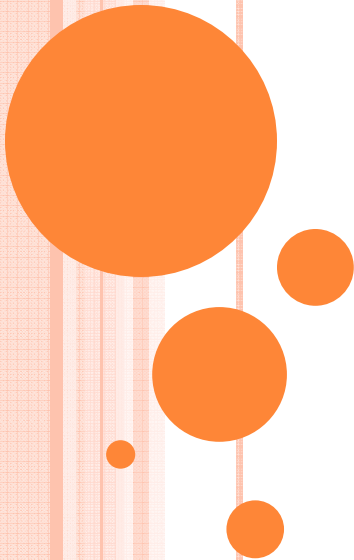


SSR - SIMPLE SEQUENCE REPEATS IDENTIFICATION FROM ESTs

Riju A

Bioinformatics Centre, IISR

riju.aikkal@gmail.com



- ❖ **Microsatellites, or Simple Sequence Repeats (SSRs)**, are polymorphic loci present in nuclear and organellar DNA that consist of repeating units of 1-6 base pairs in length.
- ❖ They are typically neutral, co-dominant and are used as molecular markers which have wide-ranging applications in the field of genetics, including kinship and population studies.



- SNP and SSR markers can be rapidly and cheaply identified thorough computational method from EST and other genomic sequence data.



SSRs are widely distributed throughout the genomes

Found in all prokaryotic and eukaryotic genomes

Three types of classes.

Perfect , imperfect and compound repeats

Perfect repeats are without interruptions.

Eg. ATATATATATATATATATATATAT

Imperfect repeats are interrupted by non-repeat nucleotides

Eg. ATATATATATATATATATATCTATAT



Compound repeats :

Two or more SSRs are found adjacent to one another.

Eg.

ATATATATATATATATATGCGCGCGCGCATGATGGGGGGGGGGGGGGGG

There may be also combination of this three.

Eg. Imperfect compound repeats.



Types of SSR

1. $(A)_n$, $(T)_n$, $(C)_n$, $(G)_n$ – mononucleotide SSRs
2. $(AT)_n$, $(CG)_n$, $(GT)_n$ – dinucleotide SSRs
3. $(ATT)_n$, $(CCG)_n$, $(GTA)_n$ – trinucleotide SSRs
4. $(CCGG)_n$, $(TATC)_n$ – tetranucleotide SSRs

Compound SSRs:

ATATATATCACACAATATATATCACACA - $(AT)_4(CA)_3$

CCGCCGATATATATCCGCCGATATATAT - $(CCG)_3(AT)_4$



Perfect Microsatellites:

- (ATT)21
- (TTC)32
- (AACAT)11

Compound Microsatellites:

- (AAGA)15(GAAA)14(GGAG)6
- (TCTT)18(TCCT)7(TCTT)7



Two types of markers

Dominant markers and Codominant markers

Allow the analysis of many loci per experiment within requiring previous information about their sequence.
Eg. RAPD, AFLPs and SMPLs

Codominant markers

allow the analysis of only a locus per experiment

More informative because the allelic variations of that locus can be distinguished
Eg. RFLPs, microsatellites



Characteristic features of SSR

It's a powerful genetic marker

Codominant, Abundance, dispersal throughout the genome, multi-allelic variation, high reproducibility and high level of polymorphisms.

High level of polymorphism is due to the mutation affecting the number of repeat units.

It requires small amount of DNA for screening.



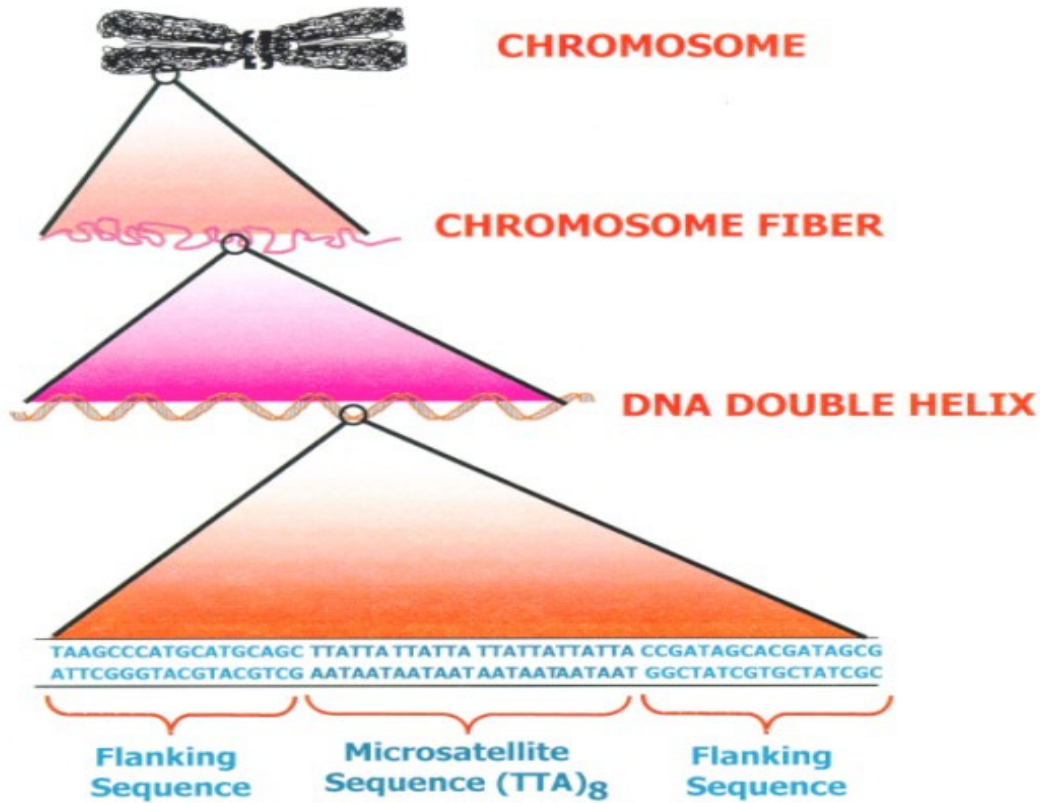
- **SSR demonstrate a high degree of transferability between species**
- **PCR primers designed to an SSR within one species frequently**
- **amplify a corresponding locus in related species**
- **Enables comparative genetic and genomic analysis**
- **Used to study gene dosage-looking for duplications or deletions of a particular genetic region**



- SSR were initially considered to be evolutionally neutral
- Plays an important role in genome evolution and hotspot of recombination
- Functional roles have been attributed to some SSRs.
- They are thought to be involved in gene expression, regulation, function, found to bind nuclear proteins and function as transcriptional activating elements
- Resides in non-coding regions but has functional significance.



MICROSATELLITES....



Majority are in non-coding region

Why do they exist?

- "junk" DNA
- A necessary source of genetic variation
- Regulate gene expression and protein function

Depends on the size of SSR .

Class I – hypervariable markers ..
>20bp or more

Class II - Ranges from >12 but <20



MICROSATELLITE MUTATIONS

- **10^{-3} to 10^{-6} events per locus per generation (point mutation 10^{-9} to 10^{-10})**
- **Varies by**
 - **repeat type**
 - **base composition of the repeat**
 - **taxonomic group**
- **most common - addition or deletion of a single repeat**



MUTATION MECHANISMS

1. Slippage in DNA at Replication (Slip-Strand Mismatching- SSM)
 - increases or decreases the repeat by one unit
2. Recombination
 - A. Unequal crossing over (UCO)
 - B. Gene conversion



POLYMORPHISM IN SSRs

1. ACGCATATATATATATATATATATATGGATCCA (TA)₁₁
2. ACTACGCATATATATATATATATATGGATCCA (TA)₁₀
3. ACTACGCATATATATATATATATATATATGGATCCA (TA)₁₂
4. ACTACGCATATATATATATATATGGATCCA (TA)₉



SIMPLE SEQUENCE REPEAT

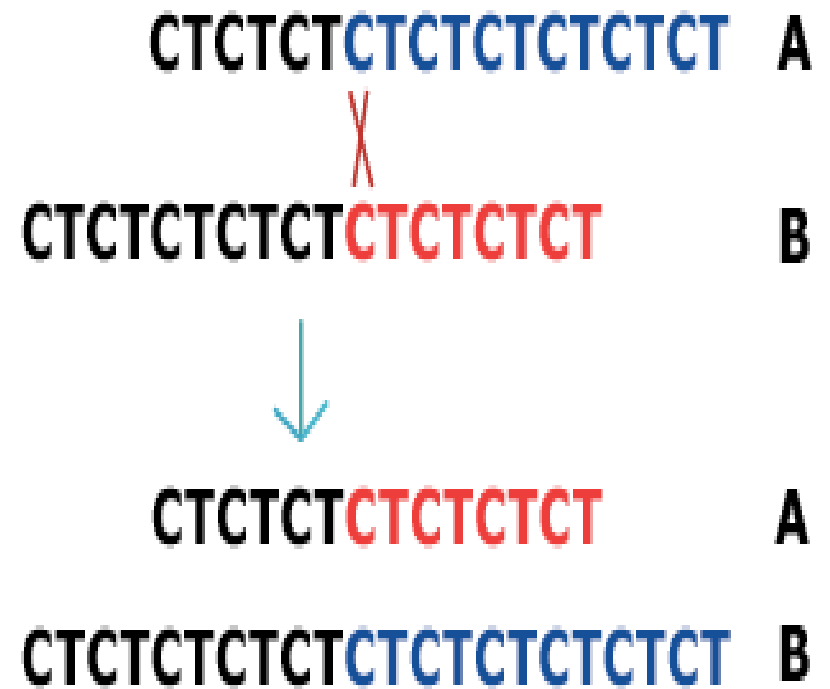
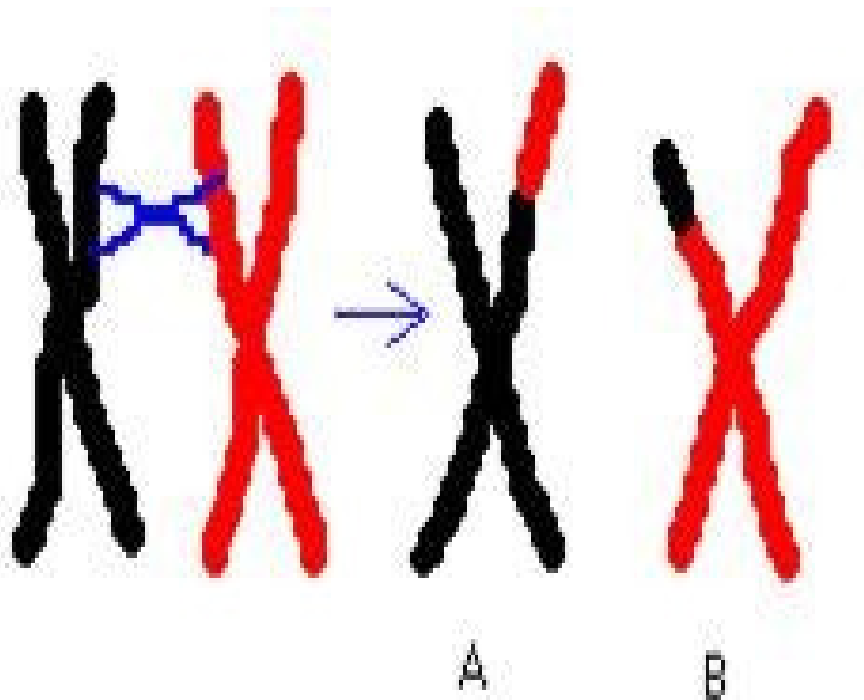
How do microsatellites mutate?

- **Replication Slippage**
- **Unequal crossing-over during meiosis**



SIMPLE SEQUENCE REPEAT

Unequal crossing-over during meiosis



WHY SSRs ARE PREFERRED GENETIC MARKERS

- Rapid processing is attainable
- Abundant throughout the genome
- Highly variable within various populations
- Small size range allows multiplex development
- Discrete alleles allow digital record of data
- Allelic ladders simplify interpretation
- PCR allows use of small amounts of DNA material
- Small product size compatible with degraded DNA



Microsatellite Discovery

1. Sequence databases screened for the presence of repeat motifs.
2. Homologous loci can often be amplified with primers from related species.



SSR tools and softwares



MICAS – MICROSATELLITE ANALYSIS SERVER

<http://sunserver.cdfd.org.in:8080/MIC/index.html>

- ❖ **MICAS is an Interactive user-friendly web-based server to find non-redundant microsatellites in a given nucleotide sequence/genome sequence.**

CDFD, Hyderabad





[MICAS - Home](#)

[What are
Microsatellites ?](#)

[Browse MICdb2.0](#)

[Scan a Sequence
for Microsatellite](#)

[Protocol](#)

[MICAS
Architecture](#)

[Suggestions /
Bug Reporting](#)

[MICAS Team](#)

[Acknowledgements](#)

MICAS - Microsatellite Analysis Server

(designed, developed & maintained at the EMBnet India Node)

Submit your Sequence to W-SSRF

W-SSRF program scans a given DNA sequence and identifies perfect microsatellite tracts. The main input to W-SSRF is the DNA sequence file.

Paste your sequence in the given window.

Sequence Input Here

Submit

Clear

Microsatellite Analysis Server MICAS

.	Motif	Repeats	Starting Position	EndingPosition
<input checked="" type="radio"/>	TG	4	45	52
<input type="radio"/>	GT	2	64	67
<input type="radio"/>	GT	2	75	78
<input type="radio"/>	AT	6	81	92
<input type="radio"/>	TA	4	93	100
<input type="radio"/>	AT	8	101	116
<input type="radio"/>	TCT	2	67	72
<input type="radio"/>	TAT	2	90	95
<input type="radio"/>	ATA	2	98	103
<input type="radio"/>	ATATA	2	96	105
<input type="radio"/>	TATATAT	2	86	99

AUTOPRIMER



Get Primer

Reset

PERFECT MICROSATELLITE REPEAT FINDER

[HTTP://SGDP.IOP.KCL.AC.UK/NIKAMMAR/REPEATFINDER.HTML](http://sgdp.iop.kcl.ac.uk/nikammarr/repeatfinder.html)

Perfect Microsatellite Repeat Finder

Enter Sequence to be examined:

```
>try this sample sequence  
CTCTCTGTCGTCGTCATAGCTACTGTCGCTGCTG
```

Search parameters

Minimum no. of repeats (default = 3):

Minimum repeat unit length (default = 2):

Maximum repeat unit length (default = 100):

Options

- Show summary report
- Show VNTR map
- Display potential primer design sequences

Search parameters:

Minimum number of repeats: 3

Minimum repeat unit length: 2

Maximum repeat unit length: 100

Summary report

Start	Stop	Unit	Repeat
1	8	CT	4
9	17	GTC	3
30	38	CTG	3

Microsatellite map (VNTR's in UPPERCASE)

CTCTCTCTGTCGTCGTCcatagctactgtCTGCTGCTG

Potential Primer Sequences

VNTR #	Start	Stop	Sequence
1	1	8	CTCTCTCTgtcgtcgtccatagctactgtctgctgctg
2	9	17	ctctctctGTCGTCGTCcatagctactgtctgctgctg
3	30	38	ctctctctgtcgtcgtccatagctactgtCTGCTGCTG

MICROSATELLITE REPEATS FINDER

[HTTP://WWW.BIOPHP.ORG/MINTOOLS/MICROSATELLITE_REPEATS_FINDER/DEMO.PHP](http://www.biophp.org/minitools/microsatellite_repeats_finder/demo.php)

Microsatellite repeats finder

Sequence:

```
AACAATGCCATGATGATGATTATTACGACACAACAACACCGCGCTTGACGGCGGCGGATGGATGCCG  
CGATCAGACGTTCAACGCCACGTAACGTAACGCAACGTAACCTAACGACACTGTTAACGGTACGAT
```

Length of repeated sequence: Minimum: 2 Maximum: 6

Minimum number of repeats: 3

Minimum length of tandem repeat: 6

Allowed percentage of mismatches: 0

Find Microsatellite repeats

Posición	Cicle	Repeats	Sequence
9	3	3	ATGATGATG
29	3	3	ACAACAACA
48	3	3	CGGCGGCGG





Welcome to Tandem Repeats Finder



[TRDB](#) is a public database of tandem repeats that allows users to run their own sequences. It has many added features not available in TRF.



[Submit a Sequence for Analysis](#)



[Download Your Own Copy of the Program](#)



[Find Out How TRF Algorithm Works](#)



[Find Out What's New](#)



[Add Your Email to Our Notification List](#)





CENTER FOR BIOINFORMATICS & COMPUTATIONAL BIOLOGY

Search [this site](#) [the web](#)

People

- Faculty
- Affiliated Faculty
- Postdocs & Students
- Scientific Staff
- Administrative Staff

Research Projects

- Genome Assembly >
- Gene Finding >
- Genome Sequencing >
- Insignia DNA Signatures
- more >

Software

- Glimmer
- MUMmer
- AMOS Assembler
- JIGSAW
- Transterm
- Bowtie
- more >

RepeatFinder: finding repetitive sequences complete and draft genomes

Overview

This directory contains two programs for finding repeats in genomic DNA sequences. The first program, described in the paper by Volfovsky et al. (2001) Genome Biology is RepeatFinder. [Get the reprint here](#). A second program, designed specifically to find repeats likely to confuse a genome assembly, is called ClosureRepeatFinder. The two programs are quite different and have different purposes; RepeatFinder is intended to be the more comprehensive approach. Both are open source; note that RepeatFinder depends on Stefan Kurtz's REPuter. Enjoy!

System requirements


Both programs are released as source code and were tested on Linux RedHat 6.x+, Sun Solaris, and Alpha OSF1, but should work on any Unix system. These have been run on newer Linux versions as well, but we are not supporting upgrades at present.

Obtaining RepeatFinder

This software is [OSI Certified Open Source Software](#) .




<http://wsmartins.net/webtroll/troll.html>



Web TROLL

find microsatellites and design primers



NEW [New \(beta\) version of WebTROLL](#)
[See video comparing the old and new \(beta\) version of WebTROLL](#)

Maximum Motif Length Minimum Repeat Length

Enter either a raw DNA sequence or a sequence (or multiple sequences) in FASTA format (at most 150,000 characters)

[Use a sample sequence](#)

- A. Castelo, W.Martins and G. Gao (2002). TROLL - Tandem Repeat Occurrence Locator, Bioinformatics Journal.



Programs

Search ▶

- ▶ alignment
- ▶ assembly
- ▶ database
- ▶ display
- ▶ hmm
- ▶ nudeic
- ▶ phylogeny
- ▶ protein
- ▶ sequence
- ▶ structure

- ▶ LIPM
- ▶ RPBS

Data Bookmarks

Jobs refresh

Tutorials refresh

- [How to use Mobyle? A step by step tutorial](#)
- [Registration information](#)

Welcome | **Programs** | Data Bookmarks | Jobs | Tutorials

etandem x

etandem

Finds tandem repeats in a nucleotide sequence

Reset

Help Pages

Run

Input section

* Sequence option (Nudeic Sequence)

Paste | DB | File clear data



WebSat
Find microsatellites and design primers

Motif Length: Mono Di Tri Tetra Penta Hexa

Repeat Minimum:

[Help: video instructions on how to use WebSat](#)

Enter either a raw DNA sequence or a sequence (or multiple sequences) in FASTA format (at most 150,000 characters)

[Use a sample sequence](#)

... or upload a file with the sequences





msatcommander

locate microsatellite repeats, design locus-specific primers, and tag 'em

[Project Home](#)

[Downloads](#)

[Wiki](#)

[Issues](#)

[Source](#)

[Summary](#) | [Updates](#) | [People](#)

News

8/21/09: 0.8.2 binaries are in the download area. See [Changelog](#) for fixes and an enhancement.

8/21/09: New changes uploaded to SVN, updating the code (mostly for windows) to deal with Python 2.5, BioPython 1.51, and wxPython. Additionally, added mispriming library to remove issues where microsatellite repeats were selected as proper primer binding sites.

Feedback

If you have constructive feedback (likes, dislikes, enhancements), please send an email to the [list-serve](#). All suggestions are welcomed and will be considered.

Purpose

msatcommander is a python program written to locate microsatellite (SSR, VNTR, &c) repeats within fasta-formatted sequence or consensus files. msatfinder will search for all di-, tri-, tetra-, penta-, and hexa-nucleotide repeats (with options to search for fewer repeat types and combinations of repeat types).

msatcommander will also design and tag primers using [primer3](#) as its primer design engine (see [References](#) for citations). Many thanks to the [primer3 team](#), in general, and [Steve Rozen](#), in particular. Additional information regarding primer3 may also be found on the [primer3 wiki](#).





University School of Biotechnology

Guru Gobind Singh Indraprastha University, Delhi

EUMICROSATDB (EUKARYOTIC MICROSATELLITE DATABASE)

How to use *EuMicroSatdb*

At present EuMicroSatdb has microsatellite data of the following species:

<i>Saccharomyces cerevisiae</i>	<i>Apis mellifera</i>
<i>Schizosaccharomyces pombe</i>	<i>Tribolium castaneum</i>
<i>Aspergillus oryzae</i> RIB40	<i>Oryza sativa</i> ssp. <i>japonica</i>
<i>Aspergillus fumigatus</i>	<i>Oryza sativa</i> ssp. <i>indica</i>
<i>Cryptococcus neoformans</i> var <i>JEC21</i>	<i>Arabidopsis thaliana</i>
<i>Encephalitozoon cuniculi</i>	<i>Ciona intestinalis</i>
<i>Eremothecium gossypii</i>	<i>Tetraodon nigroviridis</i>
<i>Candida glabrata</i> CBS138	<i>Danio rerio</i>
<i>Debaryomyces hansenii</i>	<i>Rattus norvegicus</i>
<i>Kluyveromyces lactis</i>	<i>Mus musculus</i>
<i>Yarrowia lipolytica</i>	<i>Gallus gallus</i>
<i>Caenorhabditis elegans</i>	<i>Canis familiaris</i>
<i>Plasmodium falciparum</i>	<i>Macaca mulatta</i>
<i>Anopheles gambiae</i>	<i>Bos taurus</i>
<i>Drosophila melanogaster</i>	<i>Pan troglodytes</i>
	<i>Homo sapiens</i>

SEARCHING PARAMETERS

- **Repeat Unit Length (mono-, di-, tri-, tetra-, penta-, hexa-)**
- **Repeat Sequence**
The sequence of the microsatellite e.g. GCC
- **Repeat No.**
(AT)₁₀ has a repeat number 10
- **Microsatellite Length**
(ATT)₂₀ has a motif length 60
- **Position**
Where the microsatellite is located on the chromosome
- **Interruption Size (Microsatellite Cluster)**
Difference between two adjacent microsatellites



COTTON MARKER DATABASE

<http://www.cottonmarker.org/>

- ❖ The Cotton Marker Database is a community collaboration to provide centralized access to all 9,001 publicly available SSRs and 312 mapped cotton RFLP sequences containing SSRs.





Cotton Marker Database

General info View Search Resources

General info

[About Us](#)
[Disclaimer](#)
[Tutorial](#)
[Community Collaborators](#)
[Submit Data](#)
[About SSR](#)
[About SNP](#)

View

[SSR Projects](#)
[eSNP Projects](#)
[SSRs](#)
[Markers](#)
[Homology](#)
[Primers](#)
[Panel](#)
[Publications](#)
[Maps](#)

Search

[SSRs](#)
[Markers](#)
[Homology](#)
[Publications](#)

Overview

The Cotton Marker Database is a community collaboration initiated and funded by **Cotton Incorporated** to provide centralized access to all 9,001 publicly available SSRs and 312 mapped cotton RFLP sequences containing SSRs. Currently, single nucleotide polymorphisms computationally mined in cotton ESTs (eSNPs) from the NCBI dbEST database have been included. The **standardized panel** screened data is available for many of the microsatellites. The standardized panel consists of 12 diverse genotypes selected from cultivated and exotic cottons.

Points of Interest

- The **SSR project** pages contain all available marker, primer, sequence, mapping, contact and publication information.
- The **eSNP project** pages contain summary information about CAP3 unigene assemblies and SNP data mining results.
- Download **standardized panel** screened marker data: 375 **BNL**, 127 **JESPR**, 204 **CIR**, 150 **STV**, and 200 **DPL**.
- View mapped microsatellites for F2, BC1, BC2 and RIL (10 cotton genetic maps) in the comparative map viewer **cMap**.
- Mine sequences for SSRs using the CMD **SSR server**.
- Search new sequences against existing SSRs using the batch upload **BLAST** or **FASTA** server tools.

INSATDB: INSECT MICROSATELLITE DATABASE

[HTTP://210.212.212.8/PHP/INSATDB/MSATFORM.PHP](http://210.212.212.8/PHP/INSATDB/MSATFORM.PHP)

- ❖ **InSatDb**, presents an interactive interface to query information regarding microsatellite characteristics *per se* of five fully sequenced insect genomes (fruit-fly, honeybee, malarial mosquito, red-flour beetle and silkworm).
- ❖ **InSatDb** allows users to obtain microsatellites annotated with size (in bp and repeat units); genomic location (exon, intron, upstream or transposon); nature (perfect or imperfect); and sequence composition (repeat motif and GC%).
- ❖ **InSatDb** is complete with the insects information, web links to find details, methodology and a tutorial. A separate *Analysis* section illustrates the comparative genomic analysis that can be carried out using the InSatDb output.





InSatDb: Insect Microsatellite Database

[Home](#)[About](#)[Analysis](#)[Database](#)[Tutorial](#)[Links](#)[Lab page](#)[Contact Us](#)

Insect :

Bombyx mori *Anopheles gambiae* *Drosophila melanogaster* *Apis mellifera* *Tribolium castaneum*

Location :

Intron Exon I-E boundary Upstream(1000 bp) Intergenic Repeat elements All

Microsatellite characteristics :

1. Repeat type **OR**

Repeat motif

2. GC% \geq \leq

UGMICROSATDB

DATABASE FOR MINING MICROSATELLITES FROM UNIGENES

[HTTP://WWW.VEENUASH.INFO/WEB1/INDEX.HTM](http://www.veenuash.info/web1/index.htm)

- ❖ **UgMicroSatdb** (Unigene MicroSatellite *database*), a web based relational database of microsatellites present in unigene sequences covering 80 genomes.
- ❖ **UgMicroSatdb** allows microsatellite search using multiple parameters like microsatellite type simple (perfect) and compound (perfect and imperfect), repeat unit length (mono- to hexa-nucleotide), repeat number, microsatellite length and repeat sequence class.
- ❖ Microsatellites can also be retrieved by specifying EST, cDNA, CDS identity or by using Gene Index, GenBank, UniGene IDs.
- ❖ The database also provides information about trinucleotide repeats encoding various amino acids. Such codon repeats can be searched by specifying characteristics of coded amino acids like charge (basic, acidic or neutral), polarity (polar or non-polar) and their hydrophobic or hydrophilic nature.
- ❖ The nucleotide sequences of the target UniGenes are also provided to facilitate primer designing for PCR amplification of any desired microsatellite. **UgMicroSatdb** is available at <http://ipu.ac.in/usbt/UgMicroSatdb.htm>.



_____ Search _____		
<input type="checkbox"/>	Repeat Unit Length	Mononucleotide <input type="button" value="v"/>
<input type="checkbox"/>	Repeat Sequence	<input type="text"/>
<input type="checkbox"/>	EST/mRNA/cDNA/CDS name (Functional annotation)	<input type="text"/>
	Structural annotation	<input type="checkbox"/> EST <input type="checkbox"/> mRNA <input type="checkbox"/> cDNA <input type="checkbox"/> CDS
<input type="checkbox"/>	GI/GB/UG IDs	<input type="text"/>
<input type="checkbox"/>	Repeat Sequence Class (Tri-repeats_only)	<input type="text"/>
<input type="checkbox"/>	Amino Acid [codon repeat]	Alanine <input type="button" value="v"/>
<input type="checkbox"/>	Amino Acid (Charge) [codon repeat]	Basic <input type="button" value="v"/>
<input type="checkbox"/>	Amino Acid (Polarity) [codon repeat]	Polar <input type="button" value="v"/>
<input type="checkbox"/>	Amino Acid (Hydro-) [codon repeat]	Hydrophobic <input type="button" value="v"/>
<input type="checkbox"/>	Other types [codon repeat]	Amino Acids with Aliphatic R-Groups <input type="button" value="v"/>
<input type="checkbox"/>	Microsatellite Length	Equal To <input type="button" value="v"/> <input type="text"/> <input type="text"/>

submit

Reset

<u>UgMicroSat_ID</u>	Microsatellite	Microsatellite Length(bp)	<u>UG_Len</u>	<u>UG_Seq</u>	<u>IDS</u>	
UMD-7-1-1-264	(CAG)5	15	883	Sequence	gb=BY955066 /gi=100403716 /ug=Ppa.1090	"gnl UG Ppa# Physcomitrella mRNA seque
UMD-7-1-1-490	(CAG)6	18	798	Sequence	gb=BY947547 /gi=100362825 /ug=Ppa.3571	"gnl UG Ppa# Physcomitrella mRNA seque
UMD-7-1-1-589	(CAG)7	21	861	Sequence	gb=BY948133 /gi=100369525 /ug=Ppa.4339	"gnl UG Ppa# Physcomitrella mRNA seque
UMD-7-1-1-601	(CAG)5	15	677	Sequence	gb=BY963326 /gi=100442831 /ug=Ppa.4425	"gnl UG Ppa# Physcomitrella mRNA seque
UMD-7-1-1-673	(CAG)5	15	846	Sequence	gb=BY949607 /gi=100377956 /ug=Ppa.5102	"gnl UG Ppa# Physcomitrella mRNA seque
UMD-7-1-1-754	(CAG)5	15	886	Sequence	gb=BY946150 /gi=100356009 /ug=Ppa.6222	"gnl UG Ppa# Physcomitrella mRNA seque
UMD-7-1-1-1105	(CAG)6	18	810	Sequence	gb=BY993357 /gi=100418147 /ug=Ppa.8239	"gnl UG Ppa# Physcomitrella mRNA seque



MICROSATELLITE REPEATS DATABASE

[HTTP://INSILICO.EHU.ES/MICROSATELLITES/INFO.HTML](http://insilico.ehu.es/microsatellites/info.html)

- ❖ Microsatellite Repeats finder has been used for detection of microsatellites in all up to date sequenced prokaryotic chromosomes.



MMDBJ: Mouse Microsatellite Data Base of Japan

<http://www.shigen.nig.ac.jp/mouse/mmdbj/top.jsp>

- ❖ Mouse Microsatellite Data Base of Japan (MMDBJ) provides information about SSLPs among different mouse strains, focusing on strains derived from Japanese wild mouse, *Mus musculus molossinus*, which are genetically remote from the standard laboratory strains.
- ❖ Two strains, MSM and JF1, are now widely used in gene mapping because of the high level of microsatellite polymorphisms for the standard laboratory strains and high reproduction ability in crosses with other strains.
- ❖ This data base includes PCR conditions for all entries of primer sets and keyword searches for the information.





Top

About

Search

Map

Download

Contents

- About
- Search
 - strain
 - polymorphism
 - chromosome
 - locus
 - MGI Position (cM)
 - Position(bp)
- Map
- Ensembl *e!*
 - Release 24.33.1
- Polymorphism (B6, 129/SVJ)
 - Chr. No.:
 - 1 2 3 4 5

GO TO Mouse Polymorphism DB

Mouse Microsatellite Data Base of Japan (MMDBJ) provides information about SSLPs among different strains, focusing on strains derived from Japanese wild mouse, *Mus musculus molossinus*, which remote from the standard laboratory strains. Two molossinus-derived strains, MSM and JF1, are in gene mapping because of the high level of microsatellite polymorphisms for the standard labor; high reproduction ability in crosses with other strains. This data base includes PCR conditions for primer sets and keyword searches for the information.

This project is being conducted with the support of the "Genome Frontier Development Promotion" assisted by the Special Funds for Promoting Science and Technology of the Science and Technology Agency, to which we would like to express our appreciation.

Database contents

Microsatellite: 9115

What's New !

October 25, 2007





SilkSatDb: The First Comprehensive Database on Insect Microsatellites

ABOUT

*Silkworm
Microsatellite
Database*

INFO

*Protocols
Frequency
Mutations
Distribution*

SilkSatDb is an interactive online relational database that catalogs information about the microsatellite repeats of the silkworm, ***Bombyx mori***. The user can query for microsatellites extracted from WGS and EST sequences of silkworm genome. It also stores information on primers developed and validated in our laboratory. The interface is coupled with Autoprimer, a primer-designing program, using which user can design primers for the loci of interest.



SilkSatDb: A Silkworm Microsatellite Database

<http://210.212.212.7:9999/PHP/SILKSAT/index.php>

- ❖ SilkSatDb is an interactive online relational database that catalogs information about the microsatellite repeats of the silkworm, *Bombyx mori*.
- ❖ The user can query for microsatellites extracted from WGS and EST sequences of silkworm genome.
- ❖ It also stores information on primers developed and validated in our laboratory.
- ❖ The interface is coupled with Autoprimer, a primer-designing program, using which user can design primers for the loci of interest.





Centre for Ecology & Hydrology

NATURAL ENVIRONMENT RESEARCH COUNCIL



Small Genomes Microsatellite Database

Molecular Evolution and Bioinformatics Section

CEH Bioinformatics Centre



[Home](#) | [view microsatellites](#) | [view genomes](#) | [search microsatellites](#) | [search genomes](#) | [data](#) | [find microsatellites](#)

Introduction

These data were generated as part of a survey of microsatellite sequences in small genomes from [GenBank](#). A companion database, [GeneSwytch](#), contains curated information on microsatellites thought to be potential contingency loci. Information found in this database was generated with the [Msatfinder](#) program.

Data on genomes and microsatellites are included for:



Small Genomes Microsatellite Database

Molecular Evolution and Bioinformatics Section

CEH Bioinformatics Centre



[Home](#) | [view microsatellites](#) | [view genomes](#) | [search microsatellites](#) | [search genomes](#) | [data](#) | [find microsatellites](#)

Database Log (microsatellites)

Categories of files in the Small Genomes Microsatellite Database

- Bacterial genomes
- Plasmid genomes
- Viral genomes
- Viroid genomes
- Phage genomes
- Organellar genomes



MISA

MIcroSAtellite identification tool (MISA)

Tool allows the identification and localization of perfect microsatellites.

Find Compound microsatellites also.



MISA

MISA - MicroSatellite identification tool

This tool allows the identification and localization of perfect microsatellites as well as compound microsatellites which are interrupted by a certain number of bases.

In order to design primers flanking the microsatellite loci, two perl scripts serve as interface modules for the program-to-program data interchange between MISA and the primer modelling software Primer3 (Whitehead Institute). For installation instructions of Primer3 see http://www-genome.wi.mit.edu/genome_software/other/primer3.html.

[Get MISA](#)

Supplemental perl tools:

[Get est_trimmer.pl](#) - perl script useful for pre-preprocessing sequences (invocation without parameters is explaining the syntax)

[Get Primer 3 interface modules](#) - perl scripts allowing the interaction with Primer3

Questions and comments: [Thomas Thiel](#)

Last updated: 5/14/02

<http://pgrc.ipk-gatersleben.de/misa>

SSRIT

SSRIT - Simple Sequence Repeat Identification Tool

This tool finds all *perfect* simple sequence repeats (SSRs) in a given sequence.

For source code for a stand-alone version please click [here](#).

For citation, please use this reference [Temnykh et al. \(2001\)](#).

Note: Netscape 2.0 or greater, or IE 4.0 or greater required.

1) Select search parameters

a) Select the maximum motif-length group you wish to find. For example, if you want to search for all SSRs up to and including pentamers (meaning, you'd like to search for dimers, trimers, tetramers, and pentamers), you should select 'pentamer' from the drop-down menu.

tetramer

b) Enter the minimum number of repeats you will allow. Entering a '5,' for example, will match SSRs with five or more motif repeats, such as ag-5 ('agagagagag').

2) Paste/Enter your sequence of interest into the textarea

The sequence(s) must be in FASTA format - meaning, there must be a title line with a '>' at the beginning for each sequence.

FOR EXAMPLE,

```
>seq1
agagattaggatcgatcgcgctctctctctctctcgatcgagatcgat
ggccatcatcatcatcattgagatatagcgcgatcgagagatctc
agaatagatatcgcgctatagagagatcgagagagagtaga
>seq2
agagataggaatatgagatagcggggggggggggcgctatacgcgctcg
gagagagatctctctctctttagagatcgactagctagatata
agactcactcactcactcactcagcgcat
```

Paste/Enter your sequence(s) here:

<http://www.gramene.org/db/searches/ssrtool>



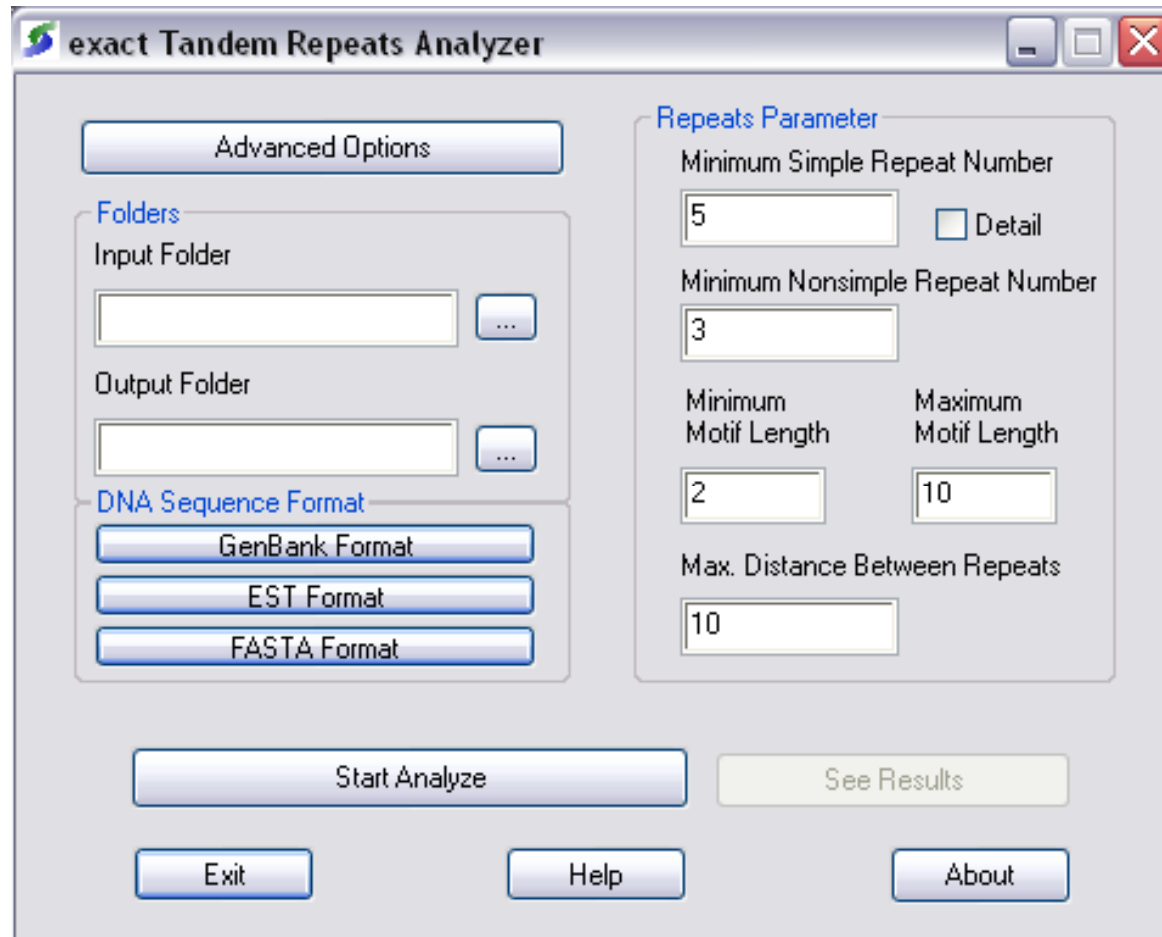
SSR IDENTIFICATION TOOLS

ETRA

Exact Tandem Repeats Analyzer 1.0 (E-TRA) combines sequence motif searches with keywords such as ‘organs’, ‘tissues’, ‘cell lines’ and ‘development stages’ for finding simple exact tandem repeats as well as non-simple repeats.



ETRA



Welcome to SSRPrimerII

SSRPrimerII accepts FASTA format sequences, searches for SSRs using SPUTNIK and designs PCR primers for the identified SSRs using Primer3. The process is described in:
Jewell E, Robinson A, Savage D, Erwin T, Love CG, Lim GAC, Li X, Batley J, Spangenberg GC and Edwards D. (2006) [SSR Primer and SSR Taxonomy Tree](#). *Biome SSR discovery. Nucleic Acids Research* **34**: W656.W659.

Sequence entry is currently limited to 4Kbp. Longer sequences are discarded without analysis.

Please contact Dave.Edwards@ug.edu.au for further details or to report problems.

1

Sequence data:

Either: Select the sequence file to upload:

Otherwise: Enter a sequence in FASTA format:

Applications of EST SSR

EST -SSR represent a potentially valuable source of gene-based markers for population genetic analyses.

Rapid and inexpensive development and high levels of cross-taxon portability.

Potential to facilitate evolutionary analyses in a wide variety of taxa, and may well represent the best way forward for the analysis of species for which only limited resources are available.

EST-SSR is the best resource for developing chromosome mapping of several crops. Linkage map can create with the help of MAPMAKER software

EST-SSR is used in the development of markers in phylogenetic analysis.



Case study using *Radopholus similis* ESTs



All Databases Pubmed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search EST for Radopholus.similis Go Clear Save Search

Limits Preview/Index History Clipboard Details

Found 7515 nucleotide sequences. Nucleotide [135] EST [7380]

Display Summary Show 20 Sort By Send to

All: 7380 Bacteria: 0 mRNA: 7380

Items 1 - 20 of 7380 Page 1 of 369 Next

1: [EY195367](#) Reports Links
 RSAA-aab21e07.g1 R.similis_EST_RSAA Radopholus similis cDNA 5- similar to ref[NP_491282.1] Calponin homology (CH) domain containing protein family member [Caenorhabditis elegans] pir[T29467 hypothetical protein F28H1.2 - Caenorhabditis elegans gb|AAB52338.1| Calponin protein 3 [Caenorhabditis elegans], mRNA sequence gi|159498513|gb|EY195367.1|[159498513]

2: [EY195366](#) Reports Links

Recent Activity

- [Radopholus.similis \(7380\)](#)
- [r.similis \(6206\)](#)

7380 ESTs in dbEST



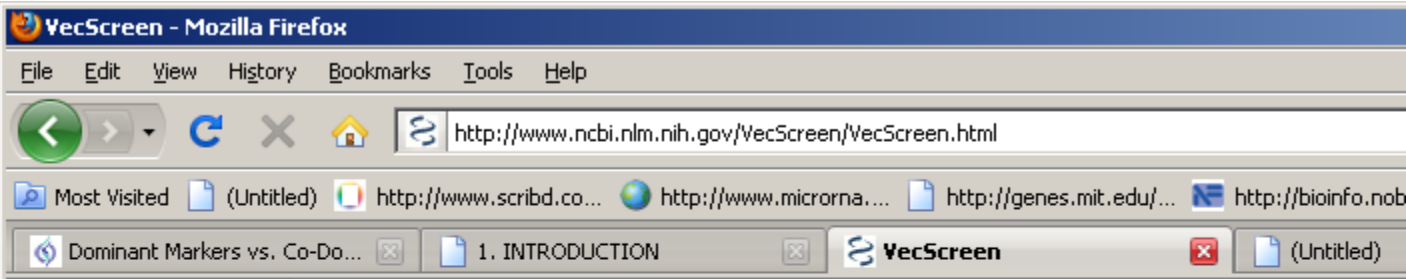
```
>gi|159498513|gb|EY195367.1|EY195367 RSAA-aab21e07.g1 R.similis_EST_RSAA Radopholus similis cDNA 5' similar to ref|NP_4  
GGCCGGGTGCGGAAGAGCCAGGGAGTGGCCACCGAGGAGACCTTCCAGAGCGTGGACTTGTTCGAGGCAC  
GTGACCTCTACTCCGTGTGCATGACCCTGTTGTCGTTGGGCCGAATTATGGAGAAGAAGGGAAAGCCGAA  
CCCATTCTCTGGATGAAGAGTAGAAGTGAGTGCAGCAAAAAGACGGACAGAGCATGTGCTATCGCTCCCAT  
TCGAGAACTCCCCGTTTTGCGAAATTTCTCCCCGTGTGCGACCTTGCAACAATACCGAGGCGTTAACTGT  
TTTCCCCTCTCTCCTCTCAAACGTGGGCATTTGAAAATAGCGATGCCGGAATAAATGGCCAATTCCA
```

```
>gi|159498512|gb|EY195366.1|EY195366 RSAA-aab21e06.g1 R.similis_EST_RSAA Radopholus similis cDNA 5', mRNA sequence  
GGCCGGGATTGGCACGTAATTCAACTGGTCTTTTTGCTGGGCGGCAATTGGCAACTTTATCCGACAGAAG  
CATGATAATTGGTGGAGCCGCCCTTTGCCGACGCACCTTTTGCAGCCCCACCCCCCATCGGGACACGGTA  
TTCGAGGTAGAATCGGACGAGGACTTCGAGAATGGGGTATACCATGCCGAAAAACCAGTGCTCATGCAGT  
TTTACGCCGATTGGTGTGGACCCTGTGCAAAATTTGGCACCCGAGGTTGATTGCCAAAAGTGAATGGACAGGA  
TGGGAAGGTATTGCTAGCGAGAGTGAATGTAGAGGGTTCGCCCGGATTTCTCGCGGAACAGTTTGATGTA  
AGCTCGATTCCCCTGTGATGTGCTGGCTGGCAGGAGAGGTTGTTGACCGTTTTGAGGGCGACGTGGAAG  
ACACGAAAGATTGACCAAATCATATCCAAATTTGGTGGAATATCAAACTGGAAAATGAATGATAAAGGGCTTG  
AACAGCCATTTAGTAGACGAAAAAAAAAAAAAAAAAAAA
```

```
>gi|159498511|gb|EY195365.1|EY195365 RSAA-aab21e05.g1 R.similis_EST_RSAA Radopholus similis cDNA 5', mRNA sequence  
GGCCGGGTGCTGCAGGCATGTGCGACCGTCCCGCCACTCATCATCGTCCCTTCCCTGTCCCCGCCCGGCA  
CACCTTCCCTGTCCAGACATTCAACTGCGGTGTGGTTCGAAAGTCTCTCCCAAAGTCAATACCCGATGCTCCG  
CCGCCATACGAGGAGTTTGTGCGTGTTCACCACCACCGCCACAAAAGGGCACCGCCATTCTGACGCGGG  
AAGAGGATGAGGAGTTGCAGAACAGACTGAACTCGGAGAGAGAGAGGGAAGTGAAGGACTGAGCGACTGGTGACATTT  
GGTTTTGTGCGAGTGTGCTGCAGCTTCGCACCATTTCCCTTTATATACGGGACTTTTCTTCATTTCTTTGTTC  
CTGACTTAACAACAATTAATAGACCA
```

```
>gi|159498510|gb|EY195364.1|EY195364 RSAA-aab21e04.g1 R.similis_EST_RSAA Radopholus similis cDNA 5' similar to ref|NP_4  
GGCCGGGGACGCCATTGTATTTTGGTTCGATGTAGCCGACCTGGAACGTATTCAGGAAGCAAAGGGAGGA  
ATTGTGGAGTCTGATGCAGGATGAACAGGTGGCAAAGTGCACCTGTGCTTGTTTTGGGCAATAAGATCGAC  
AAGCCGAATGCTCTCAGCGAAGACCAGCTCAAGTACTACCTCGGCATCCAACAATACTGCACAGGAAAAG  
GCCAAGTTGCGCGCTCAGATCTGGCCACTCGTCCTTTGGATGTGTTTCATGTGCTCAGTCCTTAGGCGACA  
GGGTTACGGCGAAGGATTCTGTTGGCTCTCACAAATATCTGGACTGATTGAACGGCCCTCGGAAAGTTGAAA  
-----
```





NCBI

PubMed Entrez BLAST OMIM Taxonomy Structure

VecScreen

▶ **Screen a Sequence Using VecScreen**

Enter your query sequence below as an Accession, GI, or [FASTA](#).

Run VecScreen Clear Input

▶ **About VecScreen**

[VecScreen](#) is a system for quickly identifying segments of a nucleic acid sequence that may be of vector origin. NCBI developed VecScreen to combat the problem of vector [contamination](#) in public sequence databases. This Web page is designed to help researchers identify and remove any segments of vector origin before sequence analysis or submission.

Failure to recognize foreign segments in a sequence can:



The screenshot shows a web browser window with several tabs: "Dominant Markers vs. Co-Do...", "1. INTRODUCTION", "Moby portal - trimest", and two "(Untitled)" tabs. The main content area is titled "Moby@pasteur" and features a navigation bar with "Welcome", "Programs", "Data Bookmarks", "Jobs", and "Tutorials". A "set email" link is visible in the top right.

On the left side, there is a "Programs" sidebar with a search bar and a list of categories: alignment, assembly, database, display, hmm, nucleic, phylogeny, protein, sequence, structure, LIPM, and RPBS. Below this are sections for "Data Bookmarks", "Jobs", and "Tutorials", each with a "refresh" button. The "Tutorials" section lists links for "How to use Moby? A step by step tutorial", "Registration information", "Sequence formats", and "Alignment formats".

The main content area displays the "trimest" program page. It has a title "trimest" and a subtitle "Remove poly-A tails from nucleotide sequences". There are "Reset" and "Run" buttons. Below the title is an "Input section" with a label "* Sequence option (Nucleic Sequence)". Underneath, there are radio buttons for "Paste" (selected), "DB", and "File". A large empty text area is provided for input.

The Windows taskbar at the bottom shows the Start button and several open applications: "My Co...", "official", "Prese...", "SNP - ...", "SSR - ...", "Moby...", "Downl...", "desktop", "kadv...", "s4-03...", "r.simili...", and "rsimilis...".



[Home](#) | [Back](#) | [Forward](#) | [Refresh](#) | [Stop](#) | [Close](#)

[Most Visited](#) | [\(Untitled\)](#) | [http://www.scribd.co...](#) | [http://www.microrna...](#) | [http://genes.mit.edu/...](#) | [http://bioinfo.noble.o...](#) | [http://www.mirz.unib...](#) | [http://](#)

[Dominant Markers vs. ...](#) | [1. INTRODUCTION](#) | [Moby portal - cap3](#) | [\(Untitled\)](#) | [\(Untitled\)](#) | [\(Un...](#)

Moby@pasteur

[Welcome](#) | [Programs](#) | [Data Bookmarks](#) | [Jobs](#) | [Tutorials](#)

CAP3 x

CAP3

Contig Assembly Program

* File of reads (Nucleic Sequence) _____

Paste | DB | File

Clipping of poor regions _____

Clipping of poor and regions _____

Done

[Start](#) | [My Co...](#) | [official](#) | [Prese...](#) | [SNP - ...](#) | [SSR - ...](#) | [Moby...](#) | [Downl...](#) | [desktop](#) | [kadav...](#) | [s4-03...](#) | [r.si](#)

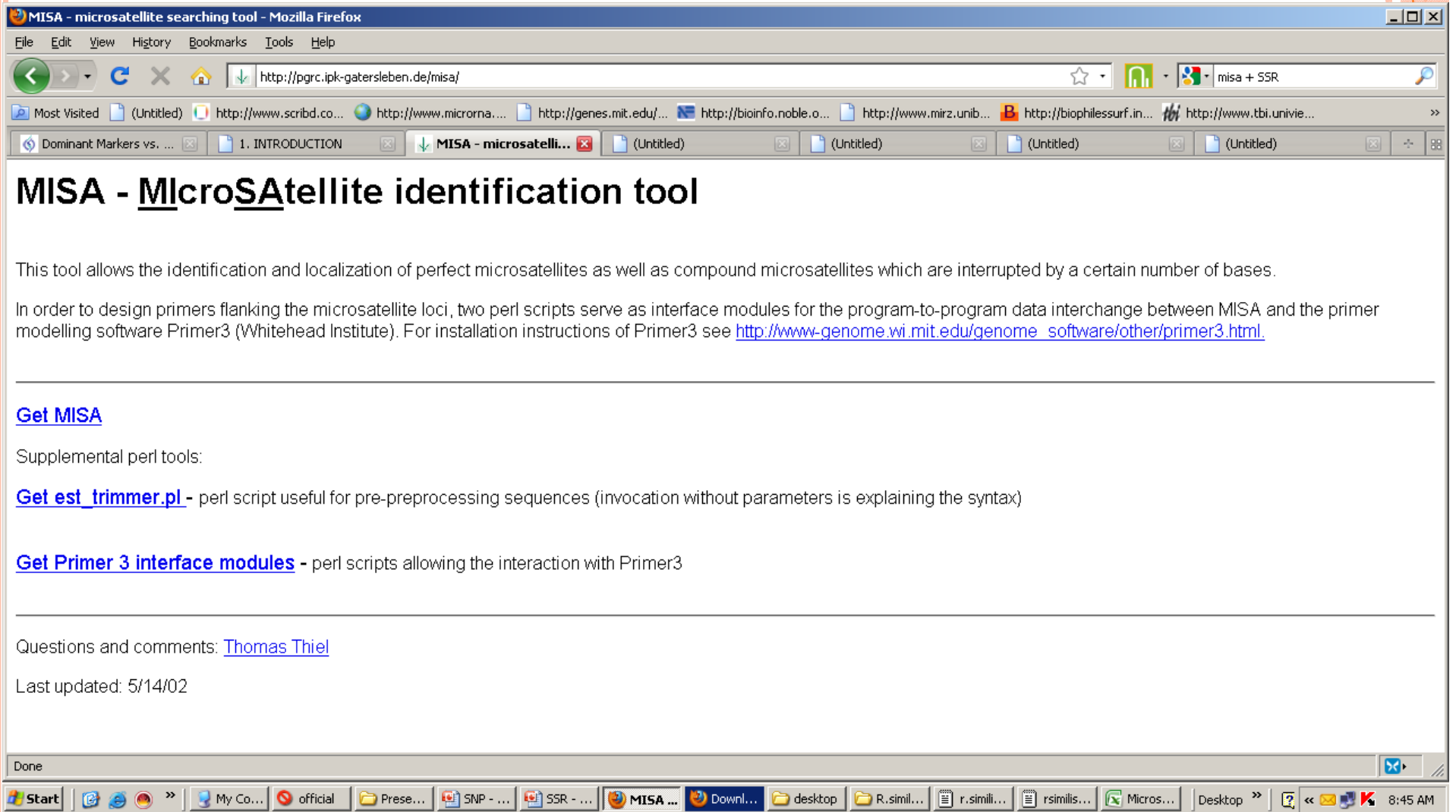
Total No of ESTs - 7380
Contigs -
Signlets -



SSR detection using MISA

MISA - MicroSatellite identification tool





MISA - microsatellite searching tool - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://pgrc.ipk-gatersleben.de/misa/

Most Visited (Untitled) http://www.scribd.co... http://www.microna... http://genes.mit.edu/... http://bioinfo.noble.o... http://www.mirz.unib... http://biophilessurf.in... http://www.tbi.univie...

Dominant Markers vs. ... 1. INTRODUCTION MISA - microsatelli... (Untitled) (Untitled) (Untitled) (Untitled)

MISA - MicroSatellite identification tool

This tool allows the identification and localization of perfect microsatellites as well as compound microsatellites which are interrupted by a certain number of bases.

In order to design primers flanking the microsatellite loci, two perl scripts serve as interface modules for the program-to-program data interchange between MISA and the primer modelling software Primer3 (Whitehead Institute). For installation instructions of Primer3 see http://www-genome.wi.mit.edu/genome_software/other/primer3.html.

[Get MISA](#)

Supplemental perl tools:

[Get est_trimmer.pl](#) - perl script useful for pre-preprocessing sequences (invocation without parameters is explaining the syntax)

[Get Primer 3 interface modules](#) - perl scripts allowing the interaction with Primer3

Questions and comments: [Thomas Thiel](#)

Last updated: 5/14/02

Done

Start My Co... official Prese... SNP - ... SSR - ... MISA ... Downl... desktop R.simil... r.simili... rsimilis... Micros... Desktop 8:45 AM



MISA - MicroSatellite identification tool

[Download MISA](#)

[Download misa.ini](#) - example of a specification file

Syntax: misa.pl <FASTAfile>

Declarations:

- <FASTAfile> is the filename containing DNA sequences in FASTA format.
- An additional file containing the search parameters is required named "misa.ini".
 - In a single line beginning with "def" a sequence of number pairs is expected, whereas the first number defines the unit sizes and the second number the lower threshold of repeats for that specific unit.
 - In a single line beginning with "int" a single number is expected defining the maximal number of bases between two adjacent microsatellites to be recognised as being a compound microsatellite.

<http://pgrc.ipk-gatersleben.de/misa/misa.html>



misa.ini Identification parameters

definition(unit_size, min_repeats): 1-10 2-6 3-5 4-5 5-5 6-5

interruptions(max_difference_between_2_SSRs): 100

definition(unit_size,min_repeats): 1-12 2-6 3-5 4-3 5-3 6-2

interruptions(max_difference_between_2_SSRs): 100

Monomer minimum 12 bp

Dimer minimum 12 bp

Trimer 15bp

Tetra 12 bp

Penta 15bp

Hexa 12 repeats



syntax

misa.pl <**FASTAfile**>

misa.pl rsimilis.fasta



	A	B	C	D	E	F	G	H	I
1	ID	SSR nr.	SSR type	SSR	size	start	end		
2	EY194446	1	c	(GGCCGG)	48	1	48		
3	EY194450	1	c	(CGGGAA)	164	4	167		
4	EY190719	1	c	(GGGAAC)	102	7	108		
5	EY191483	1	c	(GCAACA)	121	8	128		
6	Contig56	1	c	(ATCACA)	90	9	98		
7	EY192307	1	c	(GGA)5cga	36	10	45		
8	EY191149	1	c	(AAATTA)	34	11	44		
9	EY190548	1	c	(T)18catcc	106	11	116		
10	Contig812	1	c	(T)17caact	49	12	60		
11	EY191543	1	c	(A)15ctaa	46	13	58		
12	Contig727	1	c	(GTCATT)2	50	15	64		
13	EY195239	1	c	(TCAAAT):	85	15	99		
14	EY191977	1	c	(AAGGCA)	89	19	107		
15	EY195150	1	p6	(TTTAAA)2	12	20	31		
16	EY191512	1	p6	(TGGATG)	12	20	31		
17	EY191257	1	p6	(TGATGC):	12	21	32		
18	EY190935	1	p6	(TCCCAA):	12	21	32		
19	CO897777	1	p6	(TAAGCT):	12	21	32		
20	EY193889	1	p6	(TCTTAC)2	12	23	34		
21	EY192676	1	p6	(TAATTT)2	12	23	34		
22	EY194162	1	p5	(TAAAT)3	15	24	38		
23	EY190902	1	p5	(AAGCT)3	15	24	38		

Ready | | Chart4 | Chart3 | Chart2 | Chart1 | rsimilis.fasta |

RESULTS OF MICROSATELLITE SEARCH

=====

Total number of sequences examined:	7380
Total size of examined sequences (bp):	1706771
Total number of identified SSRs:	465
Number of SSR containing sequences:	413
Number of sequences containing more than 1 SSR:	41
Number of SSRs present in compound formation:	32

Distribution to different repeat type classes

Unit size	Number of SSRs
1	408
2	15
3	37
4	5

Frequency and most abundant repeats were found out.

For mononucleotides, although A, T, C and G are possible,

A and T are grouped into a single category, G and C since an A repeat on a strand is the same as a T repeat on the opposite strand, and a C on a strand is the same as a G on the opposite strand, resulting in two unique classes of mononucleotides,

A/T and C/G

all dinucleotide motifs were grouped into the following four unique classes

(i) AT/TA, (ii) AG/GA/CT/TC, (iii) AC/CA/TG/GT, and (iv) GC/CG.



The trinucleotide repeats are grouped into 10 unique classes

AAG/TTC contains AAG/AGA/GAA/CTT/TTC/TCT SSRs

(Jurka and Pethiygoda 1995; Katti et al. 2001)



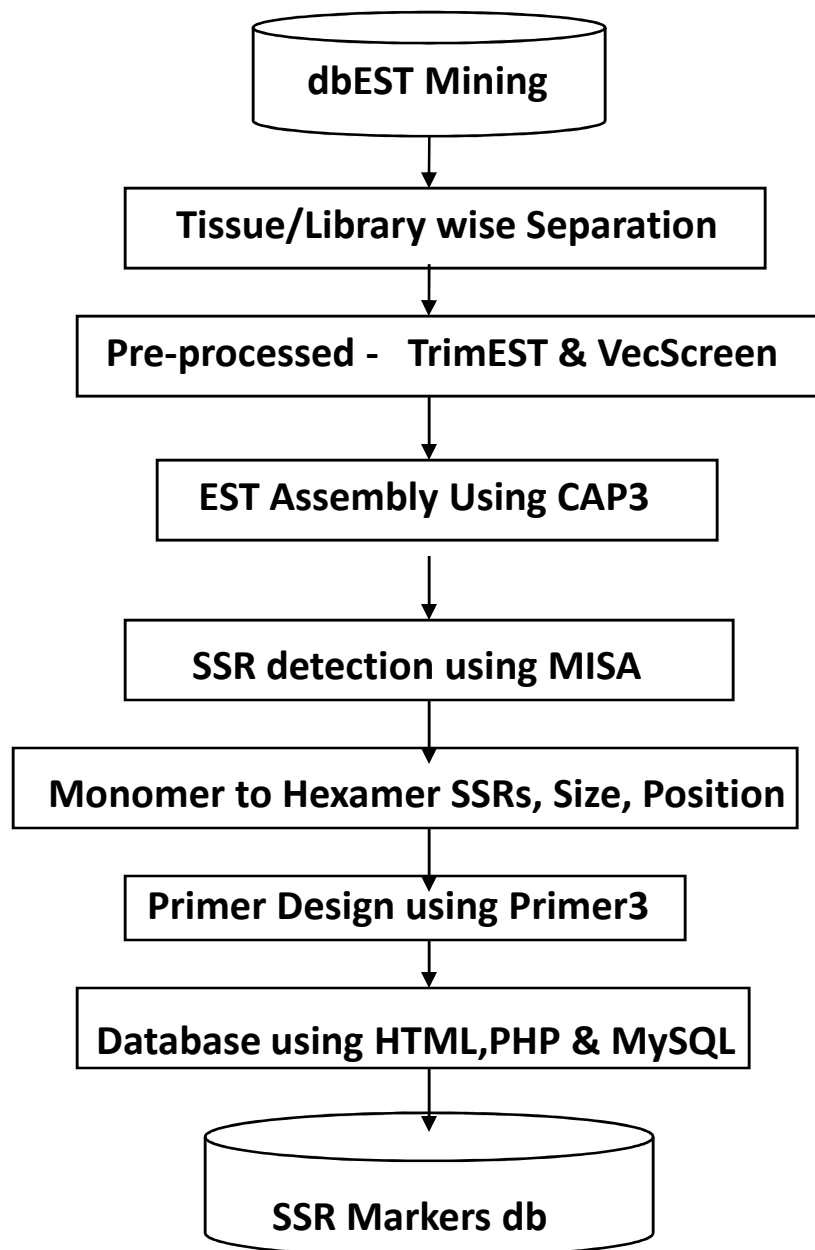


Fig. 1. Flowchart of various steps involved in *in silico* SSR discovery and database development

Thank You