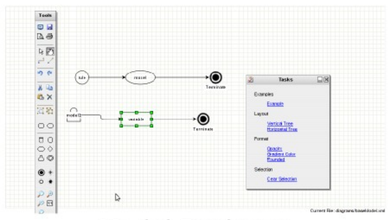
Exploring Transposons From MITE Family in The Human Genome with Logical Models

Isabelle STÉVANT[1], Catherine BELLEANNÉE[2]
[1] Master's degree Systems Biology Modelling, University of Rennes1
[2] Symbiose Project Team, INRIA/Irisa, Campus de Beaulieu 35042 Rennes Cedex

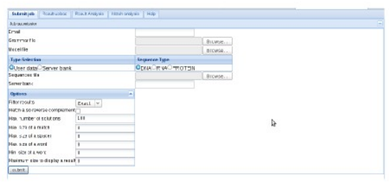
Logol : A New Pattern Matching Software Suite

A complete software suite online

Pattern matching software for DNA, RNA and Proteins
Design your model with the graphical editor
Submit your model on a given sequence or on a DataBase
View and export your results with the Result Analyser



Model Designer



Result Analyser

Beyond Regular Patterns

Dedicated constrained string language
Express sequence variability through specific operators
Based on String Variable Grammar (SVG) by David Searl
Able to model copies of a variable sequence

➔ Exemple of Logol operators :

$[0,5] : Substitutions (between 0 & 5 allowed)
£[0,2] : Insertion/deletion
#[0,30] : Size (between 0 & 30 letters)
-''wc'' X : reverse-complement of a variable sequence X

Available on Genouest Platform:
http://webapps.genouest.org/LogolDesigner/

Biological Model : *miHsmar1*

ANR Modulome : collaboration of Biology and Computing in order to have a best comprehension of genome structures
GICC team, Tours : Study of *miHsmar1* Transposon
Symbiose team, INRIA Rennes : Pattern Matching Tools development

What is a Transposon?

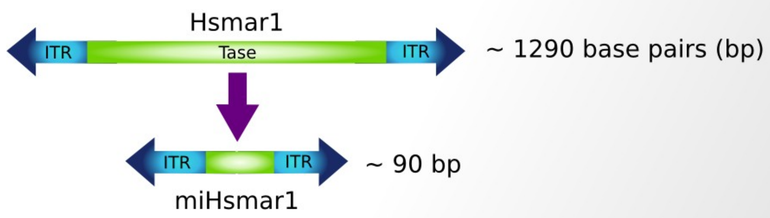
DNA fragment able to move within the genome
First discovered by Barbara McClintock in 1951 on maize
Classification :

- Class I : « copy/paste »
- Class II : « cut/paste » (ex: *Hsmar1*)
- Unclassifiables : MITEs (Miniature Inverted-repeat Transposable Elements) (ex: *miHsmar1*)

Origin of *miHsmar1*

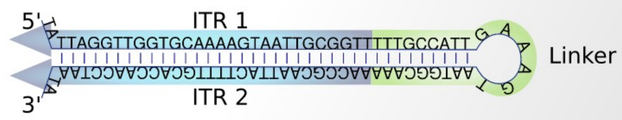
➔ *Hsmar1* :

Homo sapiens mariner 1
First discovered Class II Transposon from *Tc1-Mariner* family
Bordered by *Inverted Terminal Repeat* sequences (ITR)
Contain a gene which code for a transposase (Tase)



➔ *miHsmar1* :

mini Homo sapiens mariner 1
Troncated *Hsmar1* with two ITRs but no gene
About 90 bp but still able to transpose
Secondary structure in "hairpin"



Different level of conservation within the miHsmar1 sequence :

Model Design

Cutting *miHsmar1*

➔ Considering the conservation degree :

Modelling ITR1 :

```
"tattaggtt":{$[0,2],£[0,2]},
"ggtgcaaaagtaattgcggtt":{$[0,12],£[0,2]}
```

Modelling the Linker as a spacer from 0 to 35 bp :

```
LOGOLVAR%Spacer%:{#[0,35]}
```

Modelling ITR2 :

ITR2 = reverse complement of ITR1

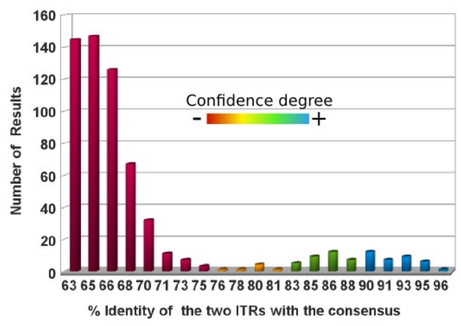
```
-"wc" "ggtgcaaaagtaattgcggtt":{$[0,12],£[0,2]},
-"wc" "tattaggtt":{$[0,2],£[0,2]}
```

First Results

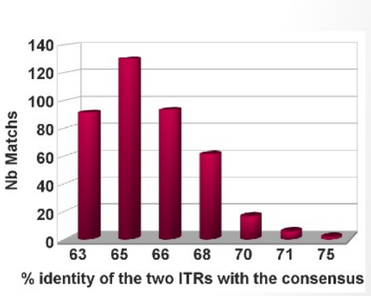
Raw Results Filtering

Extraction of the sequences found by Logol
Mass alignement of each ITRs sequence with the consensus
Deletion of the results with less than 60% of identity

Results distribution functions of the ITRs identity



Chromosome 17
Homo sapiens



Chromosome Shuffled

Conclusion

Thanks to its high sensitivity, Logol was able to detect 614 *miHsmar1* on the Human chromosome 17.
The results on a shuffled chromosome allow us to determine a detection threshold for high confident *miHsmar1*.
Automatization of the pipeline will allow to launch a study on a whole genome.