IASG [16] and homologous marker [17] were used. It was observed that both subsets of heterologous ISAG-FAO recommended primer for cattle and buffalo diversity analysis gave less validation results i.e. 10% and 13.33% respectively. Cross species amplifiability is due to conservation of cattle STR and its flanking regions in other species [31]. Though some of the primers showed validation up to 36.67% (Table 3). In the cross species amplifiability of bovidae species, such data are usually expected due to null alleles and genomic changes during speciation [32]. In validation of homologous STR, it was found that both subsets reported higher percentage of monomorphic (28.57%) and polymorphic (24.30%) loci. The validation results are limited as the first draft genome assembly of buffalo is based on cattle and it is not completely finished.

The findings of this study has limitations which need to be addressed. As genome of water buffalo is just draft assembly based on cow assembly *Btau 4.0*, thus *de novo* assembly is needed to have the buffalo specific chromosome wise microsatellite profile. The current database is

**Table 1** Frequencies of STRs based on their sizes

Size of STRs	Number of STRs	Contribution in percentage
<10	153185	16.82
11–13	276065	30.32
14–16	207831	22.83
17–25	162575	17.86
>25	110873	12.18

**Table 2 Chromosome wise distribution of STRs**

Chromosomes	Simple						Compound
	Mono	Di	Tri	Tetra	Penta	Hexa	
Chromosome 1	38690	12112	7672	454	1005	10	5858
Chromosome 2	35354	10911	7062	419	877	10	4899
Chromosome 3	32231	10406	6376	408	702	6	4611
Chromosome 4	30416	9552	6308	404	734	13	4382
Chromosome 5	22149	7486	4568	314	489	6	3365
Chromosome 6	20993	6941	4525	263	532	8	3132
Chromosome 7	21961	6884	4553	291	616	7	3282
Chromosome 8	22977	7190	4787	252	512	7	3333
Chromosome 9	20647	6278	4079	260	440	3	2937
Chromosome 10	20550	6425	4156	228	530	4	3082
Chromosome 11	19223	5662	3852	206	393	8	2701
Chromosome 12	19447	6285	3748	238	441	6	2725
Chromosome 13	17237	5220	3285	199	362	3	2721
Chromosome 14	15197	4703	2841	197	313	3	2162
Chromosome 15	14760	4892	2988	149	342	4	2219
Chromosome 16	15128	4630	3152	185	359	6	2353
Chromosome 17	14489	4585	2721	194	270	2	2160
Chromosome 18	11619	3814	2129	160	181	3	1746
Chromosome 19	13496	4373	2804	155	358	5	2098
Chromosome 20	12474	3949	2465	159	274	3	1832
Chromosome 21	11830	3467	2106	124	203	5	1588
Chromosome 22	12015	3893	2424	117	278	4	1705
Chromosome 23	18764	5898	3686	180	380	2	2704
Chromosome 24	7742	2571	1467	124	136	3	1127
Chromosome M	1	0	0	0	0	0	0
Chromosome U	50940	15373	9458	790	1279	26	9045
Chromosome X	15239	5241	3287	224	382	9	2704
Total	<b>535569</b>	<b>168741</b>	<b>106499</b>	<b>6694</b>	<b>12388</b>	<b>166</b>	<b>80471</b>

based on chromosome number of cattle which is certainly not the same in case of buffalo. For example cattle chromosome 4 is actually buffalo chromosome 8. In fact only chromosome number common between cattle and buffalo are just 5 viz 1, 2, 17, 18, and X [3]. The splitting

and translocation has rendered syntenic relationship between these two species which are well documented. Nevertheless the microsatellites in our database with option of primer designing at desired place over “chromosome” will be of immense use especially over radiation

**Table 3 STRs validation result of homologous and heterologous primer pairs of water buffalo**

	Heterologous		Homologous	
	ISAG-FAO recommended STRs from cattle	ISAG-FAO recommended STRs from buffalo	Nagarajan et al, monomorphic STRs from buffalo	Nagarajan et al, polymorphic STR loci from buffalo
Total no. of primer pairs reported	30	30	7	107
No. of positive primers (Forward)	7 (23.33%)	7 (23.33%)	2 (28.57%)	49 (45.79%)
No. of positive primers (Reverse)	11 (36.67%)	8 (26.67%)	4 (57.14%)	37 (34.58%)
No. of positive primers (common to both forward and reverse)	3 (10.00%)	4 (13.33%)	2(28.57%)	26 (24.30%)

hybrid of buffalo to resolve the problem and current issue of *de novo* assembly. Besides this, these markers can be further used for QTL, gene mapping as well as biodiversity analysis in setting the conservation priorities. The markers present in our database need further wet lab validation. Being first database of water buffalo microsatellite especially at juncture where *de novo* genome assembly is yet to be done, the use of these markers are highly warranted in order to “finishing” of water buffalo genome assembly. This will further lead to next version of buffalo microsatellite database base with proper buffalo specific chromosome wise data which is hitherto missing but critically needed. Such endeavour will fetch not only increase in buffalo productivity but also greater food security especially in third and new world countries.

## Conclusion

Being first buffalo STR database in the world, this would not only pave the way in resolving current water buffalo genome assembly problem but shall be of immense use for global community in QTL/gene mapping critically required to increase knowledge in the endeavour to increase buffalo productivity, especially for third world country where rural economy is significantly dependent on buffalo productivity.

## Availability and requirement

*BuffSatDb*, the buffalo microsatellite marker database is freely accessible for research purposes for non-profit and academic organizations at <http://cabindb.iasri.res.in/buff-satdb/>.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

DK and AR conceived this study. S, VA & MAI created the work-flow, database, web-tool and performed data analyses. MAI, S, DK and AR drafted the manuscript. All authors read and approved the manuscript.

## Acknowledgements

Financial assistance received by National Agricultural Innovation Project, Indian Council of Agricultural Research, New Delhi entitled “Establishment of National Agriculture Bioinformatics Grid in ICAR” is gratefully acknowledged. The technical assistance of Jai Bhagwan in maintaining the web server and AR Paul in designing the logo of BuffSatDb are thankfully acknowledged. Authors acknowledge the critical input of all the three anonymous reviewers and the editor in improvement of the manuscript.

## Author details

<sup>1</sup>Centre for Agricultural Bioinformatics, Indian Agricultural Statistics Research Institute, Library Avenue, New Delhi 110012, India. <sup>2</sup>Division of Biometrics & Statistical Modelling, Indian Agricultural Statistics Research Institute, Library Avenue, New Delhi 110012, India. <sup>3</sup>Genes & Genetic Resources Molecular Analysis Lab, National Bureau of Animal Genetic Resources, Karnal, Haryana 132001, India.

Received: 16 June 2012 Accepted: 9 January 2013

Published: 19 January 2013

## References

1. Scherf BD: *From World watch list for domestic animal diversity*, World Watch List. Rome, Italy: FAO; 2000.
2. FAO: *FAO Statistical Yearbook 2009*. Rome; 2009. (<http://www.fao.org/economic/the-statistics-division-ess/publications-studies/statistical-yearbook/fao-statistical-yearbook-2009/en/>).
3. Tantia MS, Vijn RK, Bhasin V, Sikka P, Vij PK, Kataria RS, Mishra BP, Yadav SP, Pandey AK, Sethi RK, Joshi BK, Gupta SC, Pathak KML: **Whole-genome sequence assembly of the water buffalo (*Bubalus bubalis*)**. *Indian J Anim Sci* 2011, **81**(5):38.
4. Womack JE: **Advances in livestock genomics: opening the barn door**. *Genome Res* 2005, **15**:1699–1705.
5. Gonçalves EC, Silva A, Barbosa MSR, Schneider MPC: **Isolation and characterization of microsatellite loci in Amazonian red-handed howlers *Alouatta belzebul* (Primates, Plathyrrini)**. *Mol Ecol* 2004, **4**:406–408.
6. Litt M, Luty JA: **A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene**. *Am J Hum Genet* 1989, **44**:397–401.
7. Luty JA, Guo Z, Willard HF, Ledbetter DH, Ledbetter S, Litt M: **Five polymorphic microsatellite VNTRs on the human X chromosome**. *Am J Hum Genet* 1990, **46**:776–783.
8. Weber JL, May PE: **Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction**. *Am J Hum Genet* 1989, **44**:388–396.
9. Georges M, Mishra A, Sargeant L, Steele M, Zhao X: **Progress towards a primary DNA marker map in cattle**. *4th World Congress Genetics Applied Livestock Production* 1990, **13**:107–112.
10. Navani N, Jain PK, Gupta S, Sisodia BS, Kumar S: **A set of cattle microsatellite DNA markers for genome analysis of riverine buffalo (*Bubalus bubalis*)**. *Anim Genet* 2002, **33**:149–154.
11. Ellegren H, Primmer CR, Sheldon BC: **Microsatellite evolution-directionality or bias**. *Nat Genet* 1995, **11**:360–362.
12. Venancio LPR, Amaral MEJ, Schneider MPC, Silva AL: **Microsatellite loci isolation from river buffalo using enriched partial genomic libraries**. *Ital J Anim Sci* 2007, **6**(2):338–341.
13. Amaral MEJ, Grant JR, Riggs PK, Stafuzza NB, Filho EAR, Goldammer T, Weikard R, Brunner RM, Kochan KJ, Greco AJ, Jeong J, Cai Z, Lin G, Prasad A, Kumar S, Saradhi GP, Mathew B, Kumar MA, Mizlara MN, Mariani P, Caetano AR, Galvão SR, Tantia MS, Vijn RK, Mishra B, Kumar STB, Pelai VAP, Santana AM, Fornitano LC, Jones BC, Tonhati H, et al: **A first generation whole genome RH map of the river buffalo with comparison to domestic cattle**. *BMC Genomics* 2008, **9**:631.
14. Thiel T, Michalek W, Varshney RK, Graner A: **Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.)**. *Theor Appl Genet* 2003, **106**:411–422.
15. Rozen S, Skaletsky HJ: **Primer3 on the WWW for general users and for biologist programmers**. In *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, Edited by Krawetz S, Misener S. Totowa, NJ: Humana Press; 2000:365–386.
16. FAO: *Molecular genetic characterization of animal genetic resources*. Rome: FAO Animal Production and Health Guidelines. No. 9; 2011.
17. Nagarajan MN, Nishanth G, Haribaskar R, Paranthaman K, Gupta J, Mishra M, Vaidhegi R, Kumar S, Ranjan AK, Kumar S: **Microsatellite markers of water buffalo, *Bubalus bubalis* -development, characterisation and linkage disequilibrium studies**. *BMC Genet* 2009, **10**:68–74.
18. Hancock JM: **Simple sequences and the expanding genome**. *BioEssays* 1996, **18**:421–425.
19. Toth G, Gáspári Z, Jurka J: **Microsatellites in different eukaryotic genomes: survey and analysis**. *Genome Res* 2000, **10**(7):967–981.
20. Katti MV, Ranjekar PK, Gupta VS: **Differential distribution of simple sequence repeats in eukaryotic genome sequences**. *Mol Biol Evol* 2001, **18**:1161–1167.
21. Morgante M, Hanafey M, Powell W: **Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes**. *Nature Genet* 2002, **30**:194–200.
22. Sharma PC, Grover A, Kahl G: **Mining microsatellites in eukaryotic genomes**. *Trends Biotechnol* 2007, **25**(11):490–498.
23. Temnykh S, Declerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch SR: **Computational and experimental analysis of microsatellite in rice (*O. sativa* L) frequency, length variation, transposon associations and genetic marker potential**. *Genome Res* 2001, **11**:1441–1452.

24. Grover A, Aishwarya V, Sharma PC: **Biased distribution of microsatellite in rice genome.** *Mol Genet Genomics* 2007, **277**:469–480.
25. Messier W, Li SH, Stewart CB: **The birth of microsatellites.** *Nature* 1996, **381**:483.
26. Primmer CR, Ellegren H: **Patterns of molecular evolution in avian microsatellites.** *Mol Biol Evol* 1998, **15**:997–1008.
27. Kariin S, Burge C: **Dinucleotide relative abundance extremes: a genomic signature.** *Trends Genet* 1995, **11**(7):283–290.
28. Shioiri C, Takahata N: **Skew of mononucleotide frequencies, relative abundance of dinucleotides and DNA strand asymmetry.** *J Mol Evol* 2001, **53**:364–376.
29. Haseneyer G, Schmutzer T, Seidel M: **From RNA-seq to large-scale genotyping-genomics resources for rye (*Secale cereale* L.).** *BMC Plant Biol* 2011, **11**:131–143.
30. Zietkiewicz E, Rafalski A, Labuda D: **Genome finger printing by Simple Sequence Repeat (SSR)-Anchored Polymerase Chain Reaction Amplification.** *Genomics* 1994, **20**:176–183.
31. Nguyen TT, Genini S, Bui LC, Voegeli P, Stranzinger G, Renard JP, Maillard JC, Nguyen BX: **Genomic conservation of cattle microsatellite loci in wild gaur (*Bos gaurus*) and current genetic status of this species in Vietnam.** *BMC Genet* 2007, **8**:77–84.
32. Kim KS, Min MS, An JH, Lee H: **Cross-species amplification of bovidae microsatellites and Low diversity of the endangered Korean goral.** *J Hered* 2004, **95**(6):521–525.

doi:10.1186/1471-2164-14-43

**Cite this article as:** Sarika et al.: *In silico* mining of putative microsatellite markers from whole genome sequence of water buffalo (*Bubalus bubalis*) and development of first BuffSatDB. *BMC Genomics* 2013 **14**:43.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

