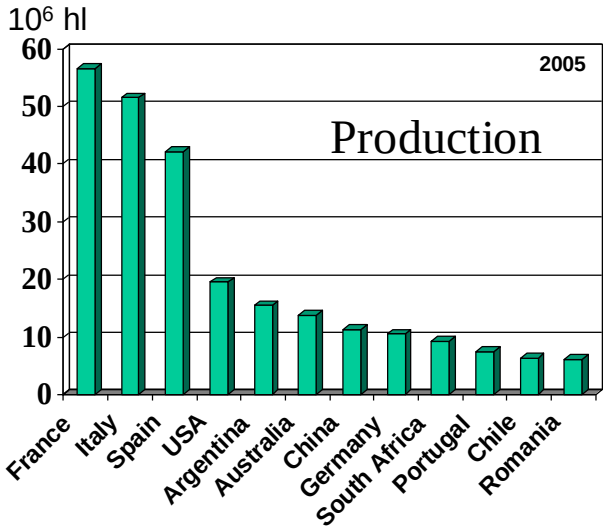
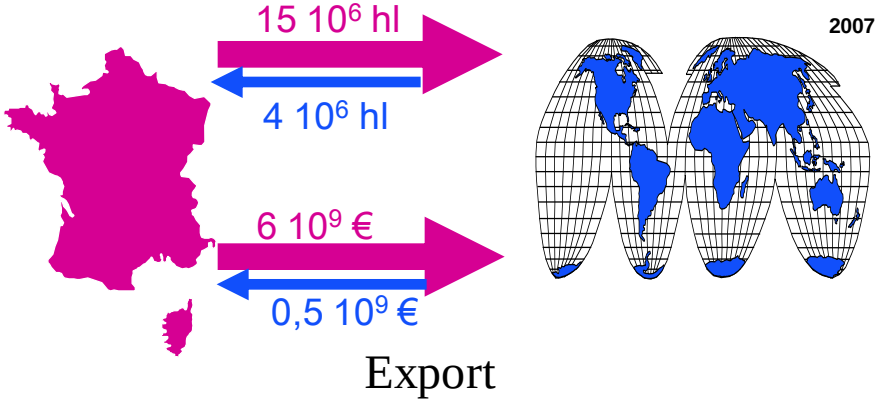
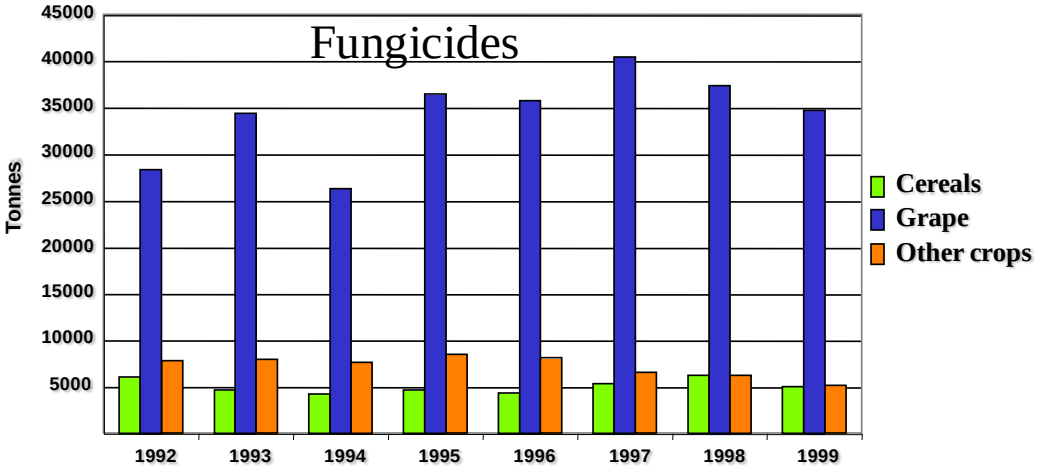
Structural and Functional Genomics in Grapevine through FLAGdb^{++}

Sandra Dèrozier, Cécile Guichard, Franck Samson, Jean-Philippe Tamby, Véronique Brunaud, Vincent Thareau, Christophe Caron, Maria Tchoumakov, Roberto Bacilieri, Anne-Françoise Adam-Blondon and Sébastien Aubourg

Why genomics on *Vitis vinifera* ?







2n = 38
475 Mb

Genome sequencing

Franco-Italian consortium
Coord : A.-F. Adam-Blondon

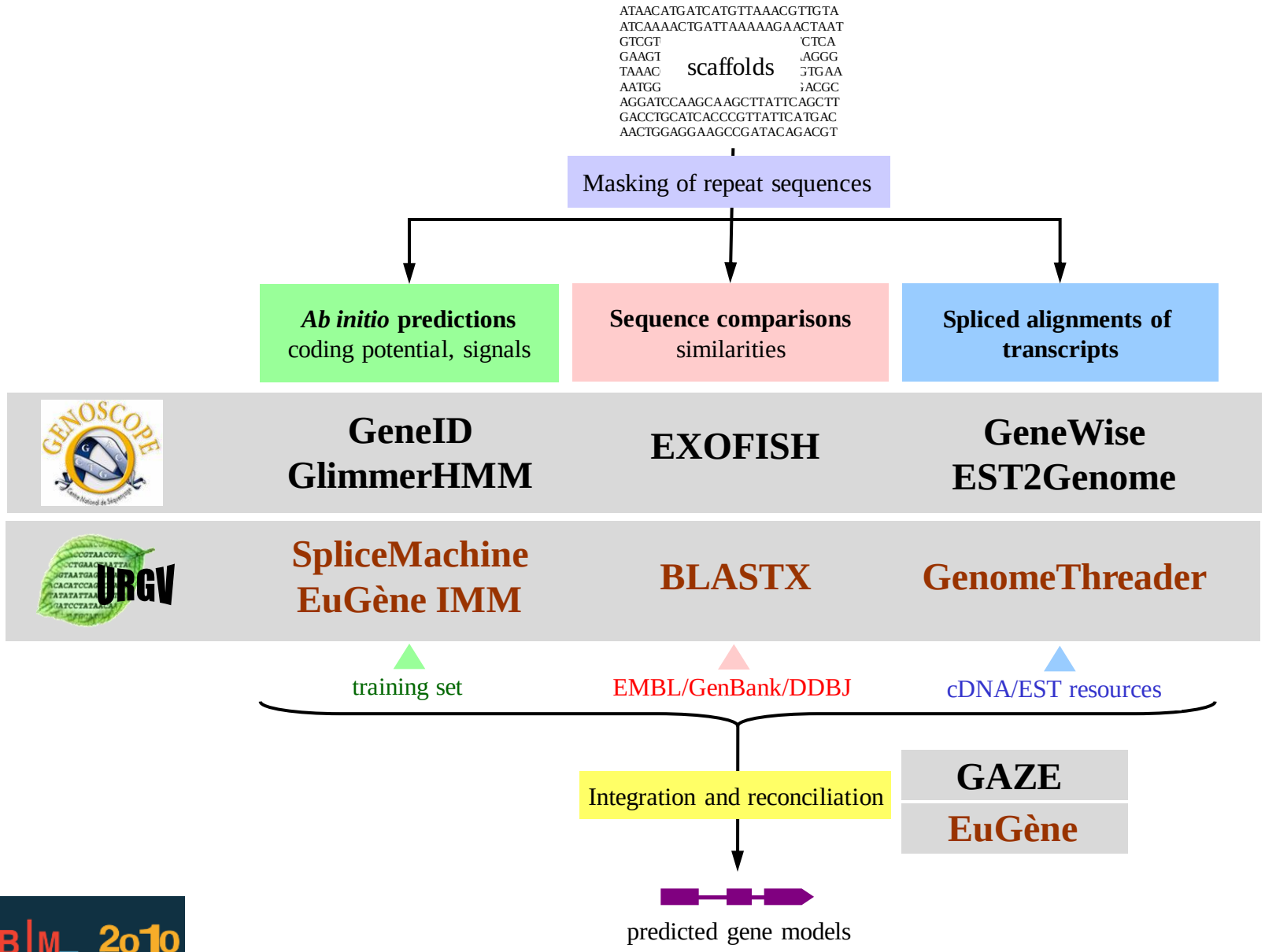
France		Italy	
Génoscope	7 X	Padova University	2.5 X
INRA		Udine University	2.5 X
		Milano University	

Sanger after WGS

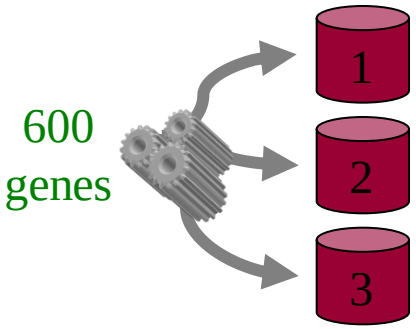
8 X assembly (Jaillon *et al.*, Nature 2007)
12 X assembly (2009)

Pinot Noir → PN40024 (93% homozygote)

Genome annotation strategy



Training, evaluation and results



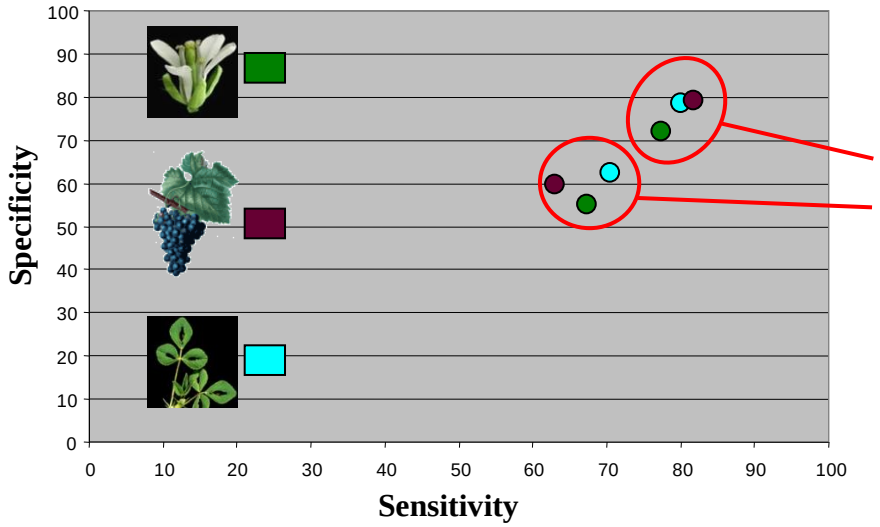
Training of EuGène (MM) and SpliceMachine ◀ 4500 splicing sites

EuGène (combiner) optimization

Prediction evaluation

EuGène requires
high computation facilities:

(MIG)



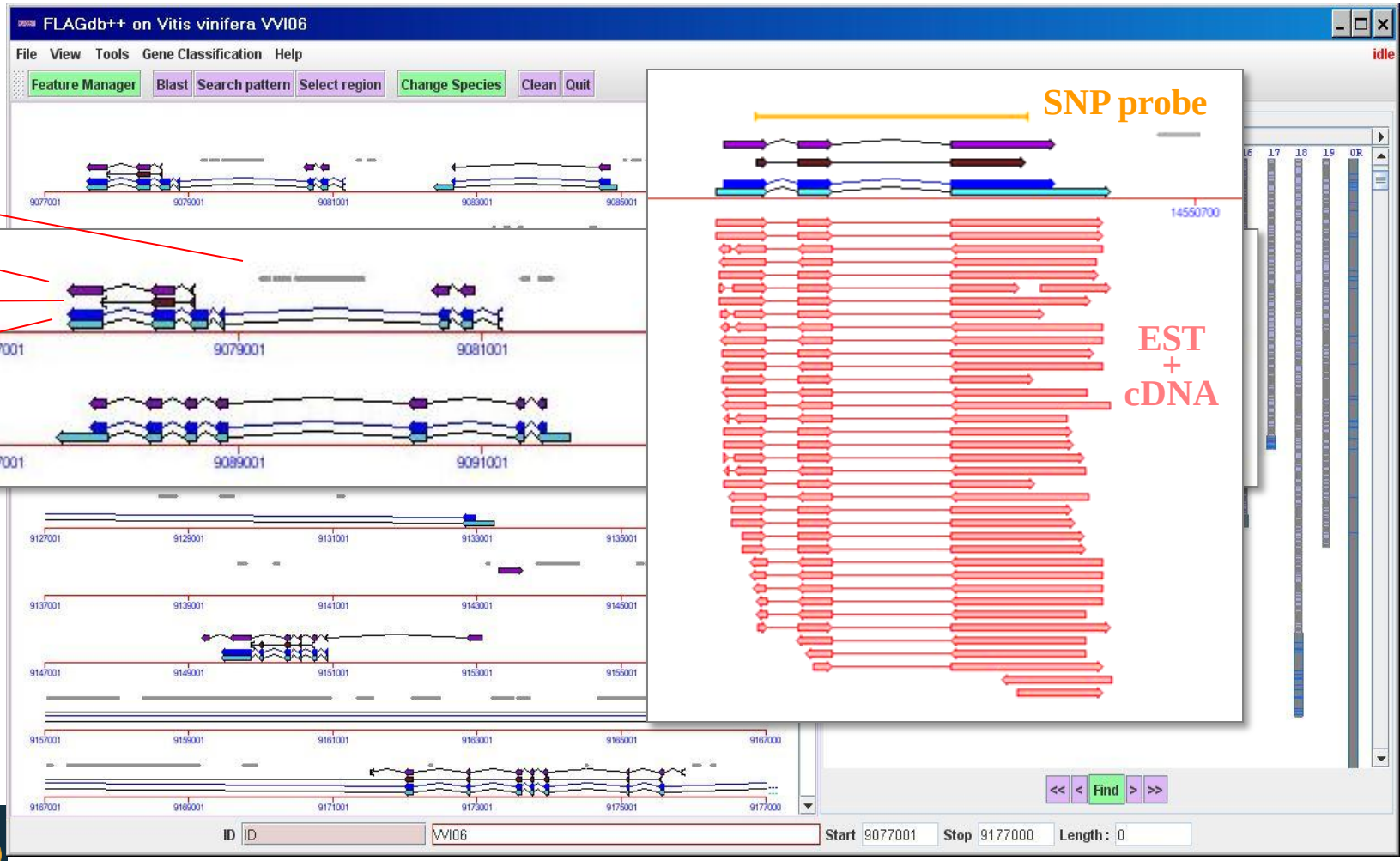
Scale : gene
Mode :
- with integration
- *ab initio* pure

EuGène predicts **44 414** coding genes in the 12x PN40024 genome
(GAZE : 26 347 genes)

Vitis vinifera in FLAGdb^{++}

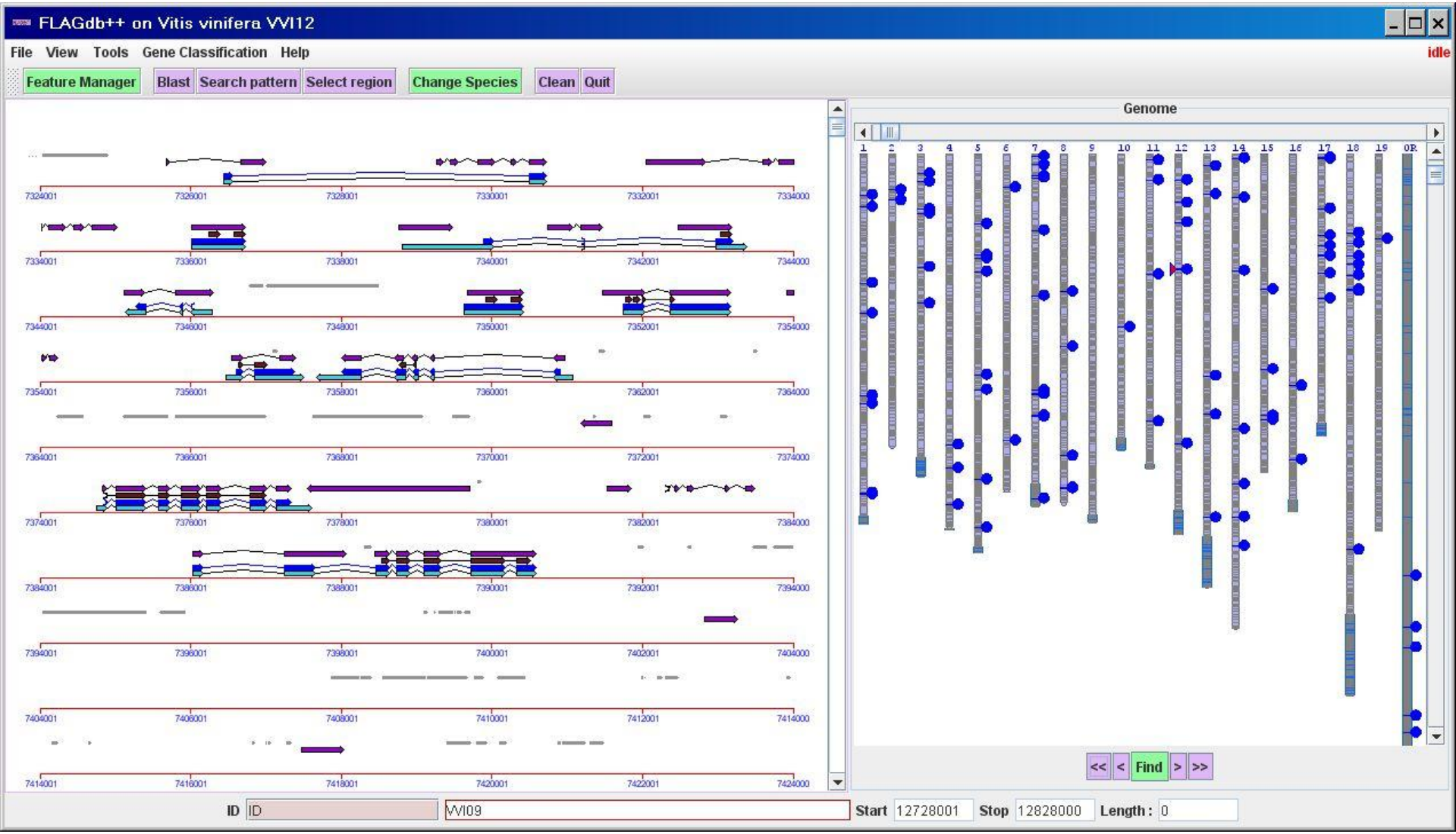
FLAGdb^{++}
Integrative Database around Plant Genomes

Relational database and JAVA application dedicated to integration and exploration of genomic data (experimental or predicted) in order to help the functional study of plant genes



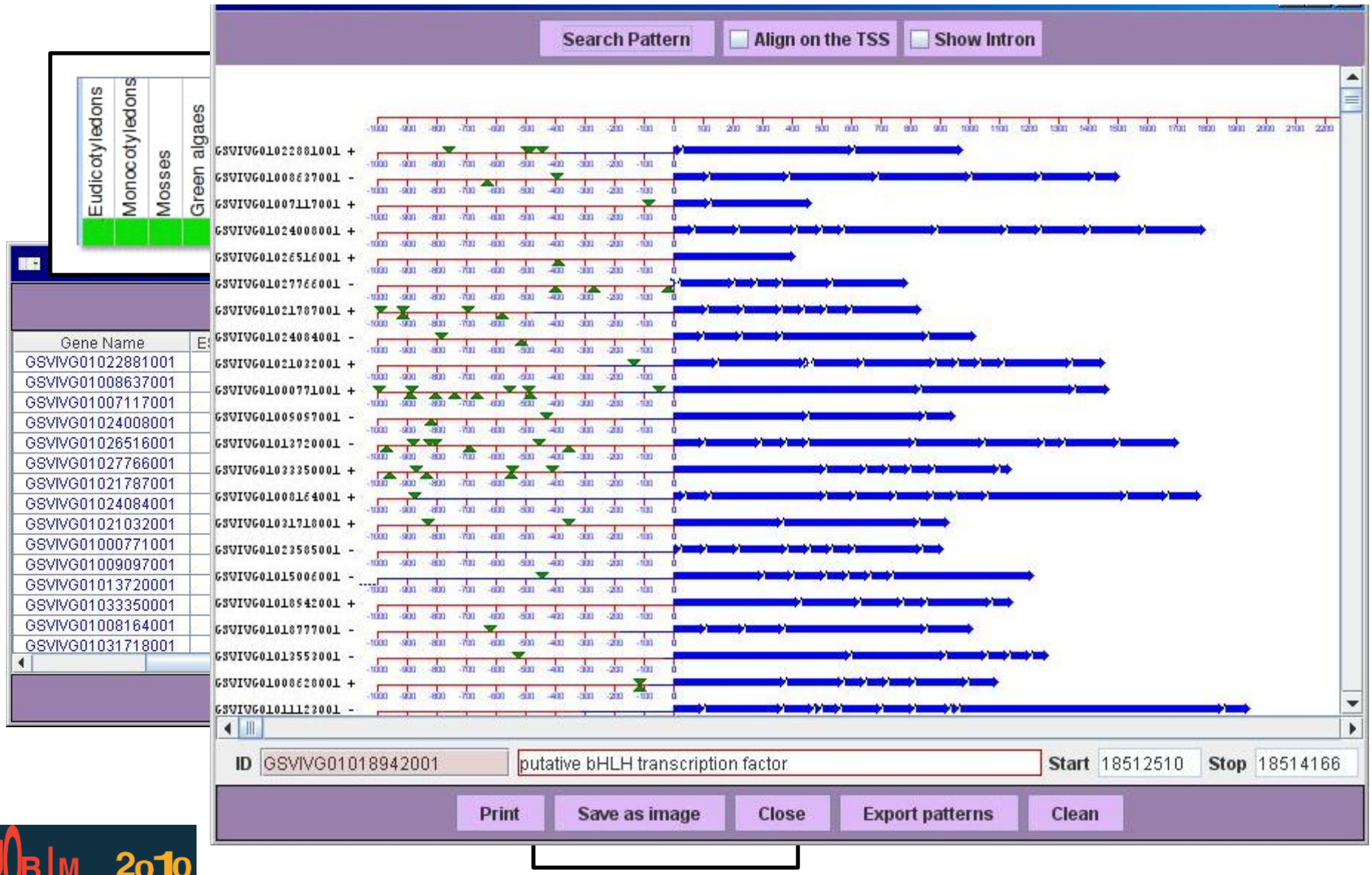
JOBIM MONTPELLIER 2010

Vitis vinifera in FLAGdb^{++}



Easy access to gene families or large groups of genes

Vitis vinifera in FLAGdb^{++}



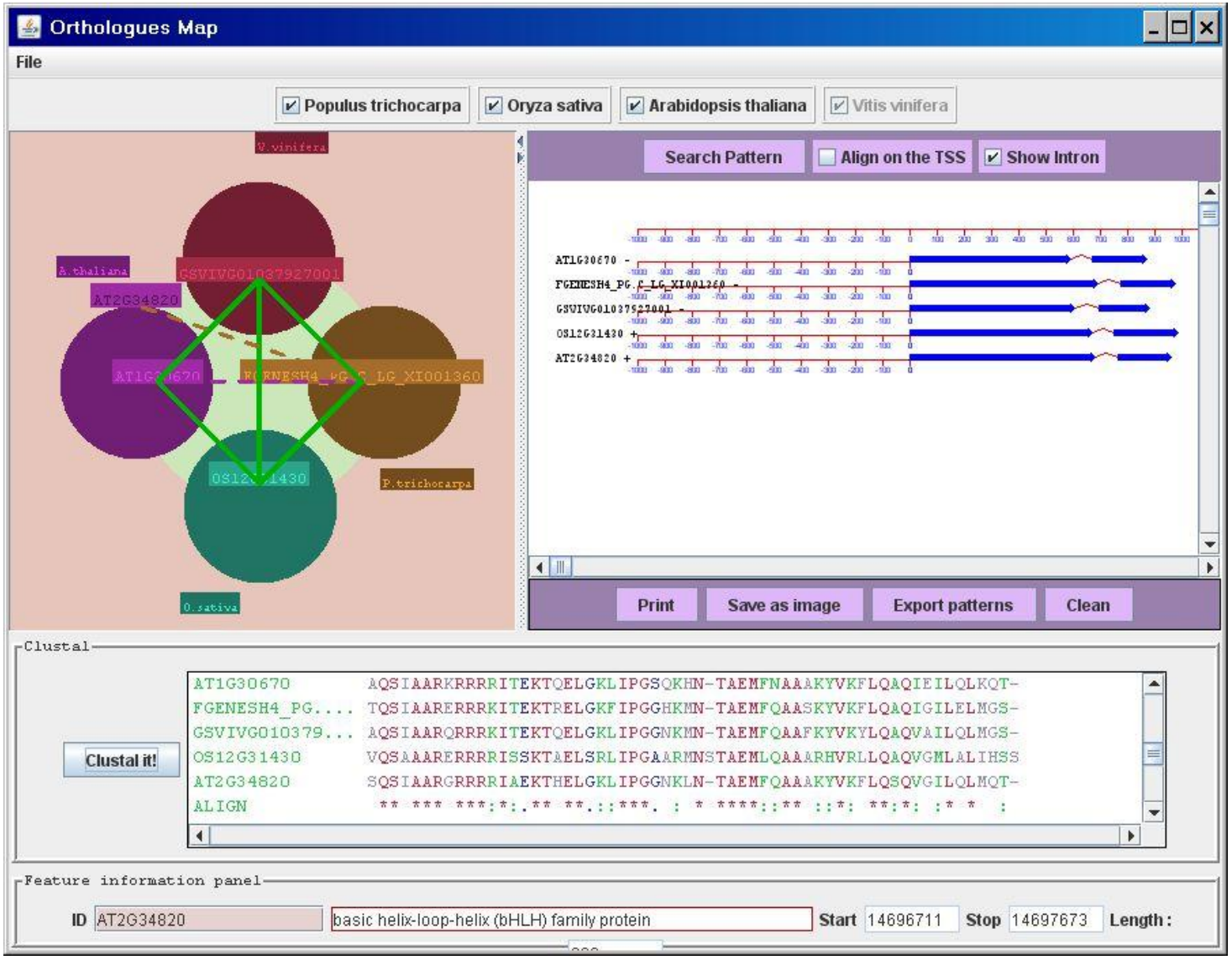
Vitis vinifera in FLAGdb^{++}

New tool dedicated to orthology relationships between the 4 plant genomes hosted in the database

- Graphical display of Reciprocal Blast Hits

- Comparison of gene structures and promoters

- Global protein alignments



From grape genome to wine flavours...

```
CCAGTGTGCAGCTTGCAGGCGCCAAACGTGATGCCTTGCTTCTCAGTTTCAAAGACGCTAAGC
TGTCTGTGGTGGAGTATGACCCAGGCACTCATGACCTGAAGACCCTGTCTCTACACTACTTTG
AGGAGCCTGAACTTCGGGATGGGTTTGTGCAGAACGTGCATACGCCTCGTGTGCGGGTGGACC
CTGACGGGCGCTGTGCAGCAATGCTCATCTATGGCACGAGGCTGGTAGTTCTGCCTTTCCGCA
GAGAGAGCCTGGCTGAGGAACATGAGGGGCTCATGGGCGAAGGGCAGAGGTCTAGCTTCTTGC
CCAGCTACATCATTGATGTACGGGCTCTGGATGAGAAACTACTCAACATCATTGACCTGCAGT
TCCTGCATGGCTACTATGAGCCCACCCTGCTTATCCTGTTTGAGCCCAACCAGACTTGGCCAG
GGCGGGTGGCTGTGAGGCAGGACACGTGCTCCATTGTGGCTATCTCGCTGAACATCACACAGA
AAGTCCATCCAGTCATCTGGTCCCTCACCAGCTTGCCTTTTGACTGTACCCAGGCCCTGGCTG
TGCCCAAGCCCATAGGTGGGGTAGTGATCTTCGCTGTCAACTCACTGTTGTACCTGAACCAGA
GTGTTCCCCCATATGGTGTGGCTCTCAATAGTCTCACCACAGGCACCACCGCTTTCCCATTGC
GTACCCAAGAGGGTGTACGGATCACCCTAGACTGCGCACAGGCGGCCTTCATCTCTTATGACA
AGATGGTCATCTCCCTCAAGGGTGGTGAGATTTATGTGCTGACCCTCATCACTGACGGCATGC
GAAGTGTCCGAGCGTTTCACTTTGACAAGGCAGCTGCTAGTGTCCTTACCACCAGTATGGTCA
CAATGGAGCCCGGATACCTGTTCCTAGGCTCTCGCCTGGGCAATTCCCTCCTCCTCAAGTACA
CGGAGAAGCTGCAGGAGCCCCCAGCCAGCTCCGTCCGTGAGGCTGCTGACAAGGAAGAGCTGC
```

a gene family story

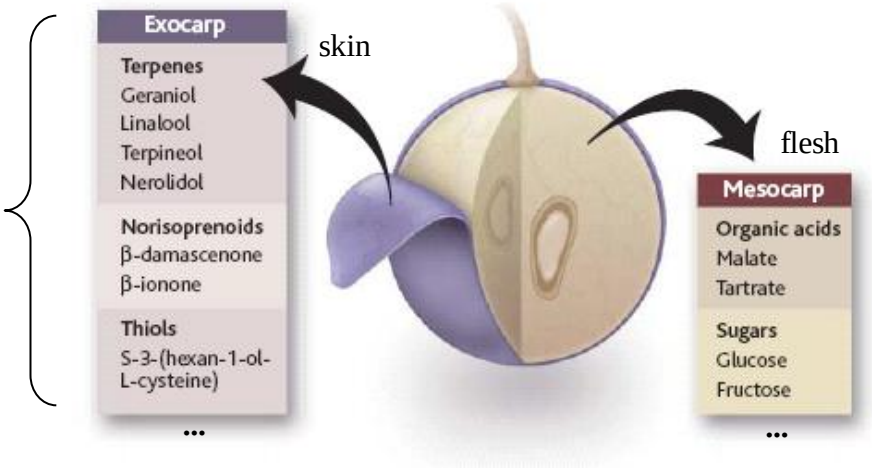
Many volatile organic compounds ► wine flavour (taste and aroma)

Stored as sugar or amino acid conjugates in vacuoles

glycosidases & peptidases

from grape, yeast and/or industrial enzymes

Volatilization of compounds essential for the flavour perception



Lund and Bohlmann (2006) Science

The terpenoid world

(22 000 different molecules !)

Gewürztraminer

▼

linalool
geraniol
nerol
citronellol
α-terpineol
…

▼

Distinctive floral character

Girard et al. (2002) Am. J. Enol. Vitic.

Cabernet Sauvignon anthers and pollen grains

▼

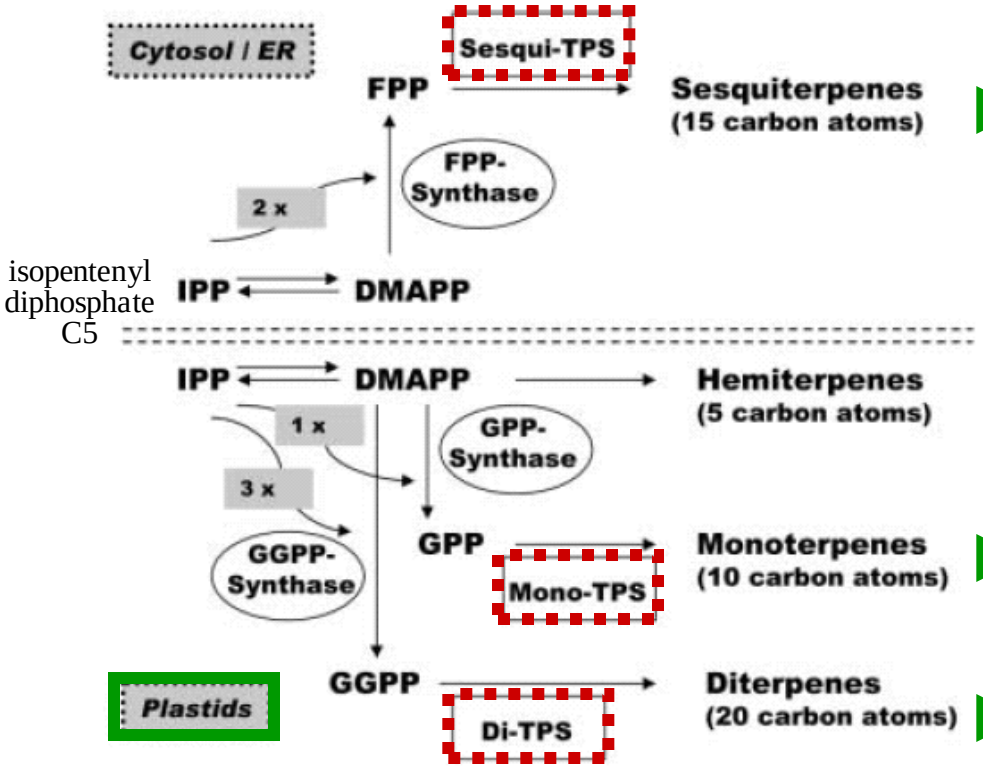
(+)-valencene
(-)-7-epi-α-selinene
…

▼

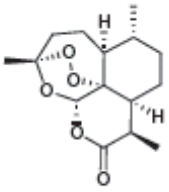
Protect reproductive tissues against pathogens
Allow pollinators to locate flowers

Martin et al. (2009) PNAS

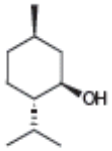
The terpenoid biosynthesis : TPS family



plant-pathogen relationships



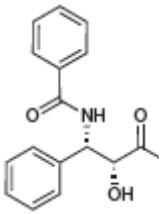
Artemisinin



Menthol

aroma
plant-pathogen relationships

gibberellic acid
plant-pathogen relationships



Paclitaxel

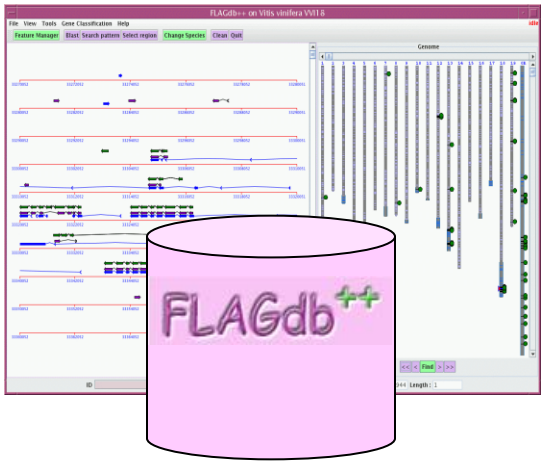
Expert annotation process

Known TPS proteins (Arabidopsis, Pinus, Citrus…)

TBLASTN

Vitis 12x genome
PN40024

152 loci TPS-like



- ✓ GAZE predictions
- ✓ EUGENE predictions
- ✓ PFAM motifs
- ✓ EST
- ✓ cDNA

From GB/EMBL & URGV

- ✓ BLASTX results
- ✓ TPS knowledge

Clean and exhaustive structural annotation of the Vitis TPS gene family

- Detect and correct erroneous annotations and missing genes
- Discriminate between pseudogenes, partial and complete genes

The Vitis TPS family

- GAZE fails to detect 54 of them
- EUGENE fails to detect 12 of them

- GAZE predicts perfectly 12 TPS (23%)
- EUGENE predicts perfectly 43 TPS (81%)

152 loci homologous to known TPS :

- **53 full** and perfect TPS genes
- **16 complete** TPS genes with only one punctual problem of sequence (sequencing error ?)
- **20 partial** TPS genes disrupted by unsequenced gap in the 12x assembly

53 **functional TPS** **89**

- **63 pseudogenes** with numerous frameshifts, stop codons and/or large deletion(s)

Arabidopsis

32 full TPS
8 pseudogenes

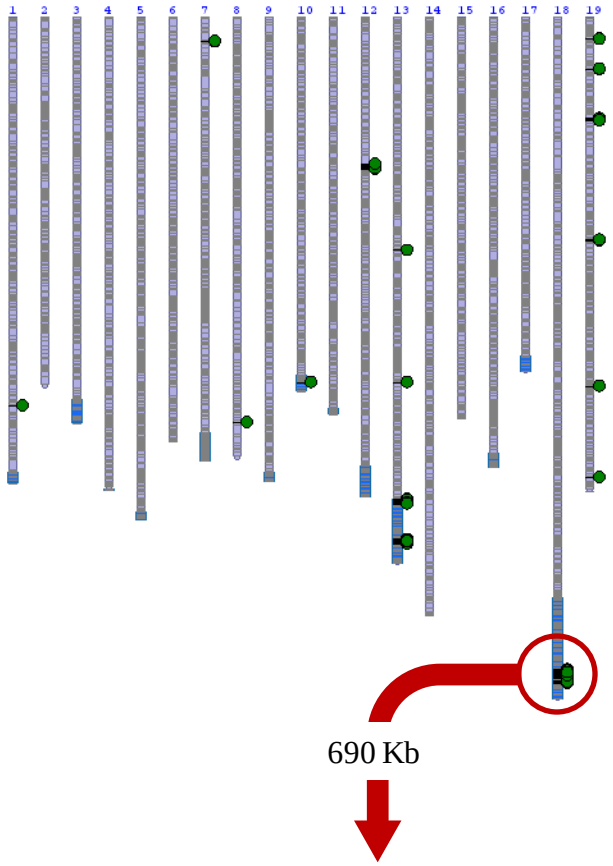
Rice

48 TPS loci
including pseudogenes

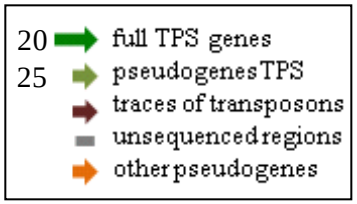
Poplar

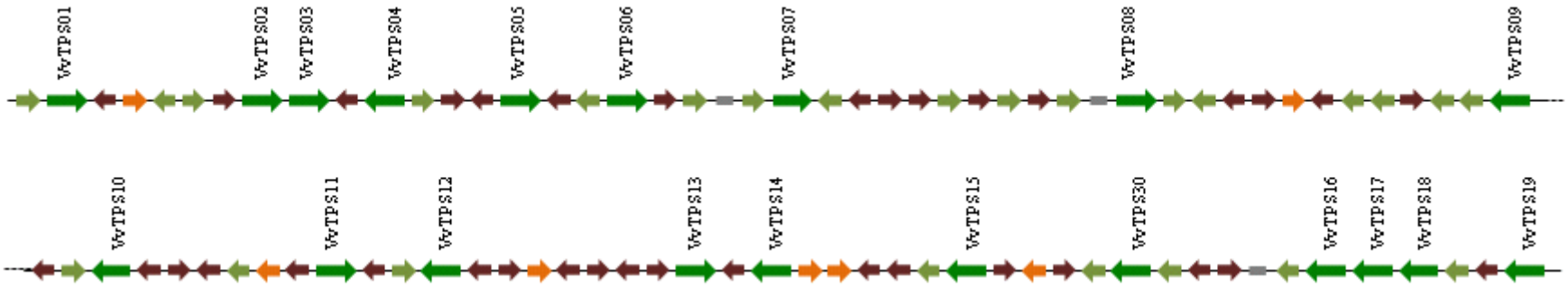
45 TPS loci
including pseudogenes
and partial genes

Topological organization of the family

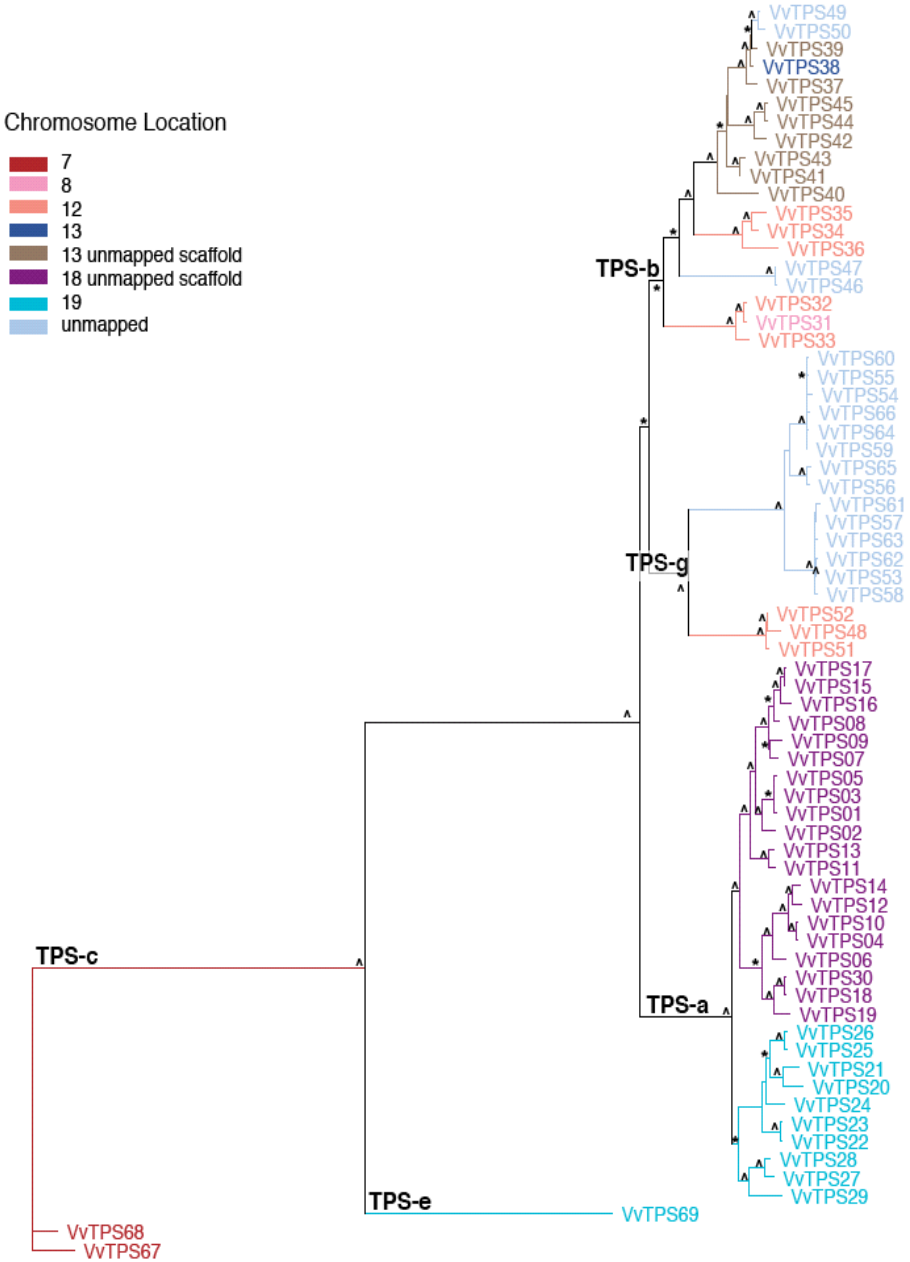


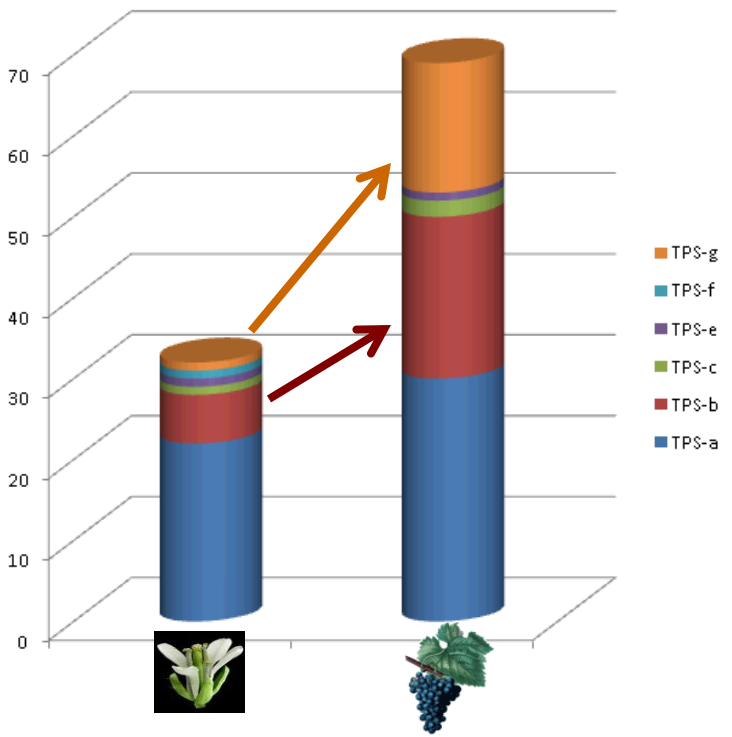
129 out of the 152 TPS loci (**85%**) are tandemly organized in 13 clusters from 2 to 45 genes.





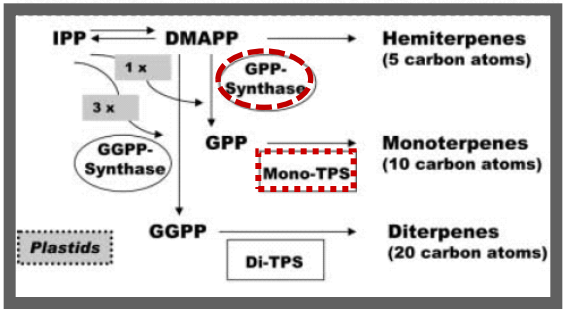
Vitis TPS classification





high diversity of monoterpenes in PN40024 ?

Geranyl diphosphate synthases (GPPS) family



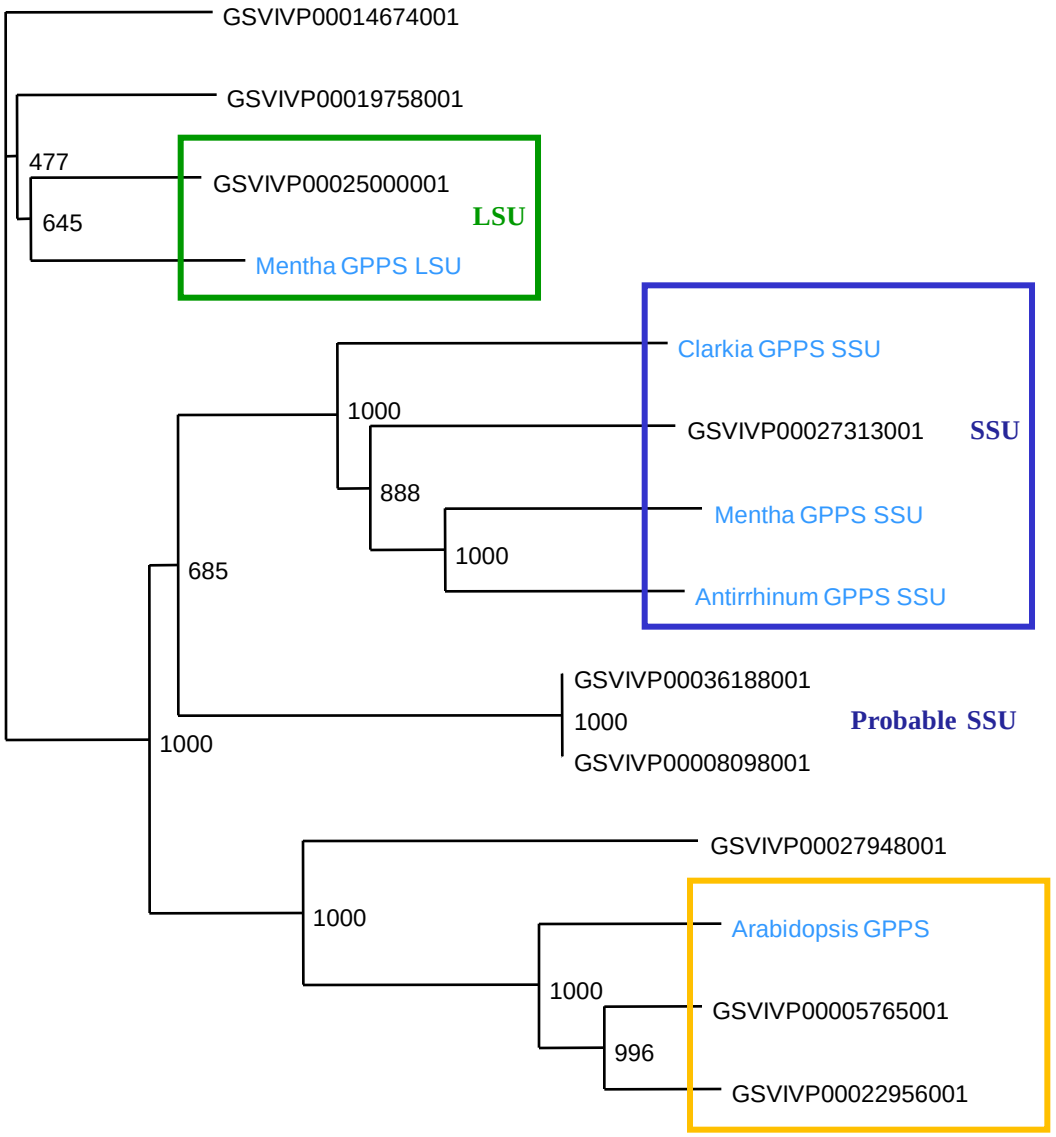
In *Arabidopsis thaliana* :
homodimeric GPPS

In *Mentha piperita* and *Clarkia breweri* :
heterodimeric GPPS
LSU + **SSU**

In PN40024, there are orthologous genes for the 2 forms of GPPS

▼

high diversity of monoterpenes



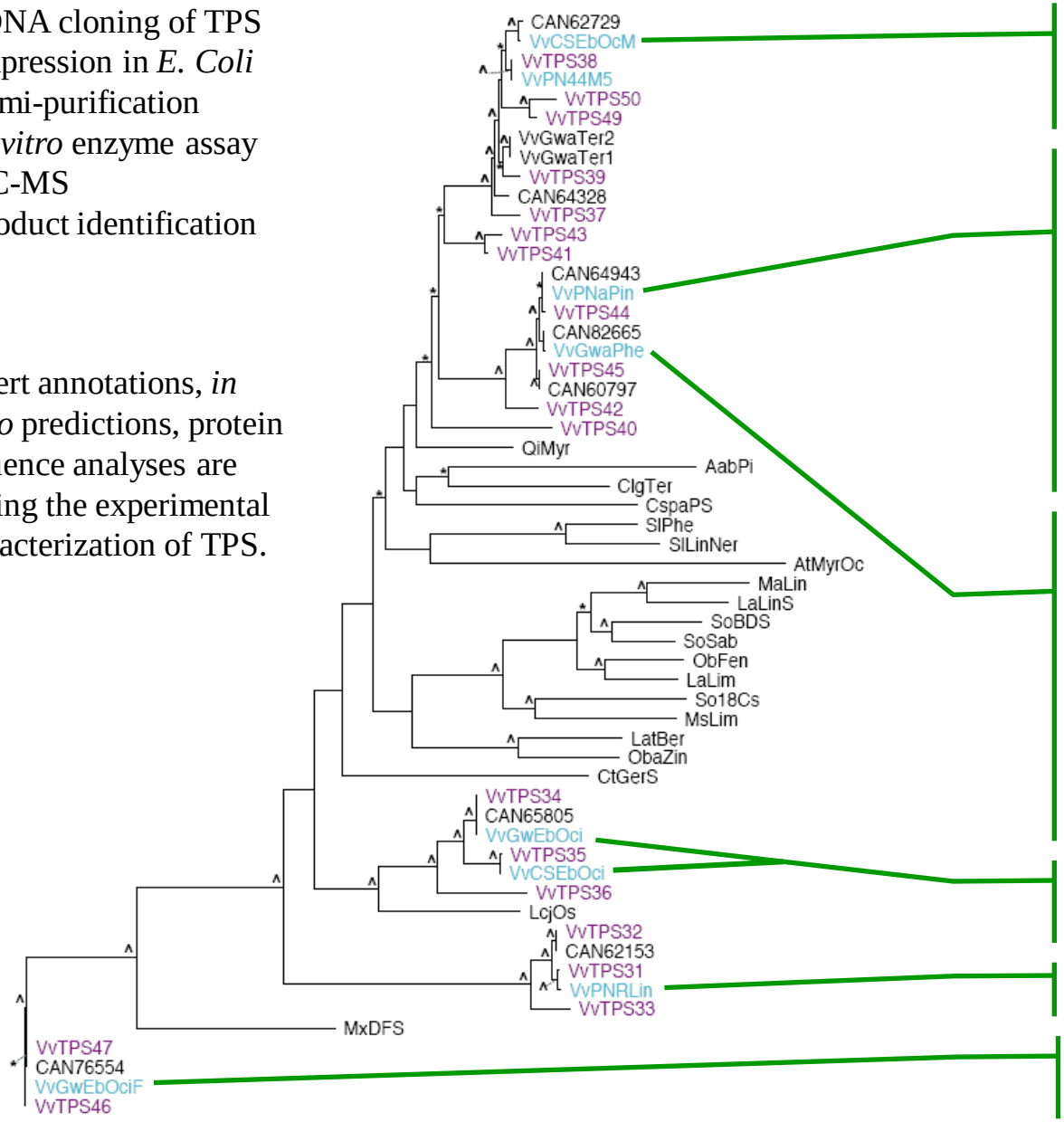
TPS gene structures

- predicted plastidial targeting peptide
  - mono- and diterpene synthases
- RRx(8)W motif
  - isomerization-cyclization reaction
- DDxxD motif
  - metal ion binding for cleavage of prenyl diphosphate substrates

Functional characterization of VvTPS genes

1. cDNA cloning of TPS
2. Expression in *E. Coli*
3. Semi-purification
4. *In vitro* enzyme assay
5. GC-MS
6. Product identification

Expert annotations, *in silico* predictions, protein sequence analyses are driving the experimental characterization of TPS.



(*E*)-β-ocimene	**53 %**
myrcene	42 %
β-pinene	3 %
limonene	2 %

(+)-α-pinene	**47 %**
(+)-limonene	25 %
(+)-camphene	9 %
myrcene	6 %
(+)-α-terpineol	5 %
(+)-sabinene	3 %
(3*S*)-linalool	2 %
(+)-β-pinene	1 %
(+)-α-phellandrene	1 %
(+)-α-thujene	1 %
α-terpinolene	1 %

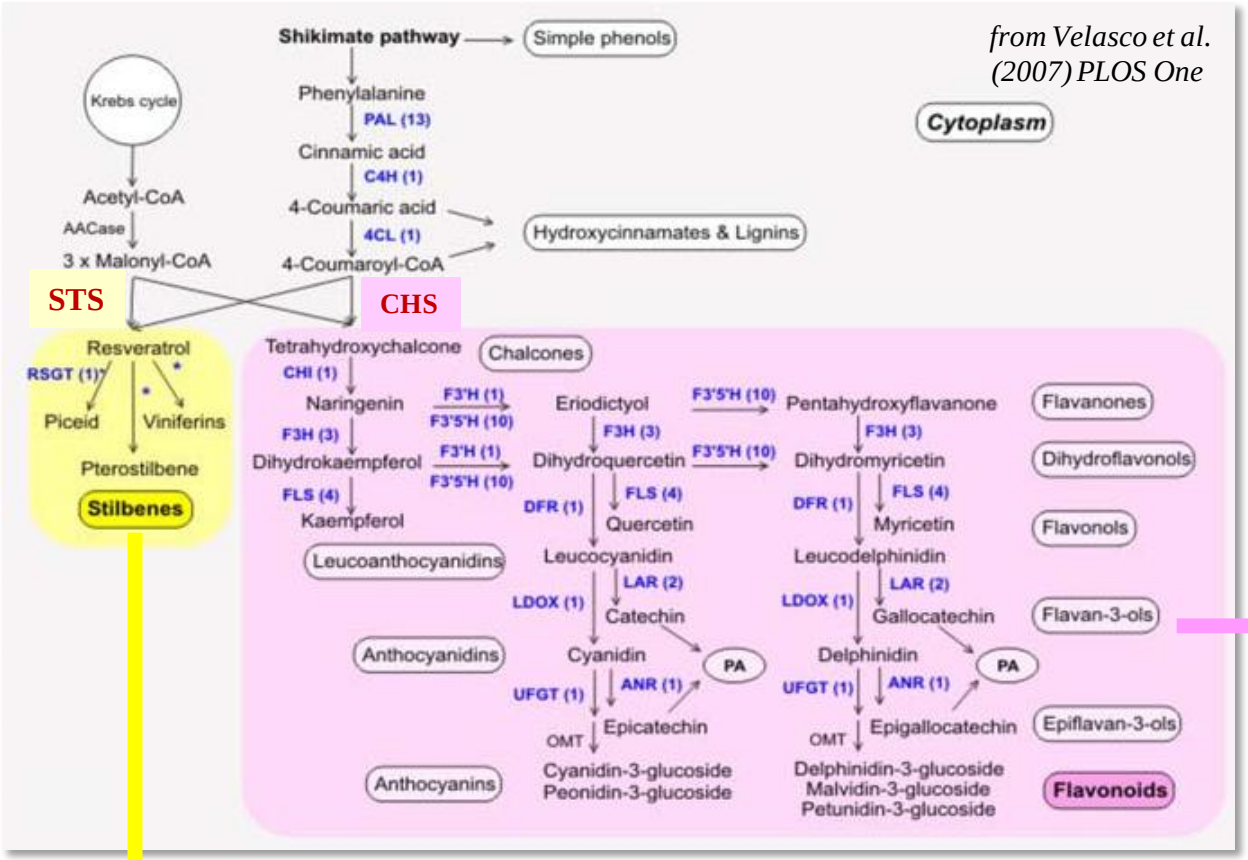
(+)-α-phellandrene	**40 %**
myrcene	15 %
terpinolene	15 %
α-terpinene	7 %
γ-terpinene	6 %
(+)-limonene	5 %
(+)-α-pinene	3 %
(+)-β-phellandrene	3 %
(+)-α-terpineol	3 %
(*E*)-β-ocimene	2 %

(*E*)-β-ocimene	**98 %**
(*Z*)-β-ocimene	2 %

***(3R)*-linalool**	**100%**

(*E*)-β-ocimene	**100%**	GDP
(*E*,*E*)-α-farnesene	**100%**	FDP

Other wine constituents…



from Velasco et al. (2007) PLOS One

Colour and astringent features of wines

Resveratrol is involved in health benefits : 'French paradox'
Prevents cancer and cardiovascular disease.
Stilbenes are also known as fungicides.

► 48 Stilbene Synthase genes !

Acknowledgements

French-Italian Public Consortium for Vitis
Anne-Françoise Adam-Blondon
Olivier Jaillon, Patrick Wincker *et al.*

Franck Samson
Cécile Guichard
Sandra Dèrozier
Jean-Philippe Tamby
Véronique Brunaud

Jérome Gouzy
Thomas Schiex

Joerg Bohlmann
Diane Martin

Christophe Caron

Philippe Hugueney

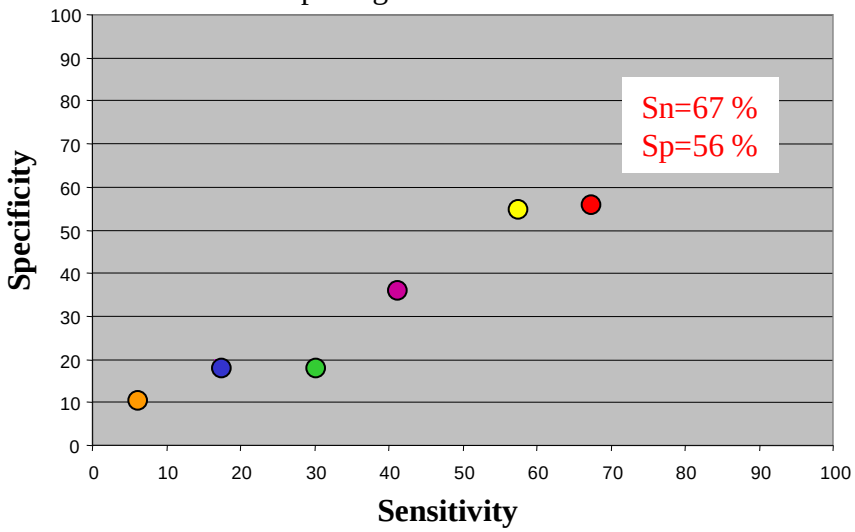
flagdb@evry.inra.fr
aubourg@evry.inra.fr
http://urgv.evry.inra.fr/FLAGdb

Why EuGène ?

An eukaryotic gene finder that combines multiple sources of evidence.
T. Schiex *et al*. (BIA, INRA Toulouse)

- FgenesP
- GenScan
- GlimmerA
- GeneMark.hmm
- FgeneSH
- EuGène

http://eugene.toulouse.inra.fr/

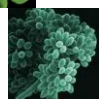


Scale : gene
Mode : *ab initio* pure

S. Aubourg *et al*. (URGV, Evry)

S. Rombauts *et al*. (PSB, Gent)

J. Gouzy *et al*. (LIPM, Toulouse)

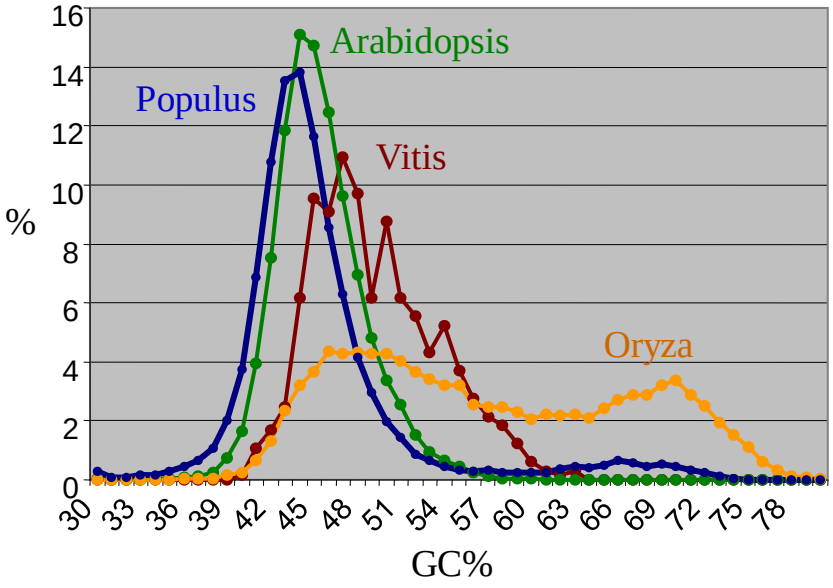


J. Amselem *et al*. (URGI, Versailles)

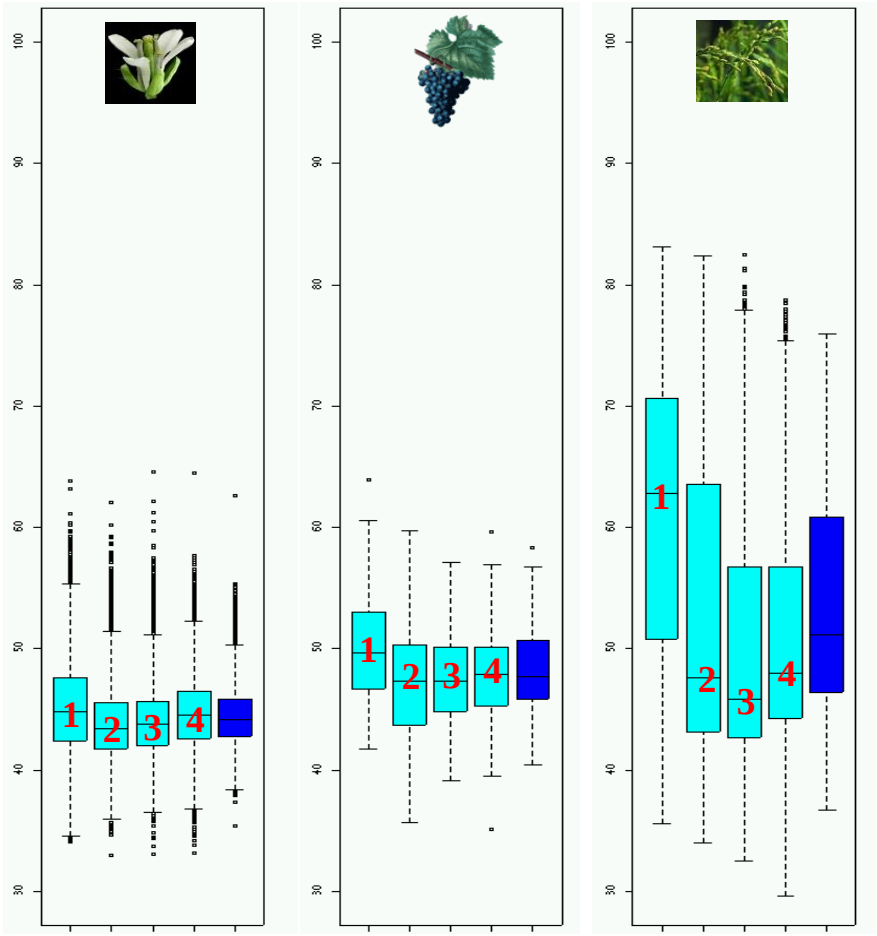
...

Gene training set

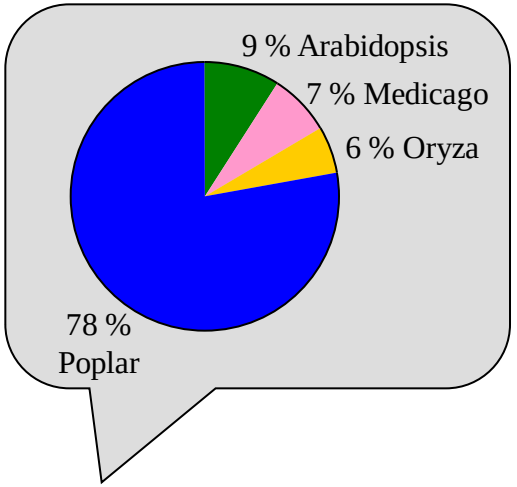
Set of 600 experimentally characterized complete genes (full-length mRNA cognate to genomic DNA) 'representative' of the whole genome + 4500 splicing sites



GC content



Integration and optimization steps



Tested formulas (plug-ins):

-1- IMM + SpliceMachine

-2- IMM + SpliceMachine + BLASTX

-3- IMM + SpliceMachine + BLASTX + TBLASTX

-4- IMM + SpliceMachine + BLASTX + TBLASTX + GenomeThreader

-5- IMM + SpliceMachine + BLASTX + GenomeThreader

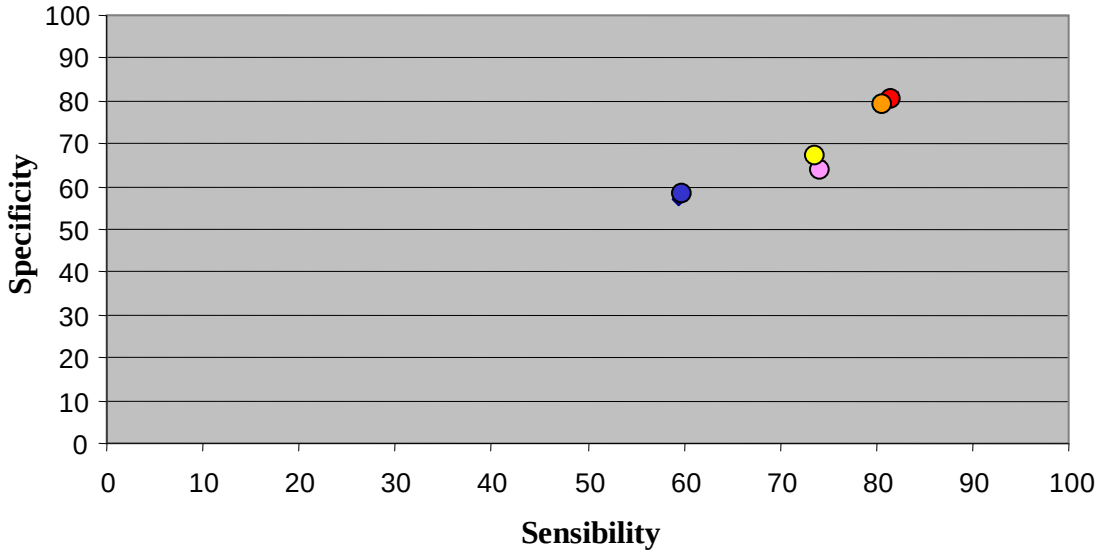
▼ ▼ ▼

- SwissProt proteomes from :
- Arabidopsis
- Oryza
- Poplar
- Medicago

Poplar genome

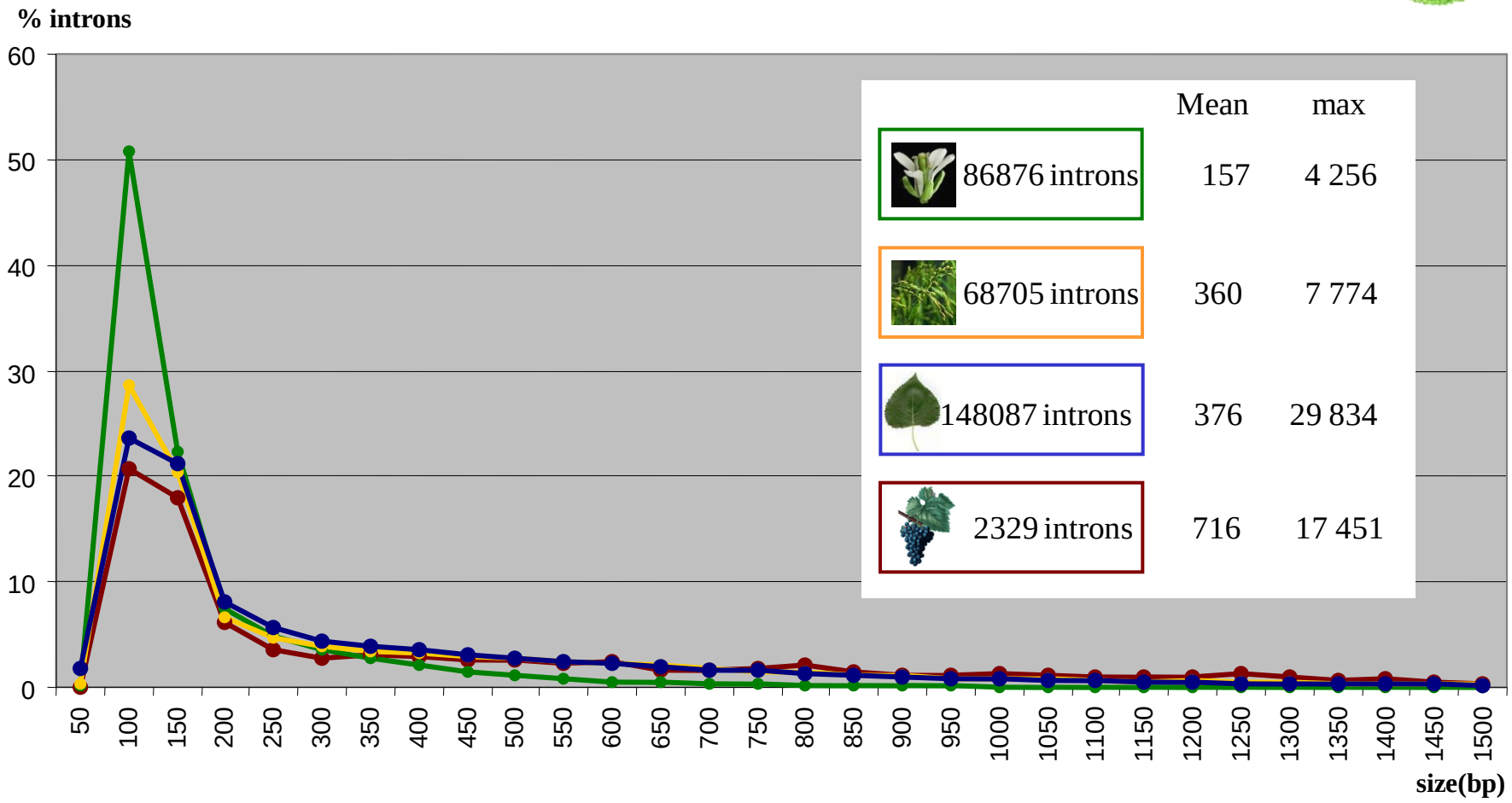
EST cDNA

EUGENE training results



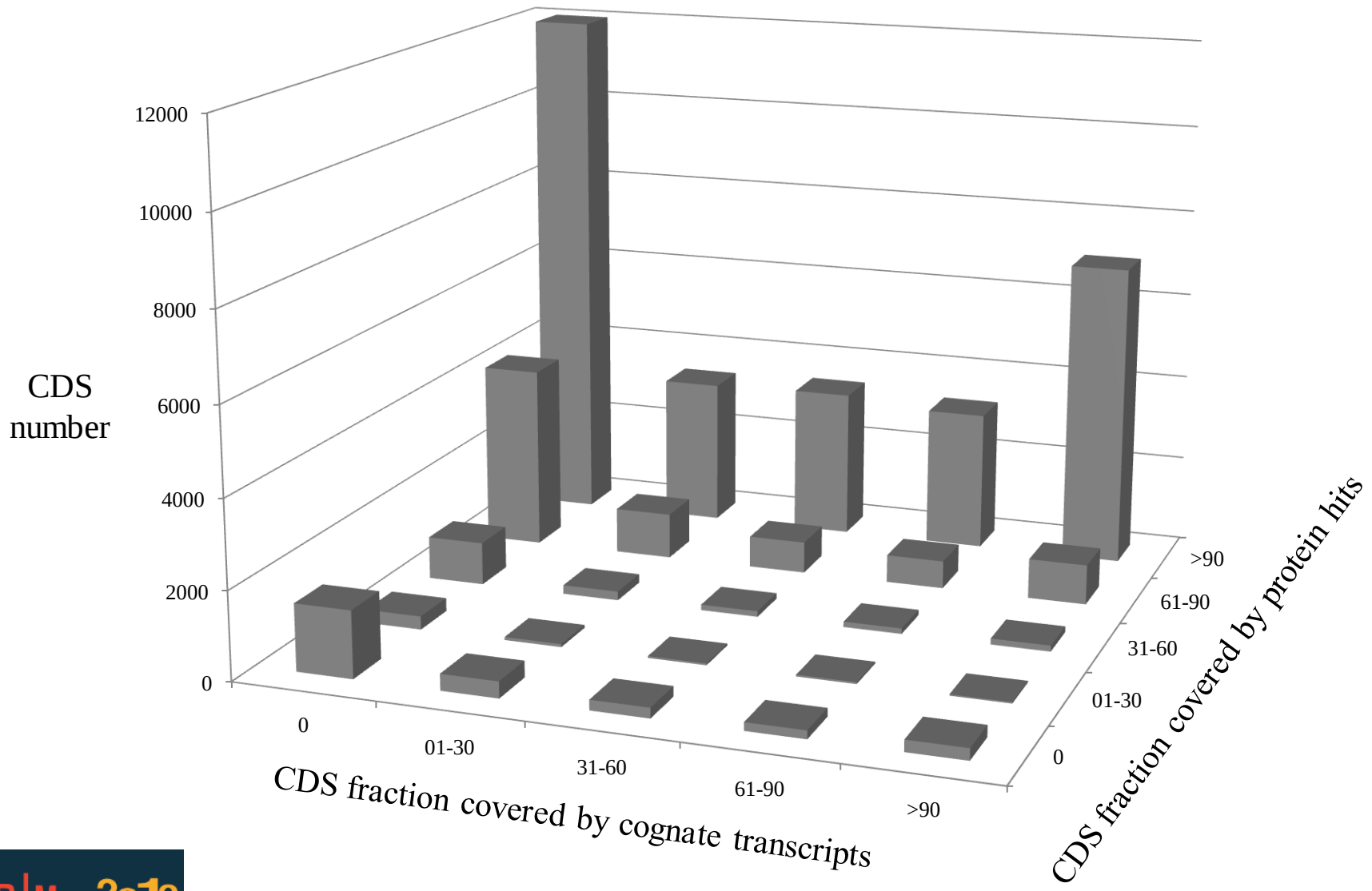
- IMM+SM
- IMM+SM+BLASTX
- IMM+SM+BLASTX+TBLASTX
- IMM+SM+BLASTX+GT
- IMM+SM+BLASTX+TBLASTX+GT

Intron size (training set)

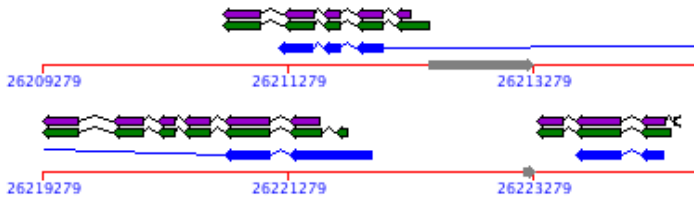


Long introns (12% include LINES elements)

EuGène results



EuGène vs GAZE (TPS family)

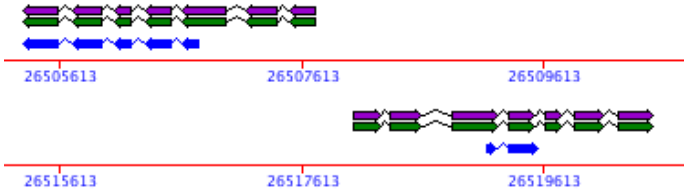


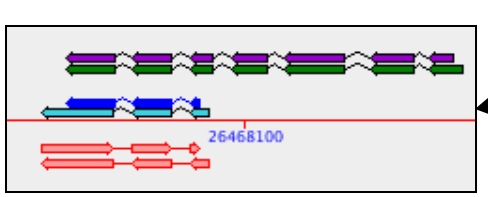
Out of 152 TPS loci :

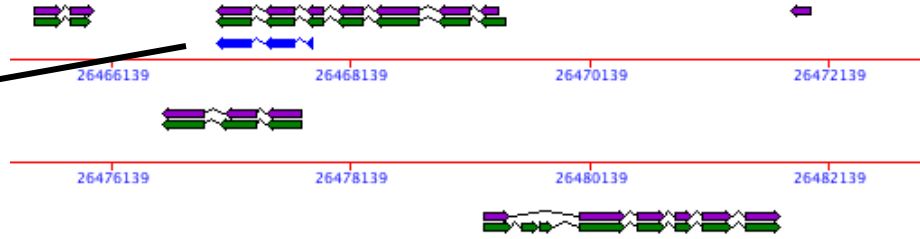
- GAZE fails to detect 54 of them (including 6 full TPS genes)
- EUGENE fails to detect 12 of them

Out of 53 complete and perfect TPS genes:
(partial and disturbed genes can not be well predicted)

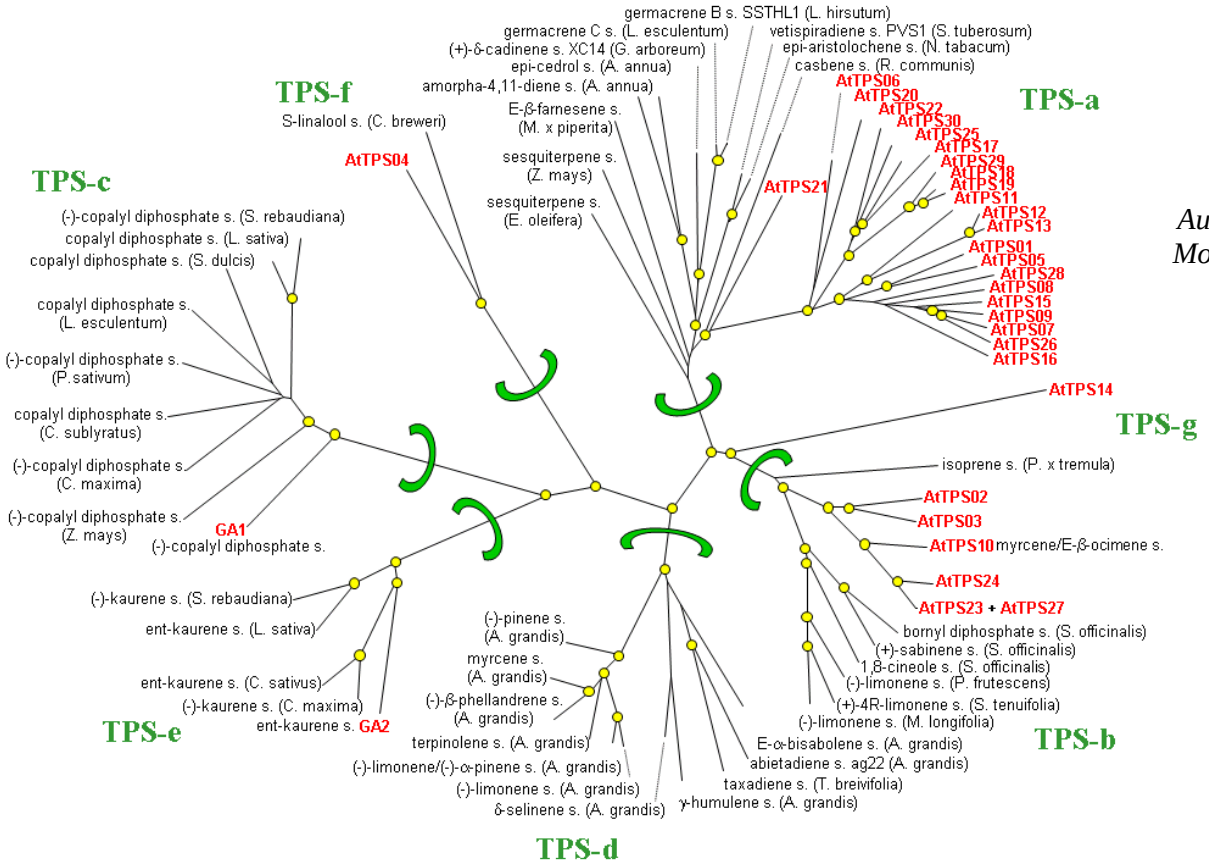
- GAZE predicts perfectly 12 TPS 23%
- EUGENE predicts perfectly 43 TPS 81%







TPS classification



Aubourg et al. (2002)
Mol. Genet. Genomics

TPS-a	Sesquiterpene and Diterpene Synthases
TPS-b	Monoterpene Synthases
TPS-c	Diterpene Synthases (copalyl diphosphate synthases)
TPS-d	TPS from conifers
TPS-e	Diterpene Synthases (ent-kaurene synthases)
TPS-f	Monoterpene Synthases (linalool synthases)
TPS-g	unknown TPS

Functional characterization of TPS genes

Gewürztraminer
Pinot Noir
Cabernet Sauvignon

Stem, leaf, berry, root and flower tissues ◀ Methyl Jasmonate

mRNA extraction

PCR with primers selected from annotated genomic sequences

cDNA cloning of vitis TPS

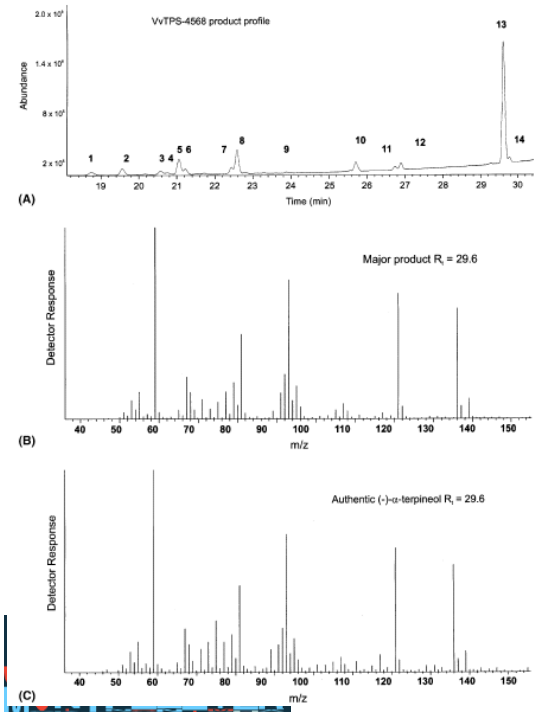
Expression of recombinant TPS in *E. Coli*

semi-purification

***In vitro* enzyme assay** ◀ Buffer and substrate selection is driven by *in silico* functional predictions

GC-MS

Identification of terpenoid products



Dr Diane Martin (UBC)


newSTS7
Vst1
Copia-like
GSVIVG00010585001
Vst2
16334693
16336693
16338693
16340693
16342693
16344692
GSVIVG00010584001
(Vst3)
16344693
16346693
16348693
16350693
16352693
16354692
16354693
16356693
16358693
16360693
16362693
16364692
16364693
16366693
16368693
16370693
16372693
16374692
16374693
16376693
16378693
16380693
16382693
16384692
16384693
16386693
16388693
16390693
16392693
16394692
newSTS8
Vst1
Copia-like
16394693
16396693
16398693
16400693
16402693
16404692
GSVIVG00010580001
Vst2
GSVIVG00010579001
(Vst3)
Unsequenced region...
16404693
16406693
16408693
16410693
16412693
16414692

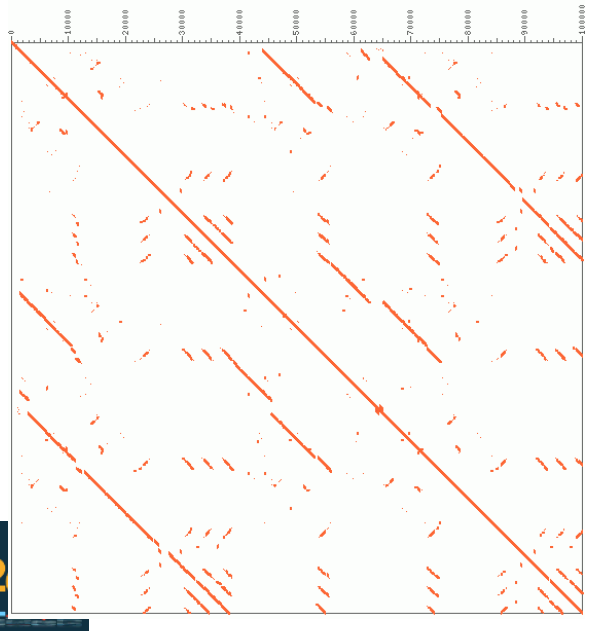
Stilbene and Chalcone Synthase families

Proteome and translated genome have been screened with sequences of 3 previously characterized Stilbene Synthases :

- **62 loci identified and expertised :**

  - 3 CHS genes highly expressed and 11 CHS-like

  - 48 STS genes (discriminated with specific residues) distributing on 2 clusters (chromosomes 10 and 16). At least 73% of them are expressed. Only 2 grape STS have been experimentally characterized. All the STS share between 91% and 99.7% of identity.



Out of 62 STS/CHS genes (2 exons):

- GAZE fails to detect 36 of them and only 2 are well predicted !

- EUGENE fails to detect only one and 40 are well predicted.