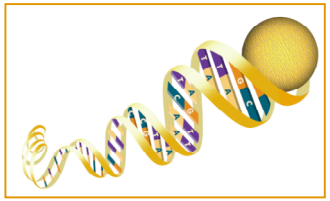
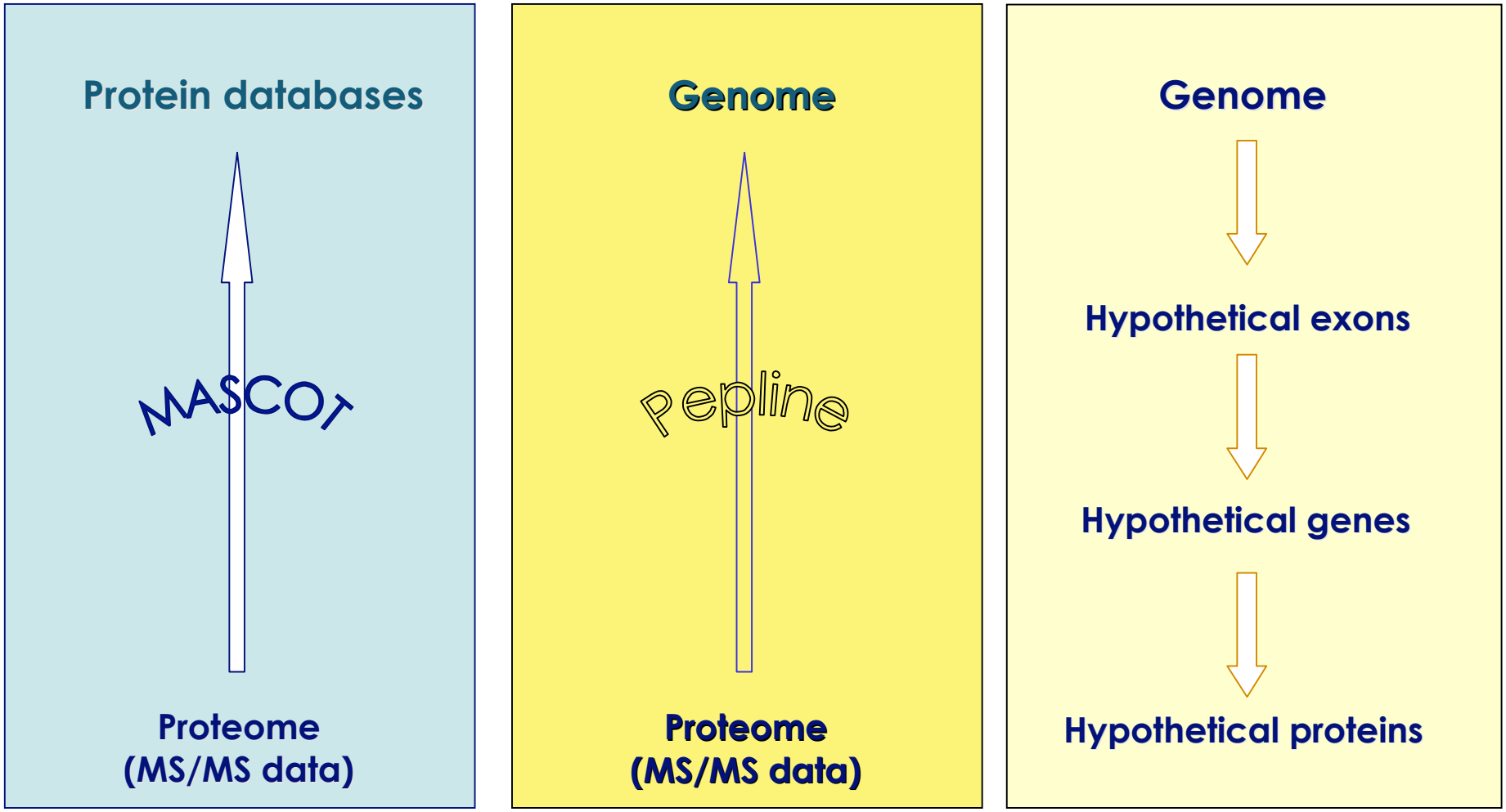
Development of New Proteomic Tools for Functional Genomics and Genome Annotation

PepLine, a New Software Pipeline using LC-MS/MS Data for Genome Annotation

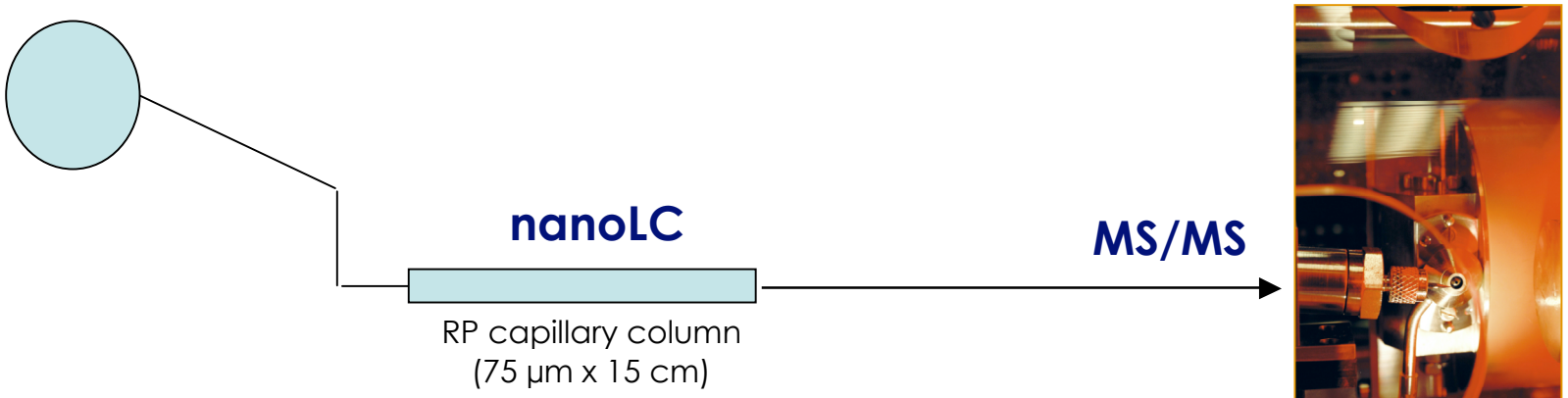


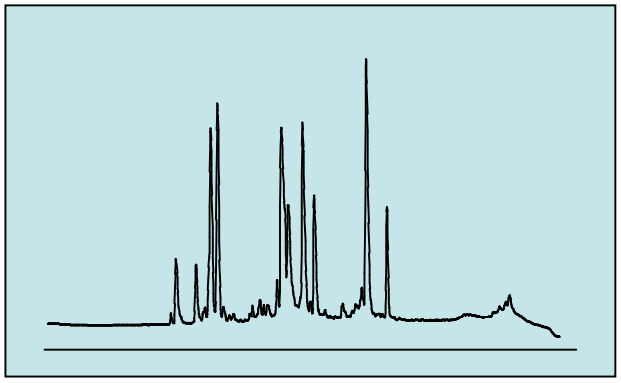
Jérôme GARIN

Genome Annotation

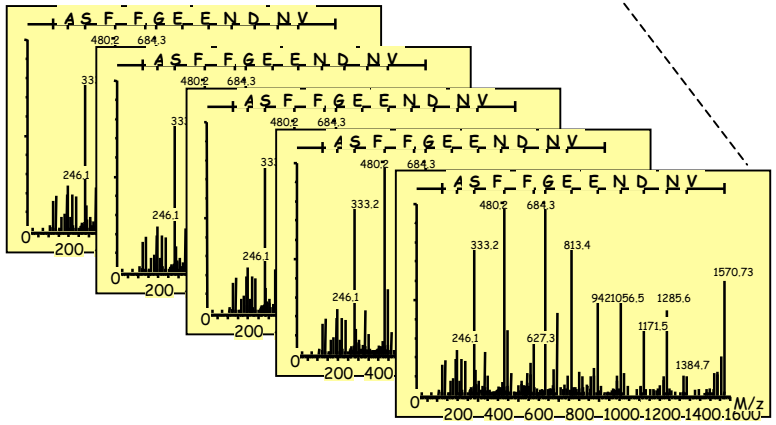


«Shot gun» Proteomics using nanoLC-MS/MS

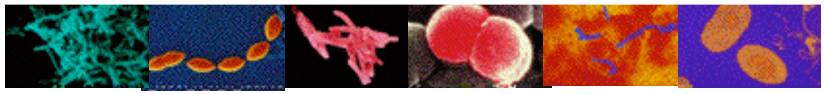




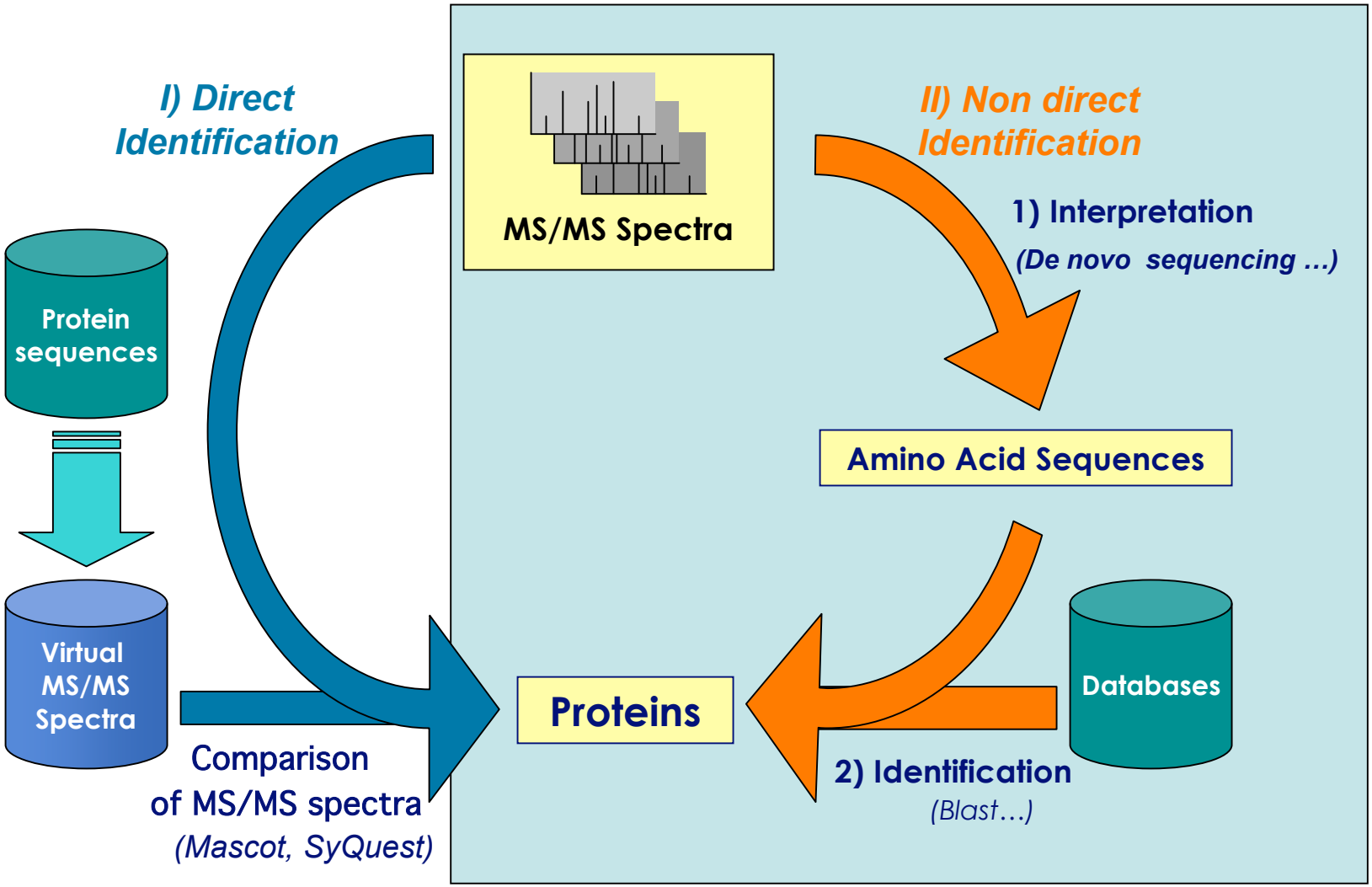
Peptide separation + concentration (UV profile)



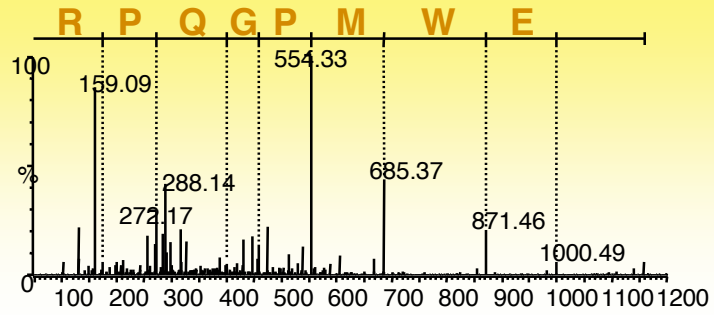
MS/MS spectra



Two Main Approaches for Protein Identification

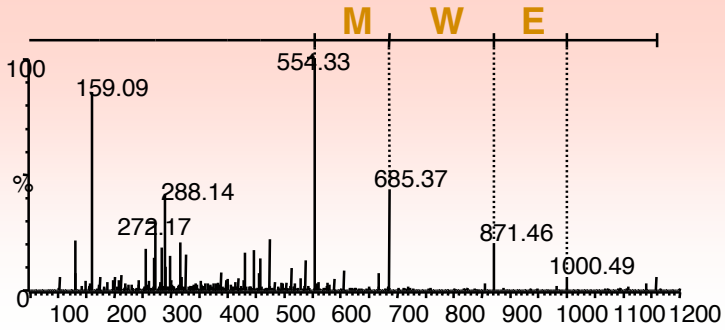


Throughput



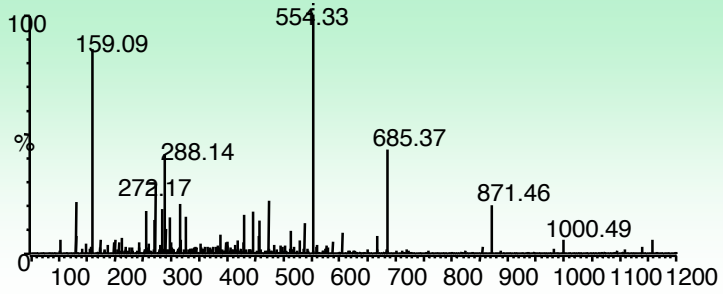
Amino Acid sequence

EWMPGQPR



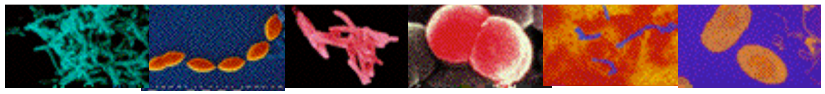
Peptide Sequence Tag = PST

(158.1)EWM(535.4)

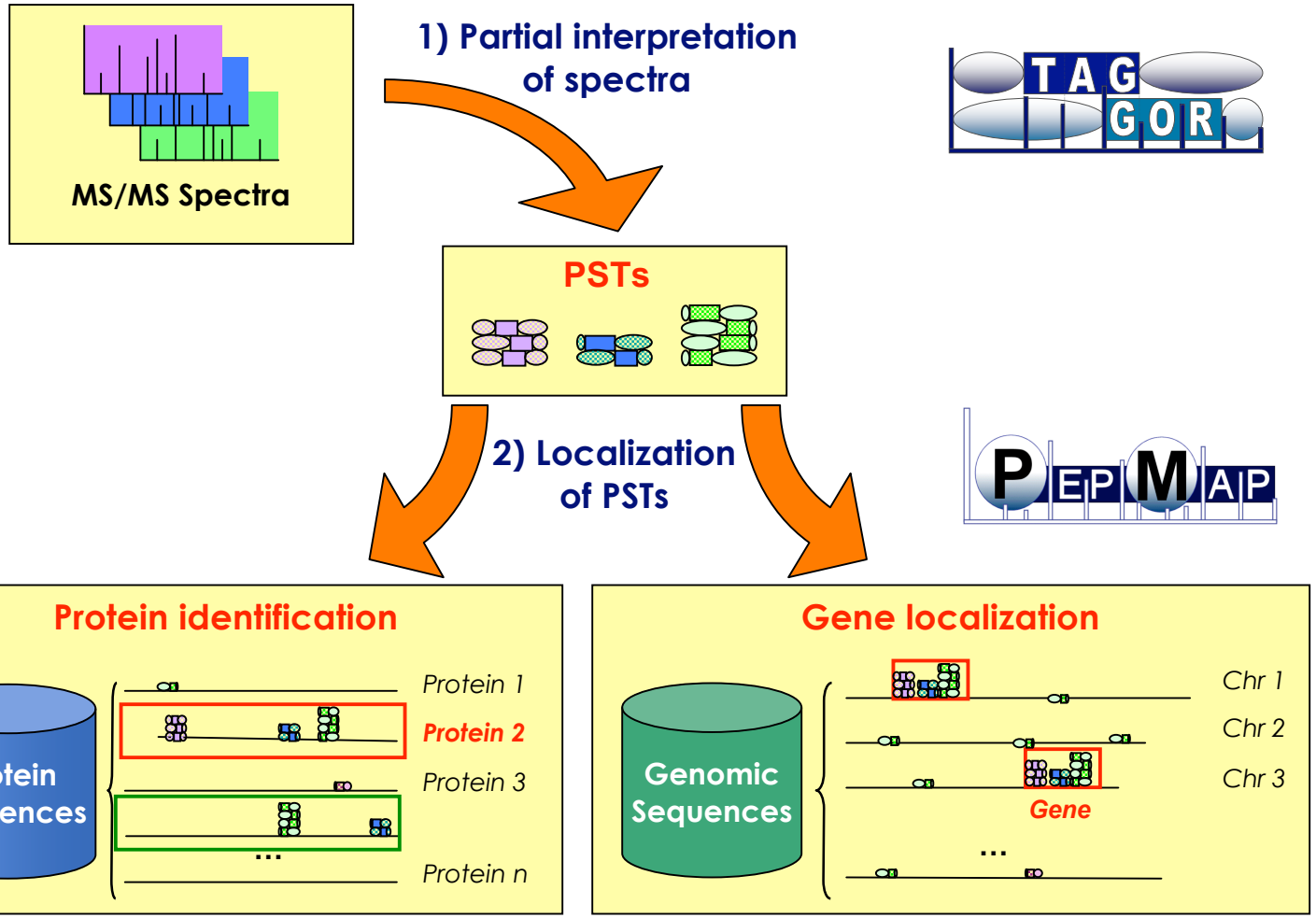


Peak List

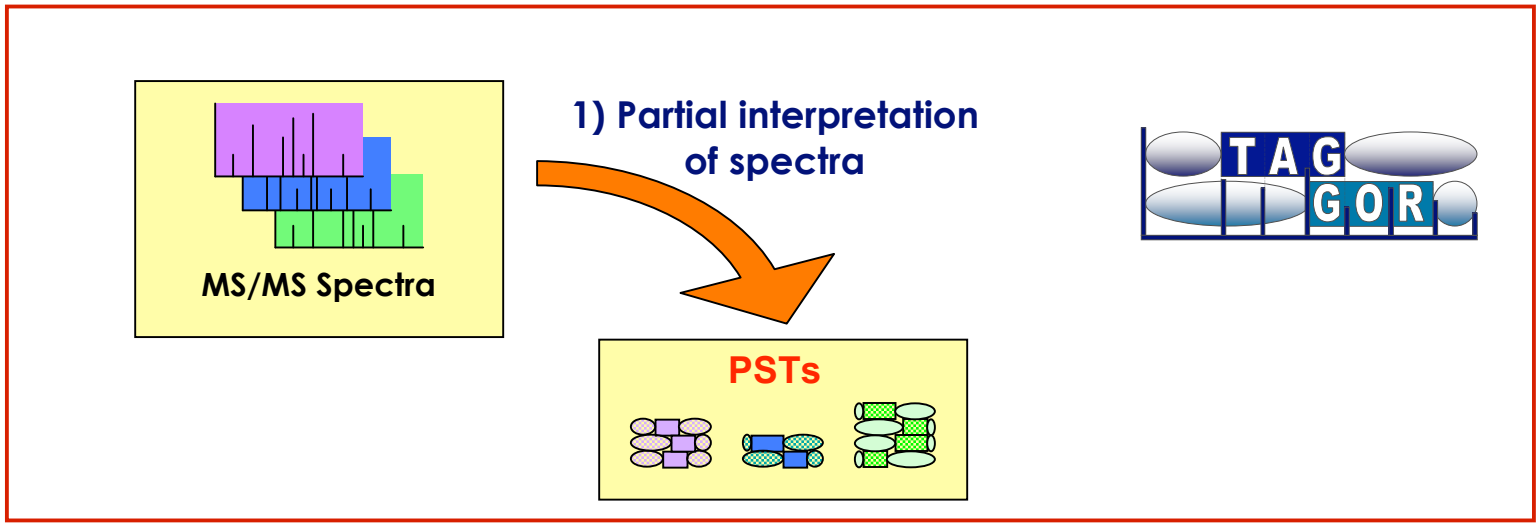
Information



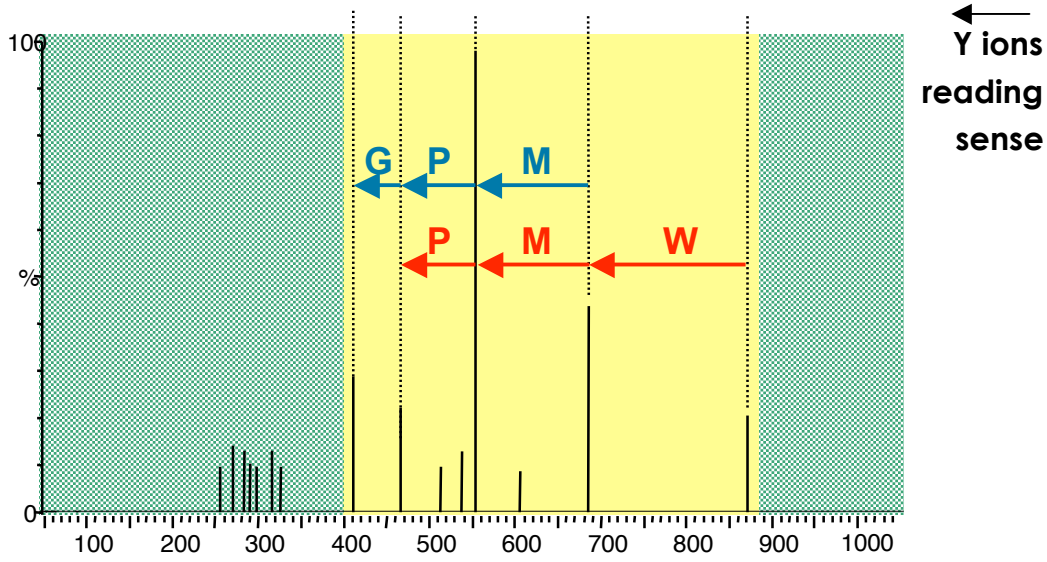
«Pepline» : the Taggor-PepMap Pipeline



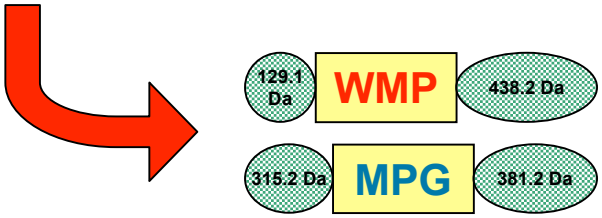
Taggor : from MS/MS Spectra to PSTs



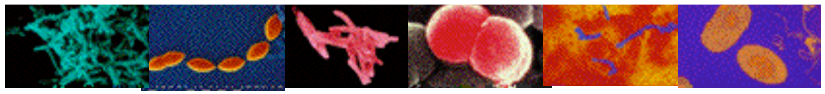
Taggor : from MS/MS Spectra to PSTs



«Brute Force» Approach



- Generates all possible amino acid triplets sequences.
- For each triplet sequence matching the MS/MS spectrum, a PST is generated.
- Overlaping PSTs can be generated
- Each generated PST is scored using peak intensities.
- Only PSTs corresponding to the 10 best scores are taken into account

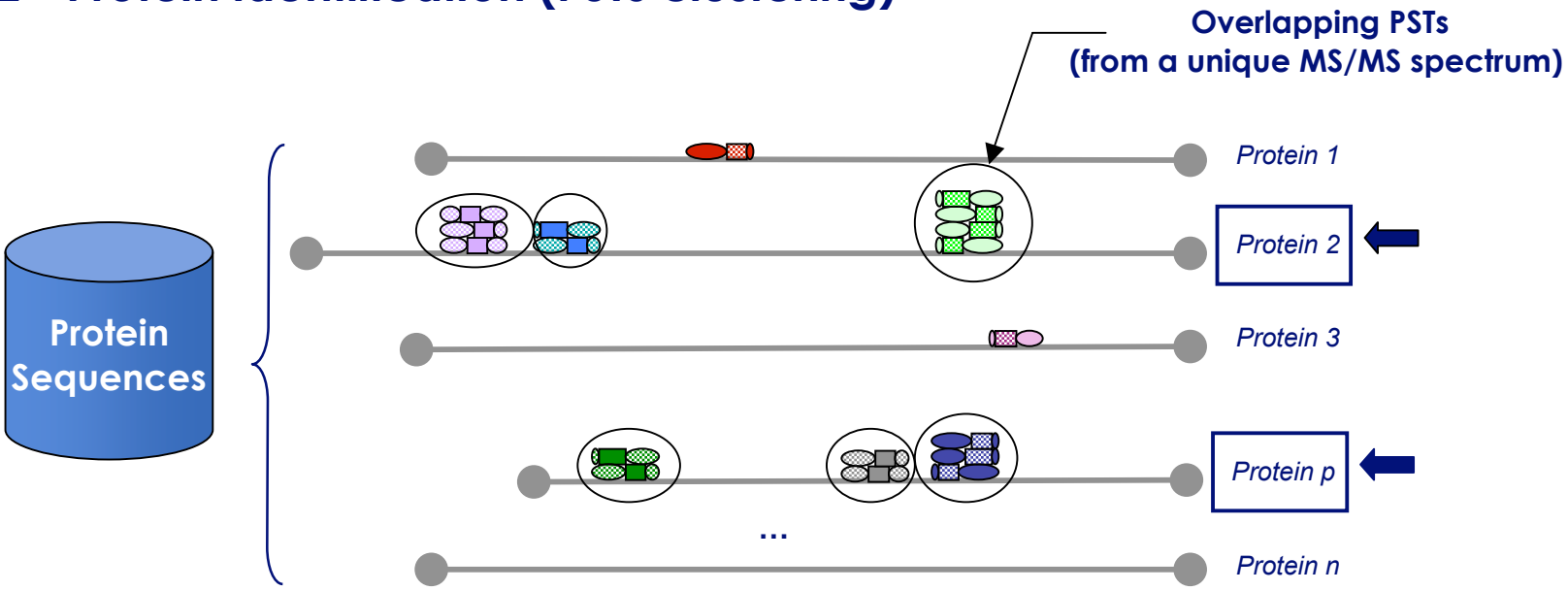


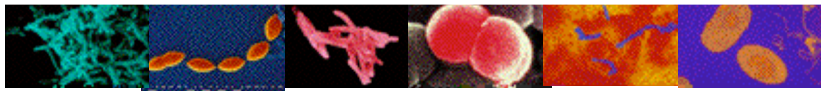
PepMap^{Pro} : from PSTs to protein identification

1 - Mapping PSTs on protein sequences

PepMap^{Pro} : from PSTs to protein identification

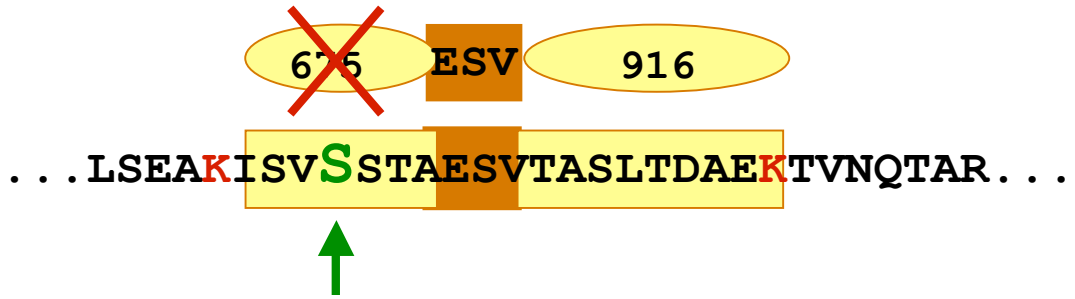
2 - Protein identification (PSTs clustering)



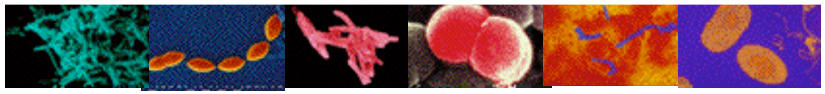


PepMap^{Pro} : from PSTs to protein identification

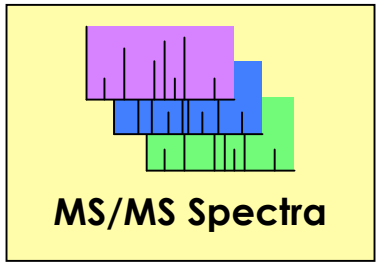
Partial matches on protein sequences



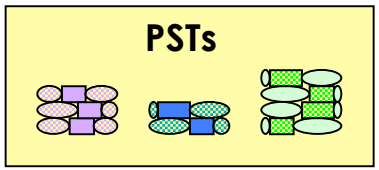
Sequence isoforms
Post translational modifications
…



Pepline^{Gene} : a Proteomic Tool for Genome annotation

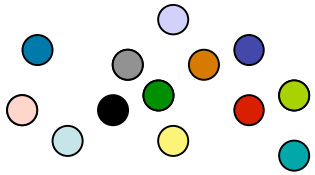


1) Partial interpretation of spectra



PepMap^{Gene} : Clustering PSTs on translated raw genomic data

PSTs



Translated chromosme sequence

+1

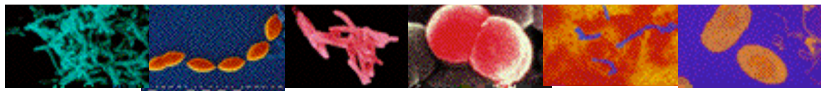
+2

+3

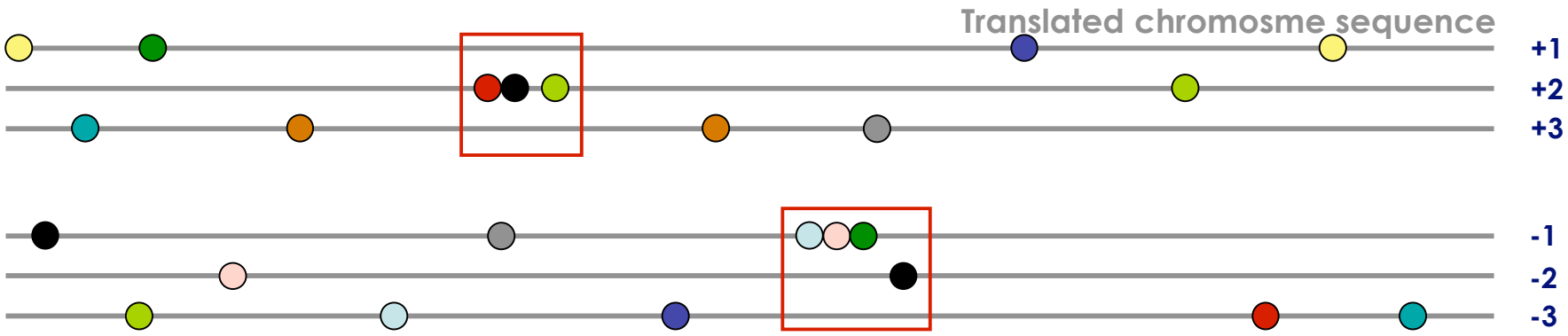
-1

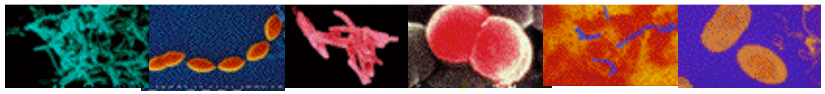
-2

-3



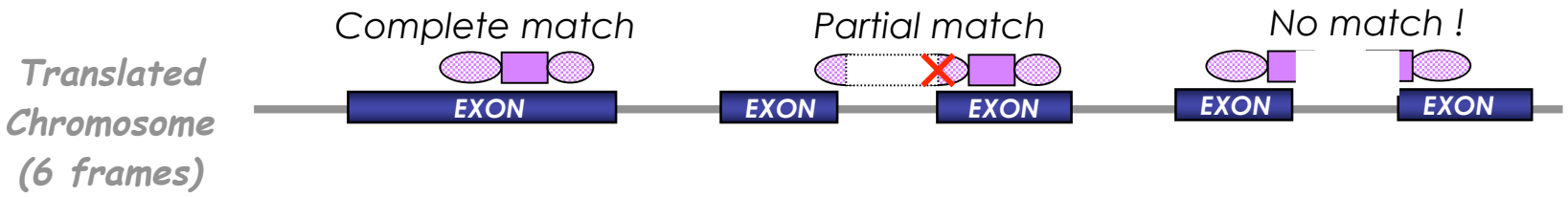
PepMap^{Gene} : Clustering PSTs on translated raw genomic data



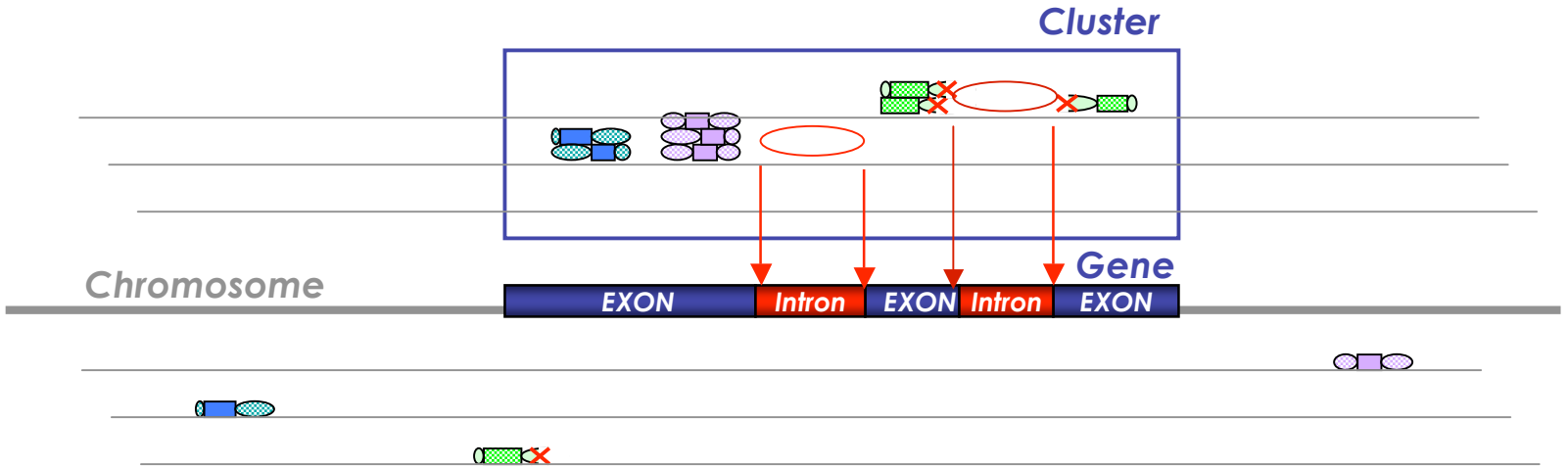


PepMap^{Gene} : from PSTs to Gene Localization

1- PSTs mapping on a translated eukaryotic genomic sequence



2 – Coding Sequence localization (PSTs *clustering*)



Pepline^{Gene} : from MS/MS data to Gene Localization

LCU0134 (nanoLC-MS/MS analysis)

pepline

11300 PSTs

Arabidopsis thaliana

Chromosome 1

Frame +1

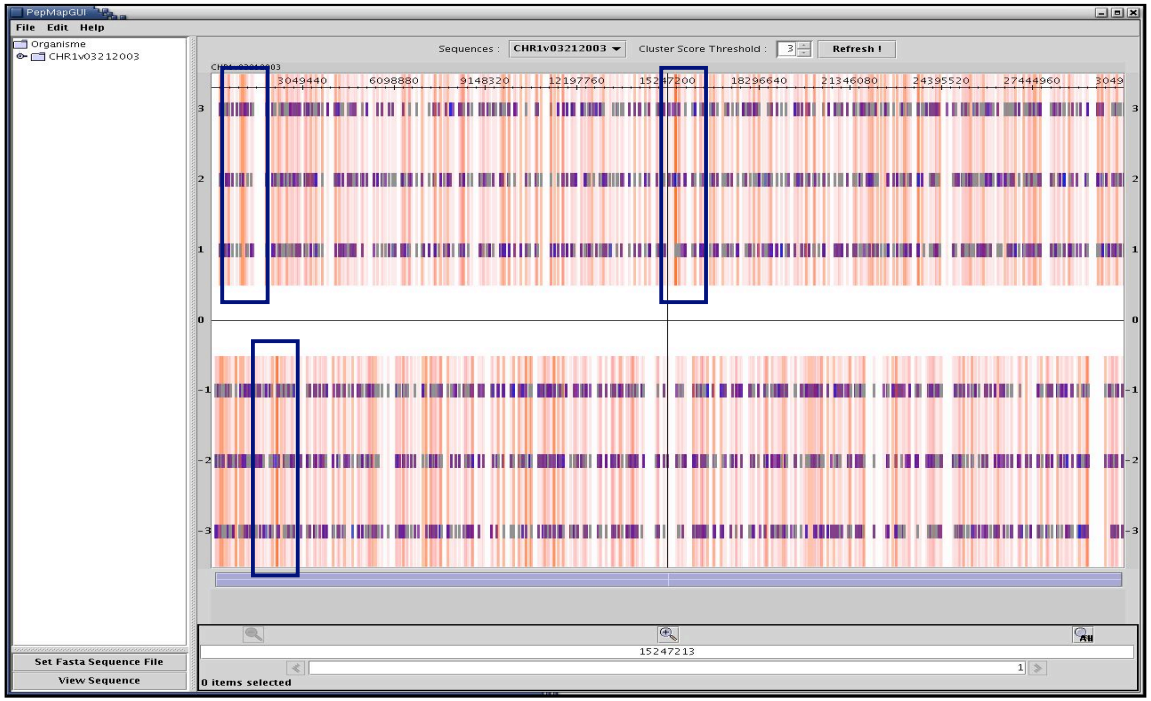
Frame +2

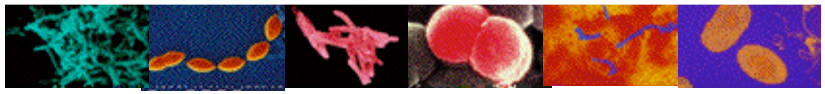
Frame +3

Frame -1

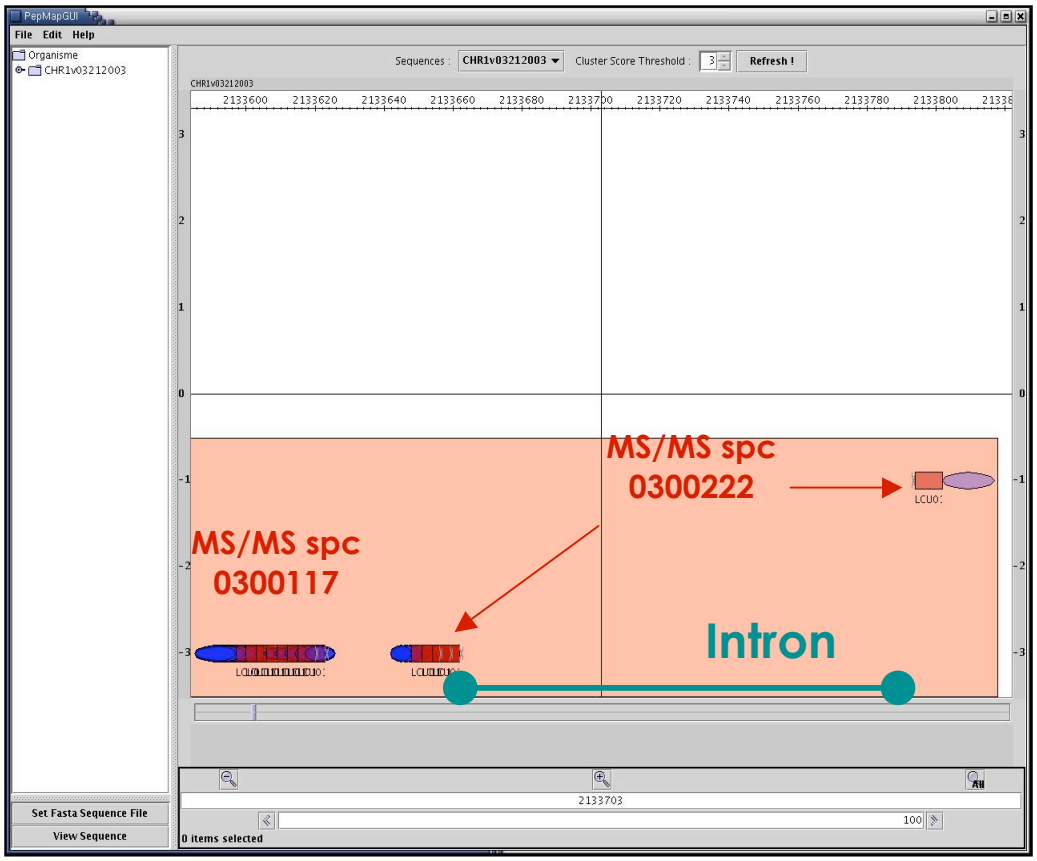
Frame -2

Frame -3





Pepline^{Gene} : from MS/MS data to Gene Localization



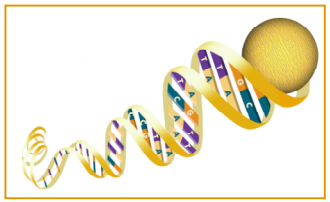
Nucleotide sequence (nt 2133100 to nt 21328001 - reverse strand)

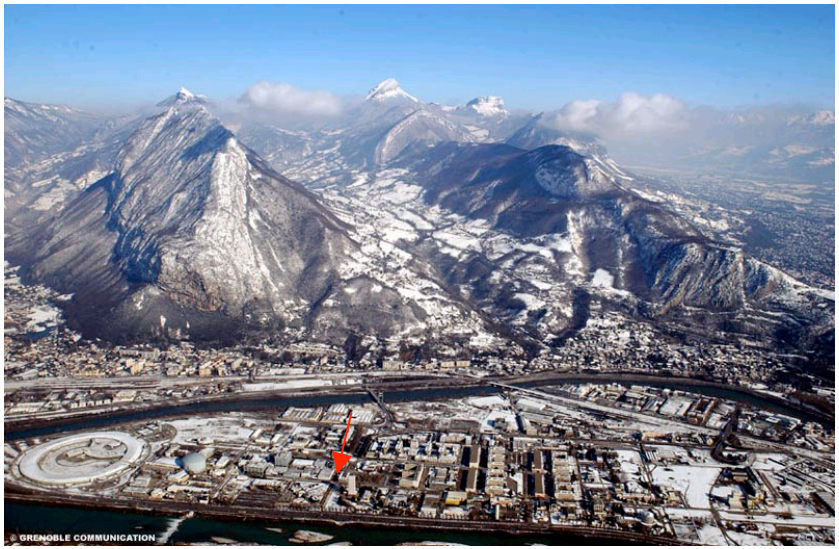
BLAST

Corresponds to protein At1g06950 : chloroplast inner envelope protein

Conclusion

- Pepline^{Pro} seems to be a promising tool for protein database mining. A version will be available soon for beta testing.
- Pepline^{Gene} allows to mine raw genomic sequences (but there are still many paramaters to tune : number of PSTs/cluster, cluster size, full match/partial match scoring …).
- The integration of Pepline^{Gene} into dedicated Bioinformatic platforms might be a potent approach for the annotation of prokaryotic and eukaryotic genomes.





Romain CAHUZAC
Myriam FERRO
Christophe BRULEY
Gilles ANDRE
Emmanuelle MOUTON
Michel JAQUINOD
Marianne TARDIF
Jérôme GARIN

Erwan RÉGUER
Estelle NUGUES
François REICHENMAN
Alain VIARI

Thierry VERMAT
Marielle VIGOUROUX
Yves VANDENBROUCK