

Database

Open Access

## EuMicroSatdb: A database for microsatellites in the sequenced genomes of eukaryotes

Veenu Aishwarya<sup>1</sup>, Atul Grover<sup>1,2</sup> and Prakash C Sharma\*<sup>1</sup>

Address: <sup>1</sup>University School of Biotechnology, Guru Gobind Singh Indraprastha University, Kashmere Gate, Delhi 110 006, India and <sup>2</sup>Shri Radha Kissen Kanoria Center for Advanced Studies in Bioscience and Biotechnology, Banasthali Vidyapith, Banasthali 304 022 (Raj.) India

Email: Veenu Aishwarya - veenu\_aishwarya@yahoo.com; Atul Grover - iatulgrover@gmail.com; Prakash C Sharma\* - deansbt@yahoo.co.in

\* Corresponding author

Published: 10 July 2007

Received: 22 November 2006

BMC Genomics 2007, 8:225 doi:10.1186/1471-2164-8-225

Accepted: 10 July 2007

This article is available from: <http://www.biomedcentral.com/1471-2164/8/225>

© 2007 Aishwarya et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Microsatellites have immense utility as molecular markers in different fields like genome characterization and mapping, phylogeny and evolutionary biology. Existing microsatellite databases are of limited utility for experimental and computational biologists with regard to their content and information output. EuMicroSatdb (**Eukaryotic MicroSatellite database**) <http://ipu.ac.in/usbt/EuMicroSatdb.htm> is a web based relational database for easy and efficient positional mining of microsatellites from sequenced eukaryotic genomes.

**Description:** A user friendly web interface has been developed for microsatellite data retrieval using Active Server Pages (ASP). The backend database codes for data extraction and assembly have been written using Perl based scripts and C++. Precise need based microsatellites data retrieval is possible using different input parameters like microsatellite type (simple perfect or compound perfect), repeat unit length (mono- to hexa-nucleotide), repeat number, microsatellite length and chromosomal location in the genome. Furthermore, information about clustering of different microsatellites in the genome can also be retrieved. Finally, to facilitate primer designing for PCR amplification of any desired microsatellite locus, 200 bp upstream and downstream sequences are provided.

**Conclusion:** The database allows easy systematic retrieval of comprehensive information about simple and compound microsatellites, microsatellite clusters and their locus coordinates in 31 sequenced eukaryotic genomes. The information content of the database is useful in different areas of research like gene tagging, genome mapping, population genetics, germplasm characterization and in understanding microsatellite dynamics in eukaryotic genomes.

### Background

Microsatellites, also called as simple sequence repeats (SSRs) or simple tandem repeats (STRs) are ubiquitous component of eukaryotic genomes. A microsatellite consists of a specific sequence of DNA which contains 1–6 bp long (mono- to hexa- nucleotide) tandem repeats viz. (A)<sub>16</sub>, (CA)<sub>20</sub>, (GATA)<sub>30</sub>. Over the years, molecular biolo-

gists have increasingly exploited these sequences for diverse applications.

With the whole genome sequencing initiatives of various eukaryotic organisms, large amount of genomic sequence data has accumulated over the last few years. These sequence resources available in the public domain have

also served as an attractive source of *in silico* mining of microsatellite sequences [1-5]. *In silico* mining of these sequences offers advantage in terms of time, labour and cost over conventional isolation from genomic libraries. Similarly, ESTs have also been screened for the presence of microsatellites [6-8]. However, finding potentially useful microsatellites occupying specific genomic regions still remains a challenge for the molecular biologists. Availability of this information can facilitate molecular mapping of desired traits and preparation of linkage maps saturated with evenly distributed SSR markers.

Popularity of *in silico* mining methods has led to the construction of various microsatellite databases in recent years, each with a different emphasis. For instance, MICdb [9] provides information on microsatellites spanning coding and non-coding regions, their frequency, size and repeat sequence. A recent version of this database covers 19 archeal, 155 eubacterial and 287 viral genomes. SilkSatDb [10] incorporates microsatellites extracted from the available ESTs and genomic sequences of the silkworm (*Bombyx mori*). This database also stores data on polymorphism status of different microsatellite loci. Similarly, mouse genomic microsatellites are collected in the Mouse Microsatellite Database of Japan (MMDBJ) [11]. CMD (Cotton Microsatellite Database) is a web-based relational database providing centralized access to publicly available cotton microsatellites. The database also provides a useful resource for mapping and related data pertaining to major cotton microsatellite projects [12]. Satellog [13] database catalogues triplet repeats associated with human disorders. Similarly Microsat2006 [14] database catalogues human microsatellite repeats. Taiwanese Polymorphic Microsatellite Database (TPMD) provides data on microsatellite mapping in the Taiwanese populations [15]. Molecular Mycology Research Laboratory, Westmead, Australia has created an SSR database [16] that stores information on microsatellite repeats in nine fungal genomes. InSatDb [17] provides size, type (perfect and

compound) and location (intron, exon, upstream or transposons) of microsatellites in five insect genomes. Some other databases [18,19], although published earlier are currently inaccessible. In conclusion, existing microsatellite databases either are very specific in their content and application or have limited utility to a wider audience. Thus, a collection of whole genome eukaryotic microsatellite data at a single platform is still not available. Recognizing this gap, we have developed a comprehensive database for easy retrieval of information on microsatellites distributed in the sequenced eukaryotic genomes. The database named as EuMicroSatdb (Eukaryotic Microsatellite database) presents a web-based user friendly interface for the extraction of both simple and compound microsatellites from 31 eukaryotic genomes assembled as chromosomes. Important features of this database are compared with those of existing databases in Table 1.

### Construction and Content

EuMicroSatdb is a platform independent relational database. The basic scheme followed for the development of EuMicroSatdb involved following steps: (1) whole genome sequences were downloaded from various sources like Ensembl [20], The National Center for Biotechnology Information (NCBI) [21], Genolevures 2 [22], International Rice Genome Sequencing Project (IRGSP) [23], Beijing Genomics Institute (BGI) [24], The Arabidopsis Information Resource (TAIR) [25] (Table 2) and scanned using a simple sequence repeat mining tool called MISA [26]; (2) filtering and restructuring of the required data using novel algorithms, VRFINE (creates "rfine" file that has sequence information of microsatellites) and VRSTRUCT (uses "rfine" file and processes it to make three separate files for repeat number, repeat motif and repeat unit length); (3) extraction of 200 bp upstream and downstream flanking sequences using a C++ program called VEXTRACT that creates two separate files, one each for upstream and downstream sequences; (4) microsatellite clustering information was generated using a Perl

**Table 1: Comparison of various microsatellite databases, available in the public domain**

Database	Details on					Coverage
	Simple Repeats	Compound Repeats	Clustering information	Genomic Positions	Flanking Sequences	
MICdb [9]	Y	Y	N	Y	Y	19 archeal, 155 bacterial and 287 viral genomes
SilkSatDb [10]	Y	Y	N	N	Y	Silkworm
MMDBJ [11]	Y	Y	N	N	N	Mouse
CMD [12]	Y	Y	N	N	Y	Cotton
Satellog [13]	Y	N	N	Y	N	Human
Database of Molecular Mycology Research Lab. [16]	Y	N	N	Y	N	9 fungal genomes
InSatDb [17]	Y	Y	N	Y	Y	5 insect genomes
MRD [18]	Y	N	N	N	N	8 eukaryotic genomes
SSRD [19]	Y	N	N	N	N	Human
EuMicroSatdb	Y	Y	Y	Y	Y	31 eukaryotic genomes

**Table 2: Details of genomes included in EuMicroSatdb**

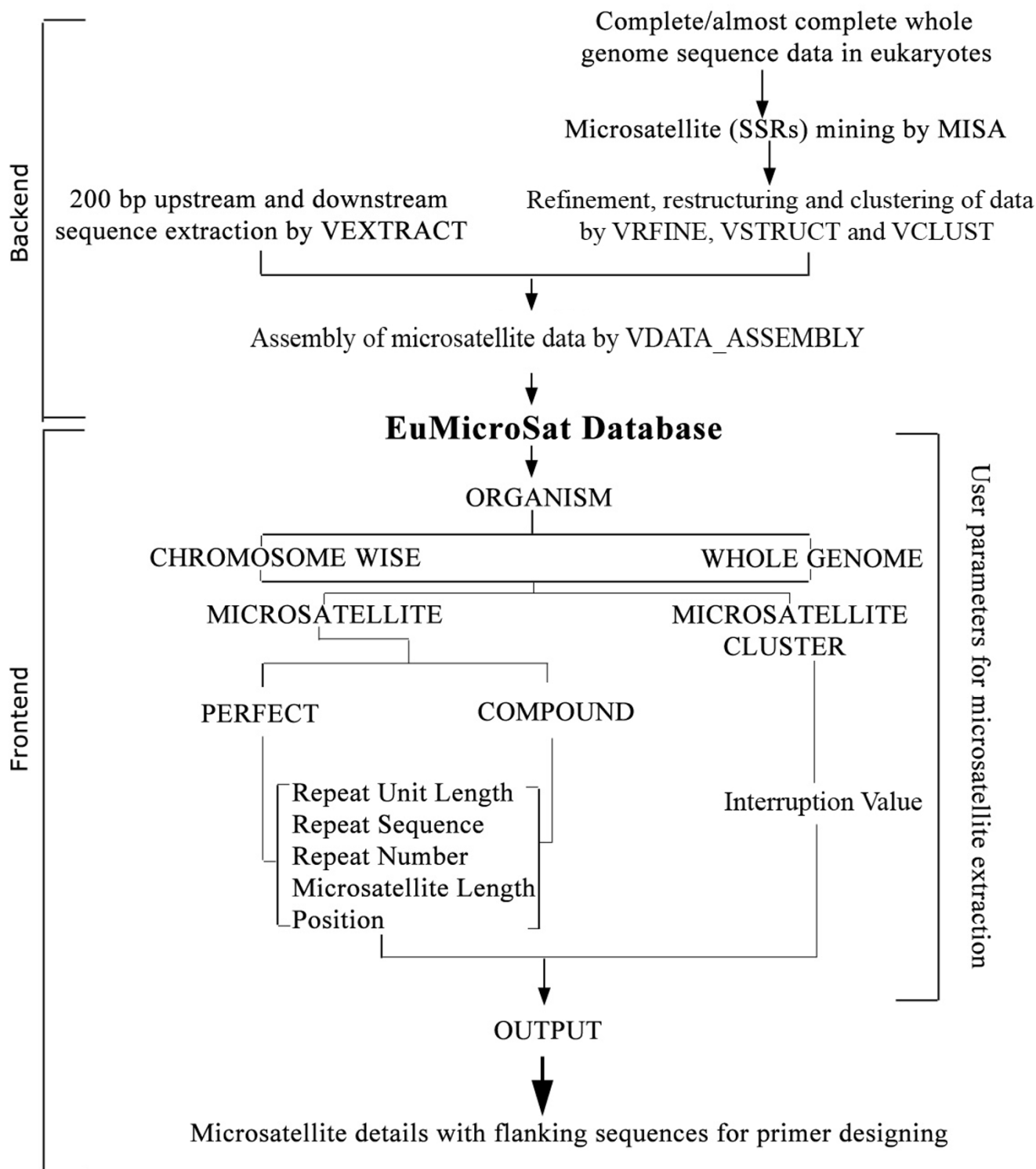
Species	Build/Assembly/Ver.	Source	Web Link
<i>Saccharomyces cerevisiae</i>	SGD 1, Nov 2005	Ensembl	<a href="http://www.ensembl.org">http://www.ensembl.org</a>
<i>Schizosaccharomyces pombe</i>	NCBI release	NCBI	<a href="ftp://ftp.ncbi.nlm.nih.gov">ftp://ftp.ncbi.nlm.nih.gov</a>
<i>Aspergillus oryzae</i> RIB40	NCBI release	NCBI	<a href="ftp://ftp.ncbi.nlm.nih.gov">ftp://ftp.ncbi.nlm.nih.gov</a>
<i>Aspergillus fumigatus</i>	NCBI release	NCBI	<a href="ftp://ftp.ncbi.nlm.nih.gov">ftp://ftp.ncbi.nlm.nih.gov</a>
<i>Cryptococcus neoformans</i> var JEC2 I	NCBI release	NCBI	<a href="ftp://ftp.ncbi.nlm.nih.gov">ftp://ftp.ncbi.nlm.nih.gov</a>
<i>Encephalitozoon cuniculi</i>	NCBI release	NCBI	<a href="ftp://ftp.ncbi.nlm.nih.gov">ftp://ftp.ncbi.nlm.nih.gov</a>
<i>Eremothecium gossypii</i>	NCBI release	NCBI	<a href="ftp://ftp.ncbi.nlm.nih.gov">ftp://ftp.ncbi.nlm.nih.gov</a>
<i>Candida glabrata</i> CBS138	Genolevures 2 Release 2, May 2006	Genolevures	<a href="http://cbi.labri.u-bordeaux.fr">http://cbi.labri.u-bordeaux.fr</a>
<i>Debaryomyces hansenii</i>	Genolevures 2 Release 2, May 2006	Genolevures	<a href="http://cbi.labri.u-bordeaux.fr">http://cbi.labri.u-bordeaux.fr</a>
<i>Kluyveromyces lactis</i>	Genolevures 2 Release 2, May 2006	Genolevures	<a href="http://cbi.labri.u-bordeaux.fr">http://cbi.labri.u-bordeaux.fr</a>
<i>Yarrowia lipolytica</i>	Genolevures 2 Release 2, May 2006	Genolevures	<a href="http://cbi.labri.u-bordeaux.fr">http://cbi.labri.u-bordeaux.fr</a>
<i>Caenorhabditis elegans</i>	WS160, July 2006	Ensembl	<a href="http://www.ensembl.org">http://www.ensembl.org</a>
<i>Plasmodium falciparum</i>	NCBI release	NCBI	<a href="ftp://ftp.ncbi.nlm.nih.gov">ftp://ftp.ncbi.nlm.nih.gov</a>
<i>Anopheles gambiae</i>	AgamP3, Feb 2006	Ensembl	<a href="http://www.ensembl.org">http://www.ensembl.org</a>
<i>Drosophila melanogaster</i>	BDGP 4.3, July 2005	Ensembl	<a href="http://www.ensembl.org">http://www.ensembl.org</a>
<i>Apis mellifera</i>	NCBI release	NCBI	<a href="ftp://ftp.ncbi.nlm.nih.gov">ftp://ftp.ncbi.nlm.nih.gov</a>
<i>Tribolium castaneum</i>	NCBI release	NCBI	<a href="ftp://ftp.ncbi.nlm.nih.gov">ftp://ftp.ncbi.nlm.nih.gov</a>
<i>Oryza sativa</i> ssp. japonica	Build 4.0	IRGSP	<a href="http://rgp.dna.affrc.go.jp/IRGSP/">http://rgp.dna.affrc.go.jp/IRGSP/</a>
<i>Oryza sativa</i> ssp. indica	2003-08-01 BGI	BGI	<a href="http://rise.genomics.org.cn">http://rise.genomics.org.cn</a>
<i>Arabidopsis thaliana</i>	ver. Jan 22 2004	TAIR	<a href="http://www.arabidopsis.org">http://www.arabidopsis.org</a>
<i>Ciona intestinalis</i>	JGI 2, Mar 2005	Ensembl	<a href="http://www.ensembl.org">http://www.ensembl.org</a>
<i>Tetraodon nigroviridis</i>	TETRAODON 7, Apr 2003	Ensembl	<a href="http://www.ensembl.org">http://www.ensembl.org</a>
<i>Danio rerio</i>	Zv6, Mar 2006	Ensembl	<a href="http://www.ensembl.org">http://www.ensembl.org</a>
<i>Rattus norvegicus</i>	RGSC 3.4, Dec 2004	Ensembl	<a href="http://www.ensembl.org">http://www.ensembl.org</a>
<i>Mus musculus</i>	NCBI m36, Dec 2005	Ensembl	<a href="http://www.ensembl.org">http://www.ensembl.org</a>
<i>Gallus gallus</i>	WASHU2, May 2006	Ensembl	<a href="http://www.ensembl.org">http://www.ensembl.org</a>
<i>Canis familiaris</i>	CanFam 1.0, July 2004	Ensembl	<a href="http://www.ensembl.org">http://www.ensembl.org</a>
<i>Macaca mulatta</i>	MMUL 1.0, Feb 2006	Ensembl	<a href="http://www.ensembl.org">http://www.ensembl.org</a>
<i>Bos taurus</i>	Btau_3.1, Aug 2006	Ensembl	<a href="http://www.ensembl.org">http://www.ensembl.org</a>
<i>Pan troglodytes</i>	PanTro 2.1, Mar 2006	Ensembl	<a href="http://www.ensembl.org">http://www.ensembl.org</a>
<i>Homo sapiens</i>	NCBI 36, Oct 2005	Ensembl	<a href="http://www.ensembl.org">http://www.ensembl.org</a>

based script **VCLUST**, and (5) all the data generated by above algorithms were reassembled into a data file using another Perl based script **VDATA\_ASSEMBL**. This file was then imported in MS-ACCESS as a table. The overall scheme of database construction is explained in figure 1. Sub-databases were constructed for individual genomes by importing all data files as tables, each representing one chromosome. Finally, an Index-database was created that communicates with these sub-databases. Front end web interface was developed using ASP that communicates with the Index database for data retrieval. The overall architecture of the database is outlined in figure 2.

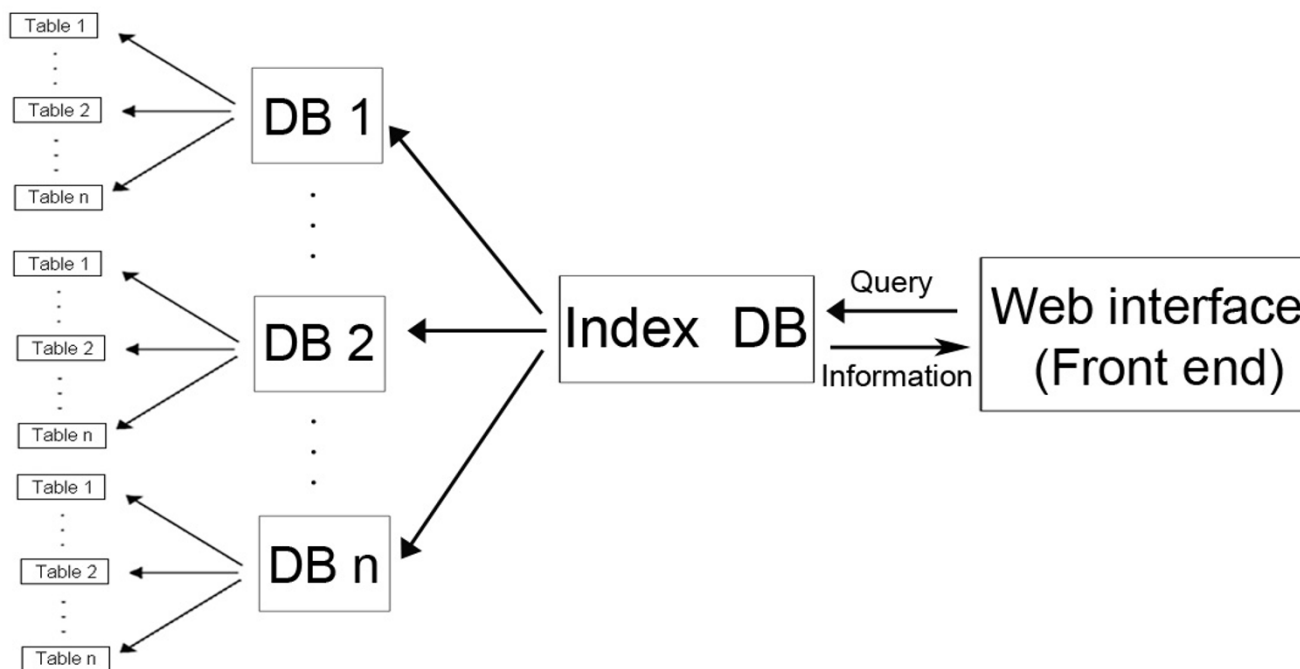
### Utility and Discussion

The EuMicroSatdb allows mining of different microsatellites along with their physical location on chromosomes in completely/almost completely sequenced eukaryotic genomes. At present, the database has over 10 million entries of microsatellites covering 31 genomes (Table 2). More genomes will be included in the database as and when their whole genome sequences are published and made available in the public domain.

User can search for perfect repeats, compound (perfect) repeats and microsatellite clusters. EuMicroSatdb database can be searched using following need based input parameters: **Repeat unit length**: the basic unit that is tandemly repeated in the microsatellite ranging from mononucleotide to hexanucleotide; **Repeat sequence**: this parameter allows the user to search microsatellite for a specific base sequence, for example, AT, GCG, etc.; **Repeat number**: is used to search microsatellites on the basis of repeat number of the microsatellite e.g. (CCT)<sub>9</sub> has a repeat number of 9, (AGAGG)<sub>10</sub> has a repeat number of 10; **Microsatellite length**: searches microsatellites on the basis of their total length in base pairs e.g. (TTGCA)<sub>5</sub> has a length of 25 bp; **Position**: defined locations on the chromosome in terms of base pairs can be specified; **Microsatellite cluster**: search can also be performed to look for adjacent microsatellites. Further, if the user wants to design primers for PCR amplification of the desired microsatellite locus, the database also provides 200 bp upstream and downstream regions of all the microsatellite loci. The search options are further explained with the help of some case studies given in a power point tutorial available on the database



**Figure 1**  
**Construction scheme of EuMicroSatdb.** Figure describing methodology used for preparation of backend (codes used in the preparation of files in the database) and front end (the web interface of the database).



**Figure 2**  
**Architecture of EuMicroSatdb.** Scheme of EuMicroSatdb describing the outline of the database.

website. Figure 3 displays the user interface for searching microsatellites by using one or more of the above-mentioned basic parameters.

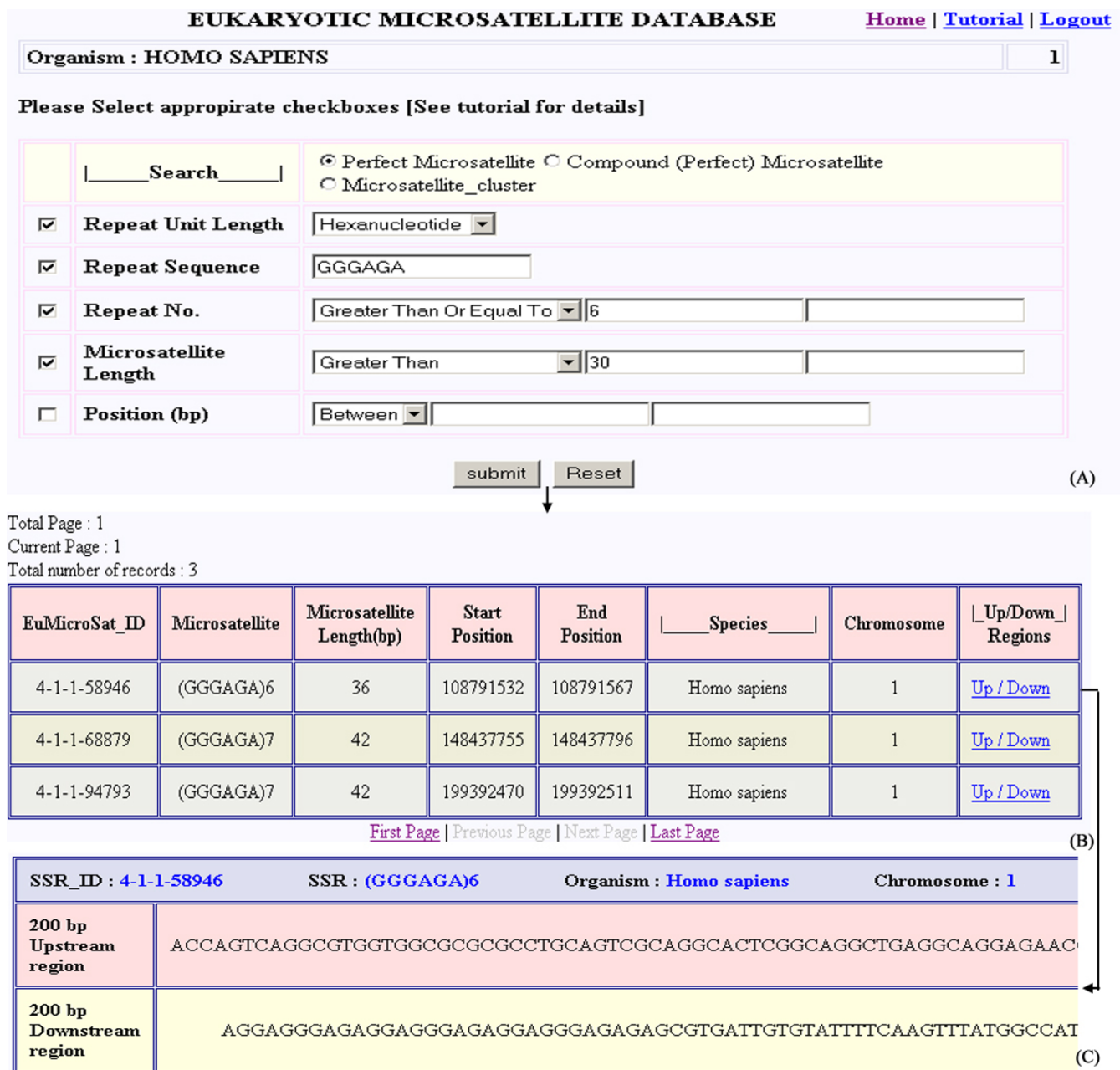
The unique feature of this database is the extraction of both simple and compound microsatellites. Compound repeats can be searched by specifying motifs desired in the combination. For example, if a user wants to search for a compound microsatellite from chromosome 1 of *Homo sapiens* which is more than 100 bp in length, has a TTTC-TC-TTTC repeat combination with the fourth association being a dinucleotide repeat, with repeat number greater than 10 for TTTC and TC, search can be made by using the parameters specified in figure 4A. The output of this query is shown in figure 4B.

User can identify the genomic regions showing high microsatellite density to study microsatellite clustering in the genome [27] by defining the size of interruption between neighbouring microsatellites as explained in the database tutorial.

EuMicroSatdb is likely to be adopted as a useful tool to study the relative occurrence and distribution of microsatellites across eukaryotic genomes. The information may have diverse applications. The user can extract the position of the microsatellite on the chromosome and thus

can link it with the gene co-ordinates, which are available on various public domains. In this way, microsatellites located in the vicinity of genes may be identified. These microsatellites will hopefully prove to be more useful for gene tagging and to investigate the role of microsatellites in gene regulation. Similarly, microsatellites spanning desired genomic regions can be selected and used for further saturation of existing molecular maps. Since microsatellite show varying levels of cross amplification among related genomes, microsatellites from the genomes included in the present database can be exploited for developing markers in the related species where sufficient STMS markers are still not available. The compound microsatellites being hypervariable can prove to be a potential source of highly polymorphic markers.

Incorporation of multiple sub-databases in EuMicroSatdb ensures faster exchange of information and unlimited expansion of the database. The database will be upgraded regularly as and when draft assemblies are updated and new genomes are sequenced. EuMicroSatdb is compatible with multi-user environment. The efficiency of data retrieval is maintained during simultaneous access by many users.



**Figure 3**  
**Web interface for simple microsatellite searching.** Web interface showing (A) various input parameters used for simple microsatellite based search, (B) output of the query, and (C) 200 bp upstream/downstream sequences for a particular microsatellite.

**Conclusion**

EuMicrosatdb has been developed for genome wide mining of microsatellite in 31 completely sequenced eukaryotic genomes considering the immense utility of these sequences for a variety of experiments. Various parameters have been carefully inducted to allow comprehensive search of simple and compound microsatellites and to

identify microsatellite clusters across the genomes. Links to retrieve flanking sequences (200 bp upstream and downstream) are provided to design primers for PCR amplification of desired motifs. EuMicroSatdb will provide a useful resource for mining microsatellites to be used in gene tagging, comparative genomics and genetic diversity based studies in different genomes

Organism : HOMO SAPIENS 1

Please select appropriate checkboxes [See tutorial for details]

Perfect microsatellite
  Compound (Perfect) microsatellite
  Microsatellite\_Cluster

<input checked="" type="checkbox"/> Repeat_Unit_Length	<input checked="" type="checkbox"/> Repeat_Sequence	<input checked="" type="checkbox"/> Repeat_No.	
<input checked="" type="checkbox"/> Tetranucleotide	<input checked="" type="checkbox"/> TTTC	<input checked="" type="checkbox"/> Greater Than	10
<input checked="" type="checkbox"/> Dinucleotide	<input checked="" type="checkbox"/> TC	<input checked="" type="checkbox"/> Greater Than	10
<input checked="" type="checkbox"/> Tetranucleotide	<input checked="" type="checkbox"/> TTTC	<input type="checkbox"/> Equal To	
<input checked="" type="checkbox"/> Dinucleotide	<input type="checkbox"/>	<input type="checkbox"/> Equal To	
<input type="checkbox"/> Mononucleotide	<input type="checkbox"/>	<input type="checkbox"/> Equal To	
<input type="checkbox"/> Mononucleotide	<input type="checkbox"/>	<input type="checkbox"/> Equal To	
<input type="checkbox"/> Mononucleotide	<input type="checkbox"/>	<input type="checkbox"/> Equal To	

Microsatellite Length : Greater Than 100

Position (bp) : Between

(A)

EuMicroSat_ID	Compound_Microsatellite	Microsatellite Length(bp)	Start Position	End Position	Species	Chromosome	Up/Down Regions
4-1-1-78104	(TTTC)17(TC)13*(TTTC)5(TC)13*	136	165756376	165756511	Homo sapiens	1	<a href="#">Up / Down</a>

[First Page](#) | [Previous Page](#) | [Next Page](#) | [Last Page](#)  
[Search by new criteria](#)
(B)

**Figure 4**  
**Web interface for searching compound microsatellite.** Web interface showing (A) various input parameters used for searching compound microsatellite, and (B) output of the query.

**Availability and requirements**

EuMicroSatdb is a platform independent relational database publicly available at <http://ipu.ac.in/usbt/EuMicroSatdb.htm>

**Authors' contributions**

VA was mainly responsible for writing codes, designing the architecture of the database and execution of the study. AG participated in designing of the database and helped in the preparation of the manuscript. PCS conceived, coordinated and supervised the study. All authors read and approved the final manuscript.

**Acknowledgements**

The EuMicroSatdb team would like to thank Prof. K.K. Aggarwal, Vice-chancellor Guru Gobind Singh Indraprastha University, Delhi for encouragement and constructive suggestions throughout the project. We also acknowledge the technical support received from Mr. Ajeet Pratap, Univer-

sity School of Information Technology, Guru Gobind Singh Indraprastha University, Delhi.

**References**

- Fujimori S, Washio T, Higo K, Ohtomo Y, Murakami K, Matsubara K, Kawai J, Carninci P, Hayashizaki Y, Kikuchi S, Tomita M: **A novel feature of microsatellites in plants: a distribution gradient along the direction of transcription.** *FEBS Lett* 2003, **554**:17-22.
- Morgante M, Hanafey M, Powell W: **Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes.** *Nat Genet* 2002, **30**:194-200.
- Zhang L, Yuan D, Yu S, Li Z, Cao Y, Miao Z, Qian H, Tang K: **Preference of simple sequence repeats in coding and non coding regions of Arabidopsis thaliana.** *Bioinformatics* 2004, **20**:1081-1086.
- Dieringer D, Schlotterer C: **Two distinct modes of microsatellite mutation processes: evidence from the complete genomic sequences of nine species.** *Genome Res* 2003, **13**:2242-2251.
- Grover A, Sharma PC: **Microsatellite motifs with moderate GC content are clustered around genes on Arabidopsis thaliana chromosome 2.** *In silico Biol* 2007, **7**:0021.

6. Grover A, Sharma PC: **Occurrence of simple sequence repeats in potato ESTs is not random: An in silico study on distribution and length of simple sequence repeats.** *Potato J* 2004, **31**:95-102.
7. Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ: **Analysis of SSRs derived from grape ESTs.** *Theor Appl Genet* 2000, **100**:723-726.
8. Varshney RK, Thiel T, Stein N, Langridge P, Graner A: **In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species.** *Cell Mol Biol Lett* 2002, **7**:537-546.
9. Sreenu VB, Alevor V, Nagaraju J, Nagarajaram HA: **MICdb: database of prokaryotic microsatellites.** *Nucleic Acids Res* 2003, **31**:106-108.
10. Prasad MD, Muthulakshmi M, Arunkumar KP, Madhu M, Sreenu VB, Pavithra V, Bose B, Nagarajaram HA, Mita K, Shimada T, Nagaraju J: **SilkSatDb: a microsatellite database of the silkworm, Bombyx mori.** *Nucleic Acids Res* 2005, **33**:D403-D406.
11. **Mouse Microsatellite Data Base of Japan (MMDBJ)** [<http://www.shigen.nig.ac.jp/mouse/mmdbj/top.jsp>]
12. Blenda A, Scheffler J, Scheffler B, Palmer M, Lacape J-M, Yu JZ, Jesudurai C, Jung S, Muthukumar S, Yellambalase P, Ficklin S, Staton M, Eshelman R, Ulloa M, Saha S, Burr B, Liu S, Zhang T, Fang D, Pepper A, Kumpatla S, Jacobs J, Tomkins J, Cantrell R, Main D: **CMD: A cotton microsatellite database resource for Gossypium genomics.** *BMC Genomics* 2006, **7**:132.
13. Missirlis PI, Mead CR, Butland SL, Ouellette BF, Devon RS, Leavitt BR, Holt RA: **Satello: A database for the identification and prioritization of satellite repeats in disease association studies.** *BMC Bioinformatics* 2005, **10**:145.
14. **Microsat2006** [[http://www.microsatellites.org/db\\_search.php](http://www.microsatellites.org/db_search.php)]
15. Chang Y-H, Su W-H, Lee T-C, Sun H-FS, Chen C-H, Pan W-H, Tsai S-F, Jou Y-S: **TPMD: a database and resources of microsatellite marker genotyped in Taiwanese populations.** *Nucleic Acids Res* 2005, **33**:D174-D177.
16. Karaoglu H, Lee CMY, Meyer W: **Survey of simple sequence repeats in completed fungal genomes.** *Mol Biol Evol* 2005, **22**:639-649.
17. Archak S, Meduri E, Kumar PS, Nagaraju J: **InSatDb: a microsatellite database of fully sequenced insect genomes.** *Nucleic Acids Res* 2007, **35**:D36-D39.
18. Subramanian S, Madgula VM, George R, Mishra RK, Pandit MW, Kumar CS, Singh L: **MRD: a microsatellite repeats database for prokaryotic and eukaryotic genomes.** *Genome Biol* 2002 [<http://genomebiology.com/2002/3/12/preprint/00111>].
19. Subramaniam S, Madgula VM, George R, Kumar S, Pandit MW, Singh L: **SSRD: simple sequence repeats database of the human genome.** *Comp Funct Genomics* 2003, **4**:342-345.
20. **Ensembl Genome Browser** [<http://www.ensembl.org/>]
21. **The National Center for Biotechnology Information** [<ftp://ftp.ncbi.nlm.nih.gov>]
22. **Génolevures 2** [[http://cblabri.u-bordeaux.fr/Genolevures/download/GL2\\_index.php](http://cblabri.u-bordeaux.fr/Genolevures/download/GL2_index.php)]
23. **International Rice Genome Sequencing Project** [<http://www.rgp.dna.affrc.go.jp/IRGSP/>]
24. **Beijing Genomics Institute** [<http://www.rise.genomics.org.cn>]
25. **The Arabidopsis Information Resource** [<http://www.arabidopsis.org>]
26. **MISA - MicroSATellite identification tool** [<http://pgrc.ipk-gatersleben.de/misa/>]
27. Grover A, Aishwarya V, Sharma PC: **Biased distribution of microsatellite motifs in the rice genome.** *Mol Genet Genomics* 2007, **277**:469-480.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

