Au delà du miroir

Biologie Moléculaire et Algorithmique Combinatoire

Projet HELIX, Inria Rhône-Alpes et UCBL, Lyon

(le reste du texte, mais pas de l'exposé, sera en anglais)

“In another moment down went Alice after it, never once considering how in the world she was to get out again”

The persons from the lac[‡], those who went to the other side of the mirror[◇], and those who came to this side[⋆]

Inria - Lyon (HELIX)	Various - Paris (Jan. 2002)
Nadia Pisanti	Julien Allali (IGM)
Sébastien Provencher	Philippe Blayo (IGM)
Marie-France Sagot	Anne Cariou (IGM)
Marina Zelwer	Frédéric Chataigner (LIAFA)
USP - Brazil	Sophie Créno (IGM)
Estela M. Rodrigues	Laurent Marsan (IGM)
	Raquel Tavares (Pasteur)

[◇]computer scientist [⋆]biologist

[‡]Laboratoire d'Algorithmique Combinatoire, Inst. Pasteur

What will this talk be about ?

a mostly non technical survey of past and current work and of some of its perspectives

the point of view adopted will reflect that of an algorithmicist

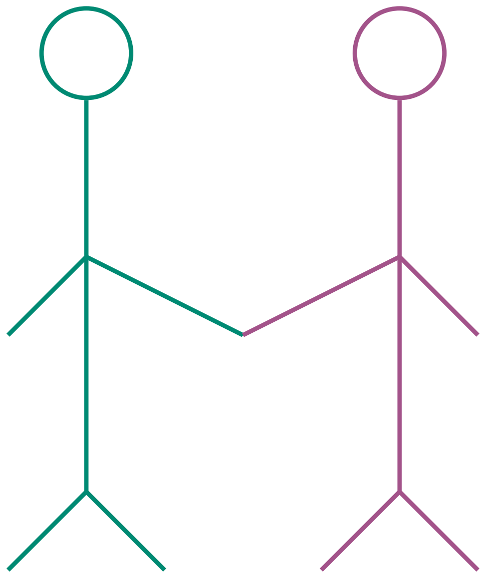
remember though that

At the beginning

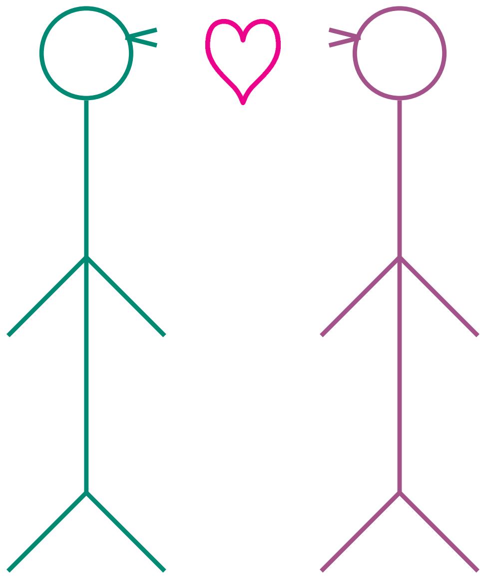
there is *always* a biological question

that we try to answer by understanding the deep structure of both the algorithmical/mathematical problems raised and the biological objects upon which they bear

Sometimes algorithmics and biology advance hand in hand

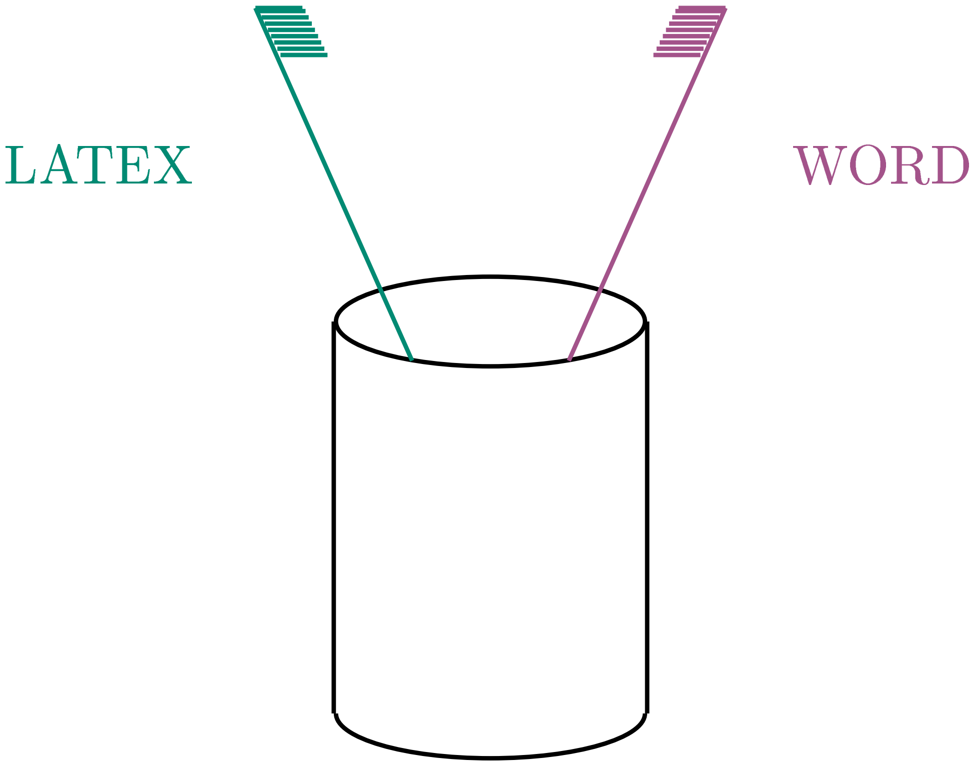


Sometimes in parallel



Each keeping its own personality

(or, later in life...)



“Understand the deep structure of problems and objects”

- **infer *unknown* biological objects or processes with the help of algorithms**
- **taking into account the specific constraints and characteristics of biology**
  - **variability and “acceptable” differences**
    - **not to be confused with presence of imprecision or errors in the data (which must also be considered)**
  - **combinatorics which makes diversity of life possible**

First (VERY broad) area of (biological) research:

REGULATION

OF

GENE EXPRESSION

more precisely, initially, we'd like to be able to answer the question:

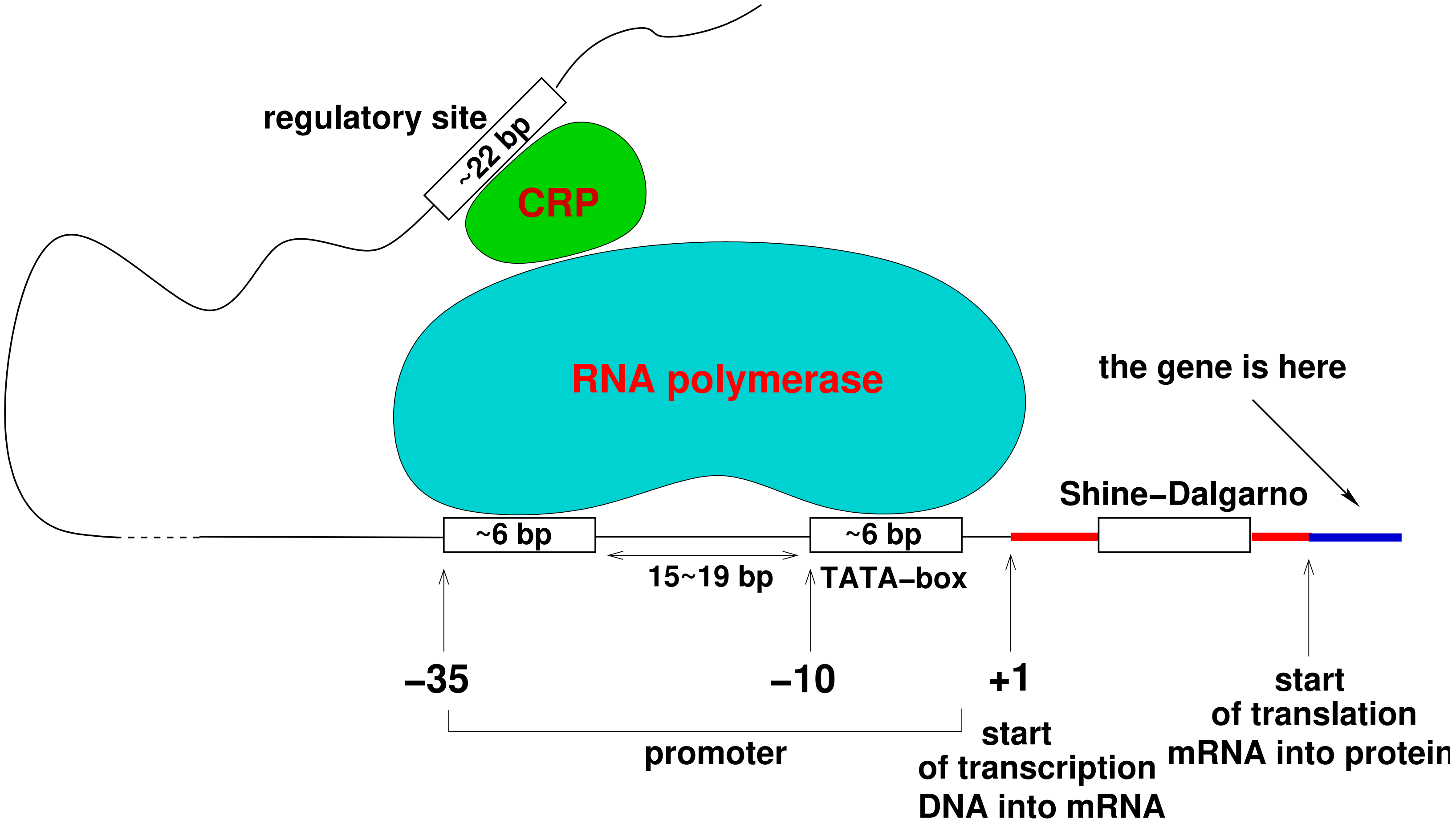
given a genome (or various genomes)

what are ALL the elements

involved in the regulation of the expression of its genes ?

in other words, can we "model" recognition and capture the combinatorics underlying regulation ?

A very simplified (and partial) view on the regulation of gene expression (in the “ simpler” case of a prokaryote)



We start slowly

by using as information

the sequence (string) only

that is, we address the problem of (sequence) motif inference

Motif search × Motif inference

Search

a known motif

a text

$\Longrightarrow$

positions in the text
where the motif
is "found"

Motif search $\times$ Motif inference

Search

a known motif a text	$\Longrightarrow$	positions in the text where the motif is "found"

Inference

a set of properties a text	$\Longrightarrow$	(all) motifs satisfying the properties

“*Not* **quite** *right, I'm afraid*, **said Alice timidly,** *some of the words have got altered.*”

set of properties

- **main property: motifs of interest = “conserved” element**
  - **remember, in biology, almost nothing is ever perfectly conserved**
    - **mutations (substitutions, insertions, deletions) may happen that do not destroy function and may even enable to modulate it**
  - **in some cases, one has also to consider man-generated “errors”**

Motif

"simple pattern"	"network expression"
TATAAT	TA[AT]N[AT]T
TTGNCA	
RFXCP	[ILMV][ASG]XXC[ILMV]H

where N or X is the wild card (don't care symbol)

$\Sigma_{DNA/RNA}$ = {A, C, G, T(U for RNA)} (nucleotides)
$\Sigma_{protein}$ = {A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y} (amino acids)

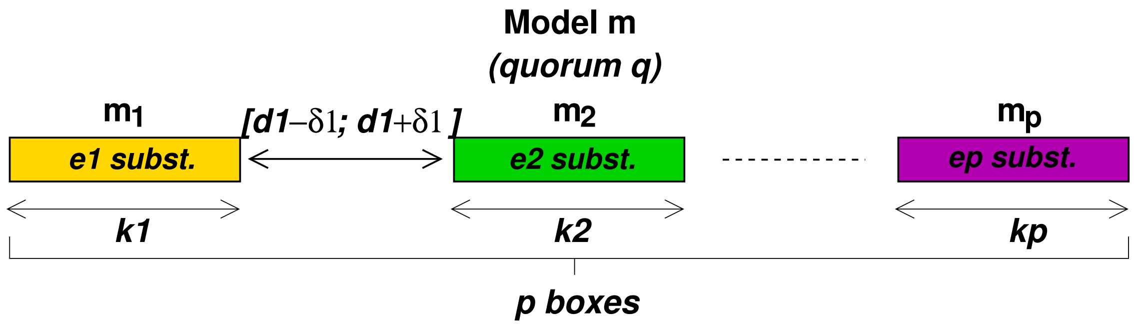
Motif + Properties = MODEL

Sought: ALL models, that is, all motifs satisfying the properties

Towards modelling the combinatorics of regulation

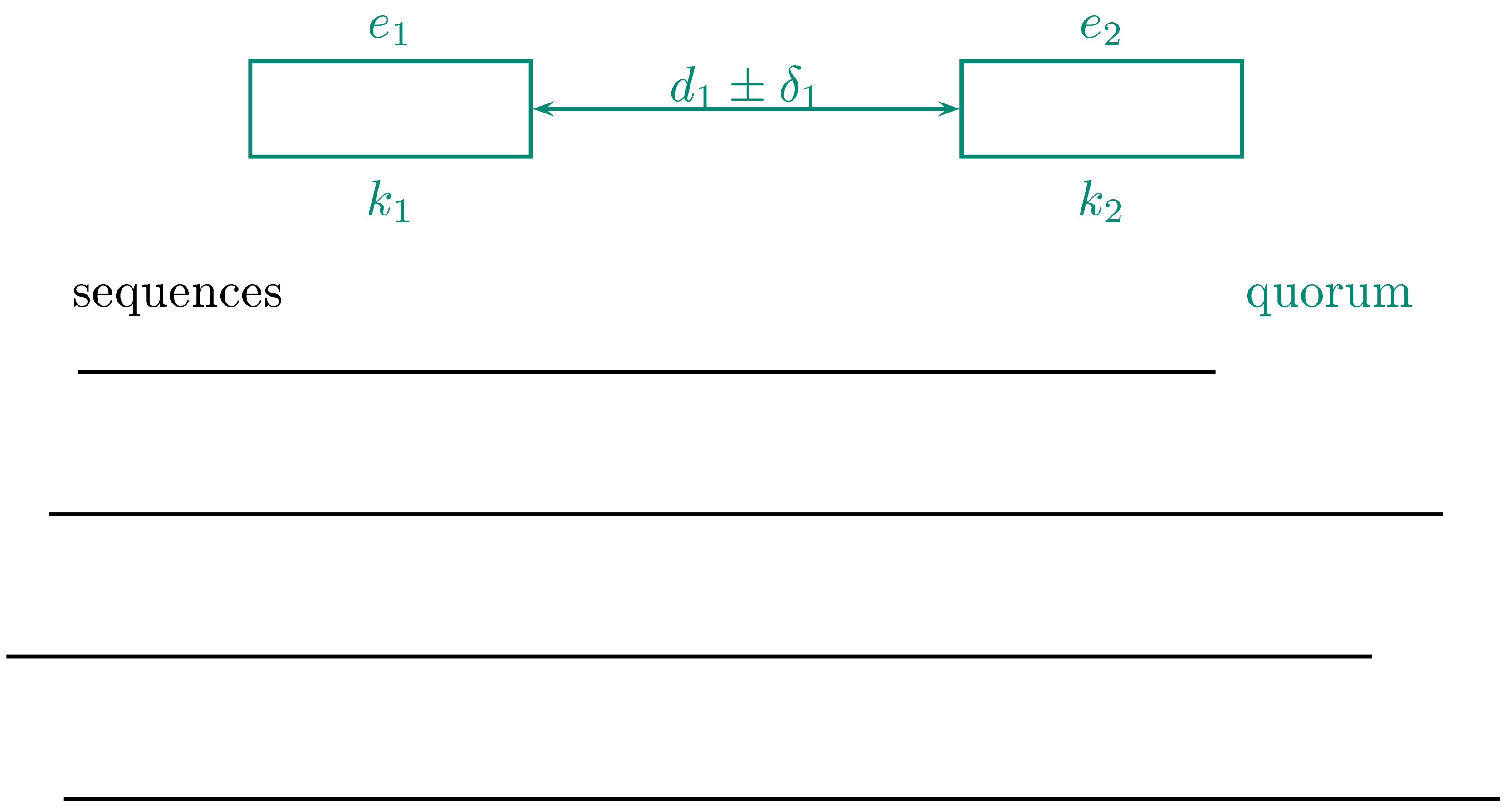
Structured models

an ordered collection of p boxes, p maximum numbers of differences (e), p intervals of lengths (k), $p-1$ intervals of distances (between successive boxes in the collection) ($d \pm \delta$) (this is called *spacers*)

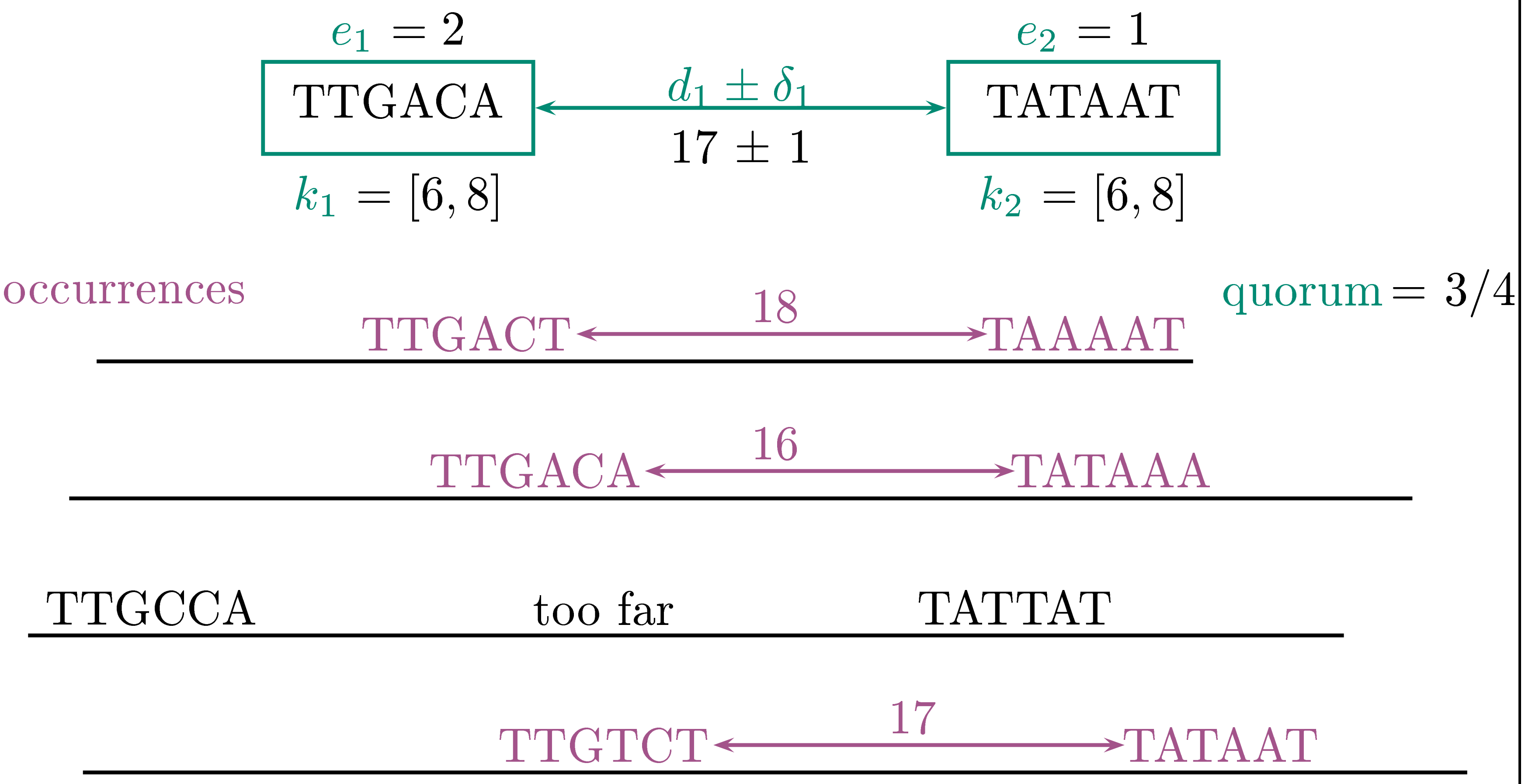


Smile (L. Marsan *et al.*, JCB, 7:345-362, 2000)

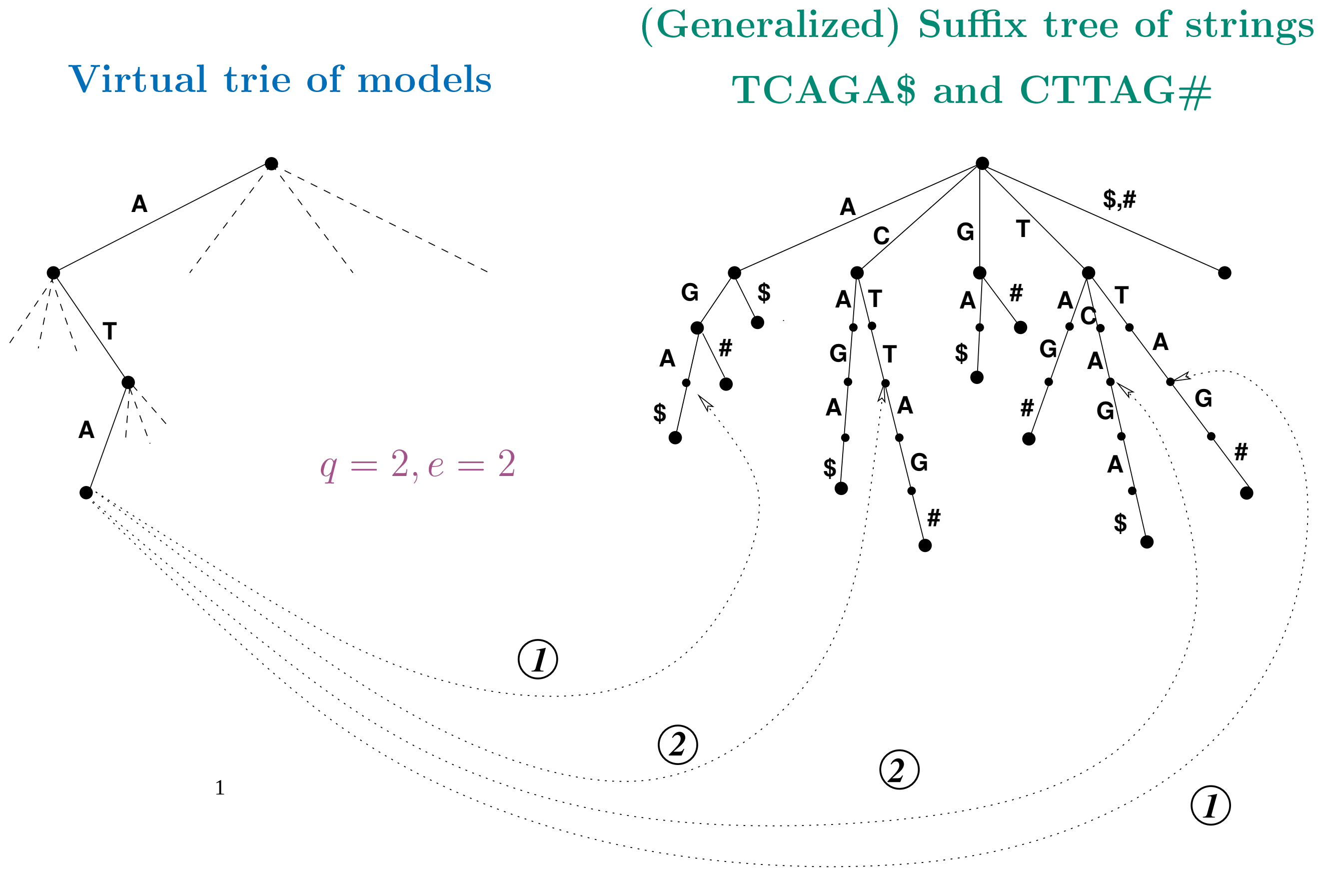
An example with two boxes



An example with two boxes

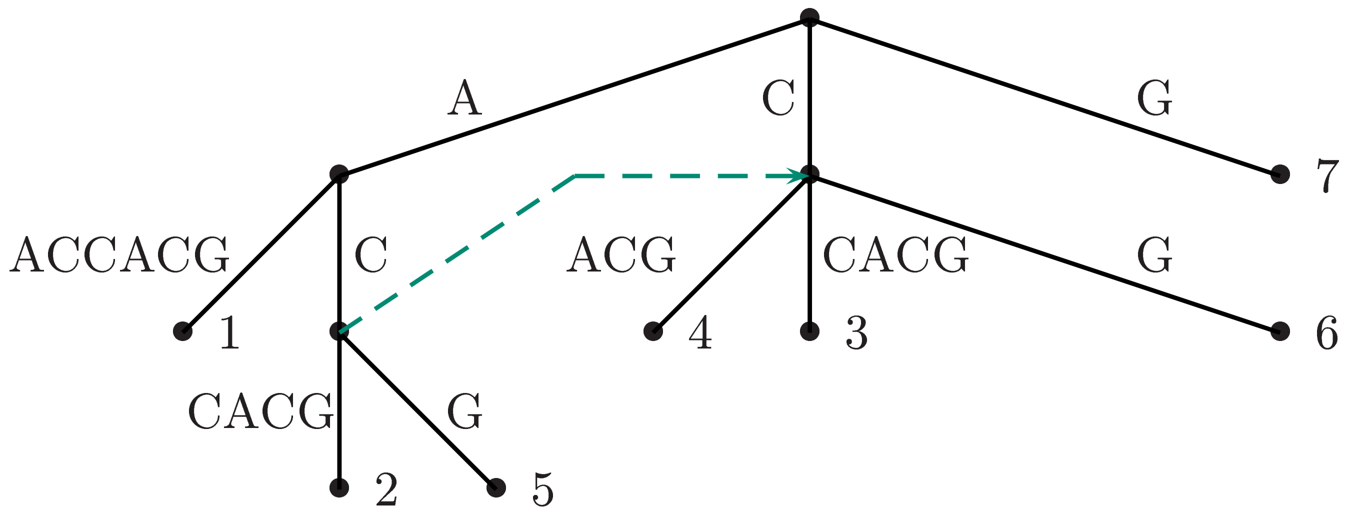


A whiff of the algorithm: A story of traversing two trees with a dynamic programming algorithm



Suffix tree: A reminder

Suffix tree for string AACCACG: a tree representing all suffixes of the string

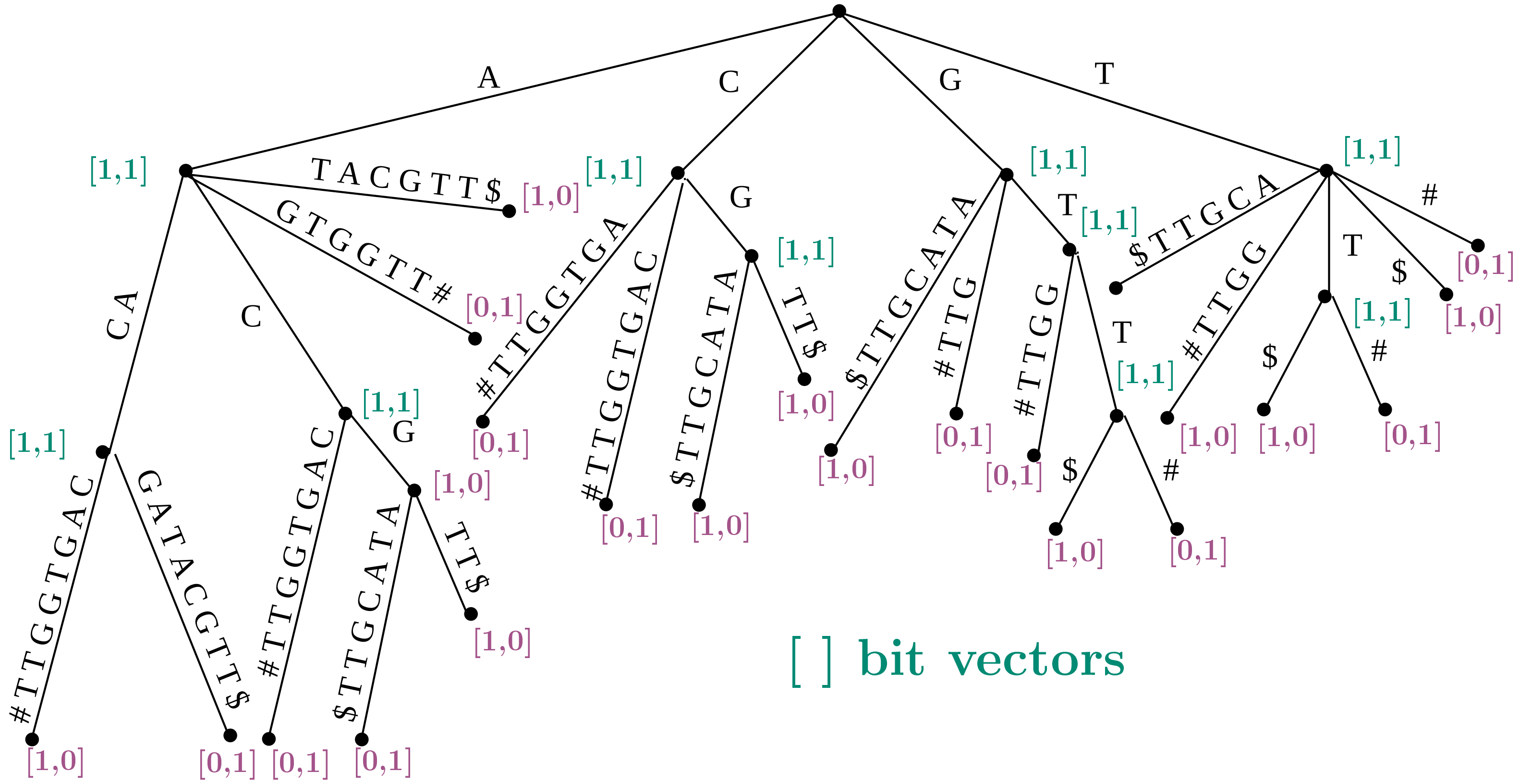


key feature: concatenation of labels of arcs from root to leaf i spells suffix of input string starting at position i

$\rightarrow$ **(suffix link) links node spelling αm to node spelling m**

Generalized suffix tree

Strings AACGATACCTT$ and AACCAGTGGTT#



[ ] bit vectors

Key ideas

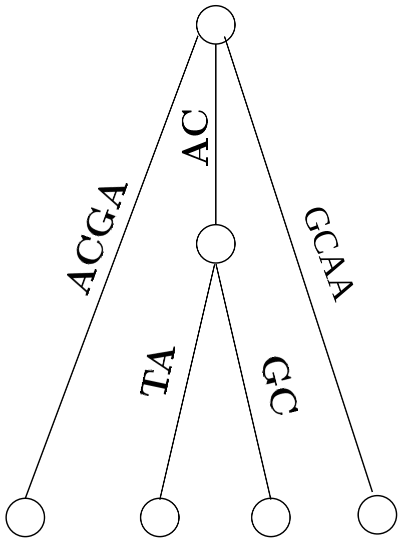
Use of an index (a generalized suffix tree) for representing the strings

in fact, what is used is a new data structure: a generalized factor tree

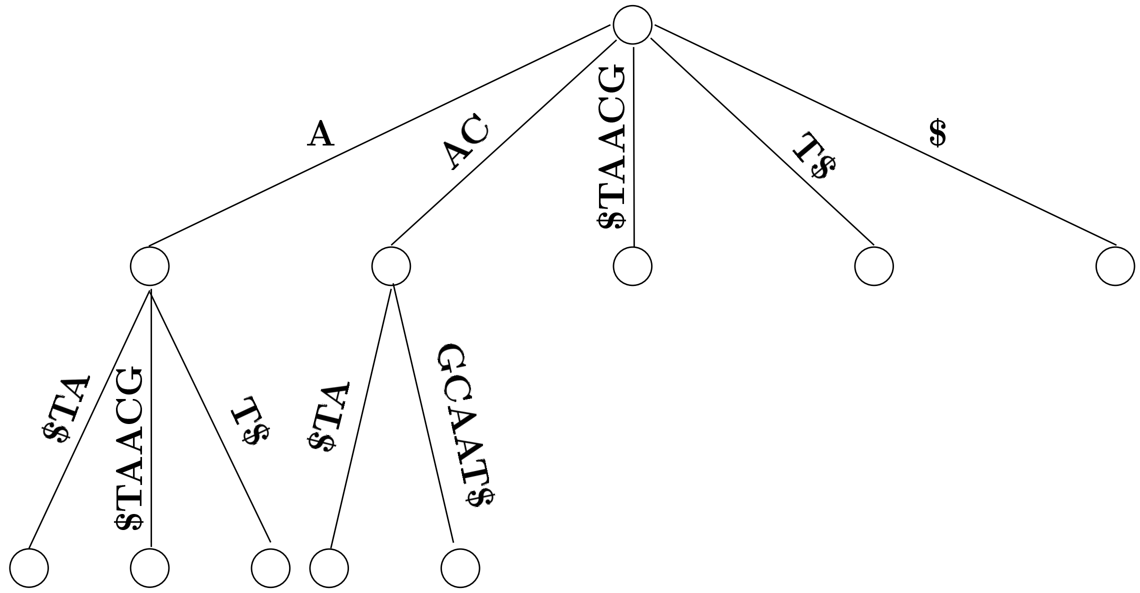
(Allali *et al.*, in preparation for Inf. Proc. Letters)

Factor tree (example for all factors of length *exactly* 4 of the string CAGCAAT)

Construction in the same time and space complexity: $O(n)$



factor tree



suffix tree for the same string

Key ideas

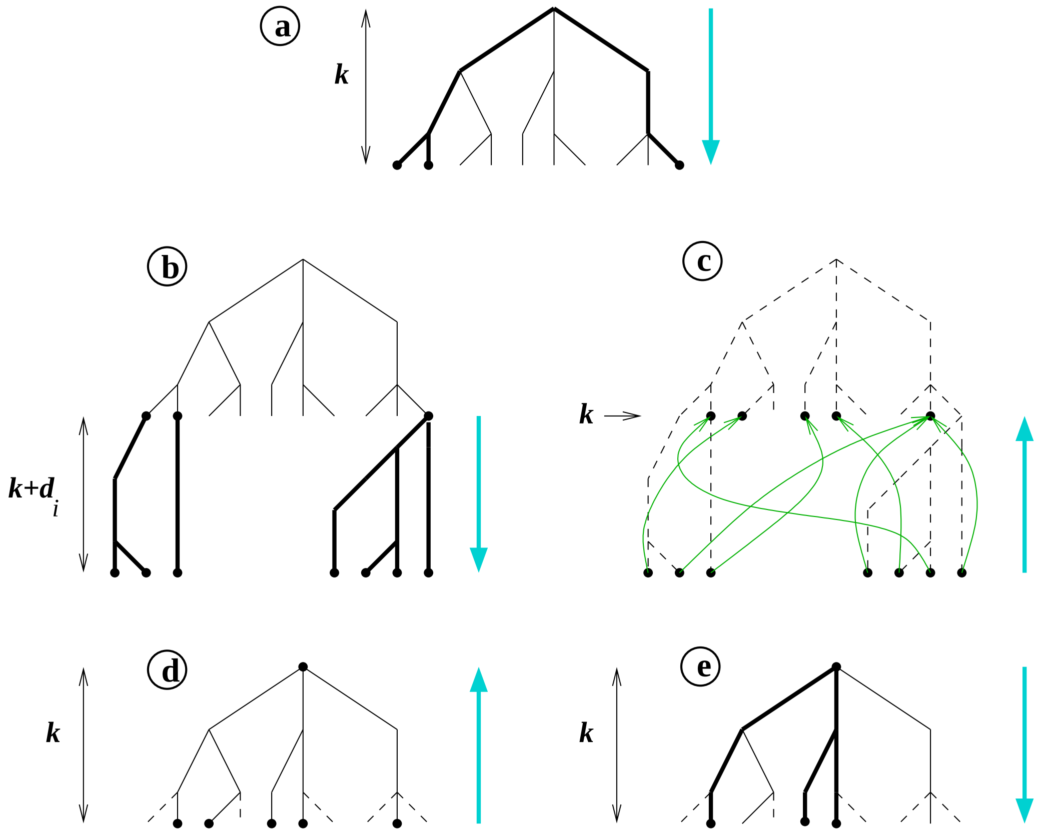
Use of an index (a generalized suffix tree) for representing the strings

in fact, what is used is a new data structure: a generalized factor tree

(Allali *et al.*, in preparation for Inf. Proc. Letters)

For models of two or more boxes, use of a generalized dynamic factor tree

Generalized dynamic factor tree



Key ideas (continuation)

Models having one prefix which does not meet the constraints will not meet it either

descent down the trees may end at first such prefix

Key ideas (continuation)

Models having one prefix which does not meet the constraints will not meet it either

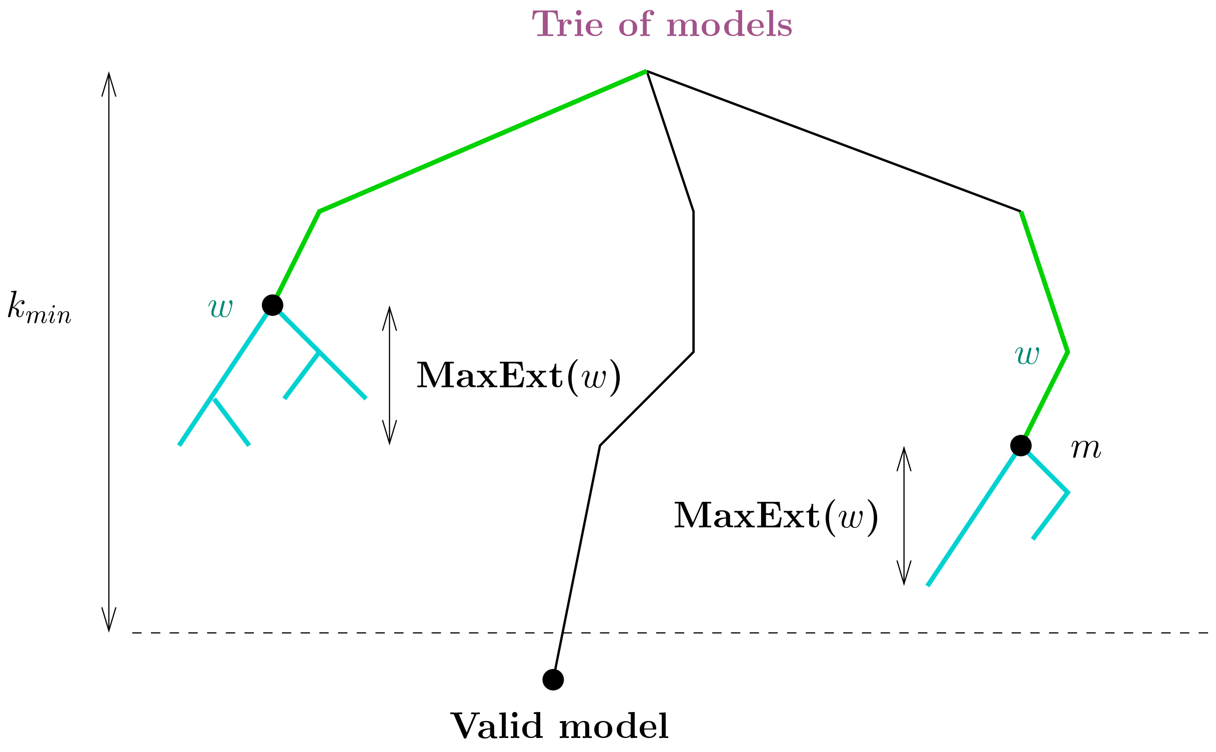
descent down the trees may end at first such prefix

Models m having a suffix which we know (from previous calculations) will not enable m to reach the minimum length constraint need not be extended further

descent down the trees may end as soon as such situation is met

(Marsan, Pisanti *et al.*, in preparation)

Maximum "extendability"



All of this has, for now, to be followed (*a posteriori*) by a statistical evaluation of the motifs found

Occurrence probability of a motif with the same properties (except quorum) in random sequences with same k-mer composition for a given value of k as estimated by simulation

A priori **occurrence probability of a motif with the same properties (except quorum) in random sequences under a Markov Model of order 0 or 1 (for now)**

(S. Robin *et al.*, submitted to JCB)

Biological applications

A comparison to other approaches

and initial application to:

Escherichia coli, Bacillus subtilis, Helicobacter pylori

both "clean" reduced datasets and big "genomic" datasets

may be found in:

A. Vanet, L. Marsan, A. Labigne, M.-F. Sagot, JMB 297:335-353, 2000

A. Vanet, L. Marsan, M.-F. Sagot, Res. in Microb. 150:779-799, 1999

L. Marsan, N. Pisanti, M.-F. Sagot, in preparation

Further, currently in progress, biological applications

Vertebrates

on a dataset used by Ingrid Lafontaine for testing a completely structural approach to promoter and regulatory sequences detection (Lafontaine, Lavery)

to identify splicing enhancer sequences (work with Claude Thermes and Yves d'Aubenton Carafa)

Synechocystis **PCC 6803**

a bacterial genome that seems not very "bacterial-like"

Mycobacterium tuberculosis & leprae **(A. Cariou)**

two very closely related genomes, one of which seems to be "shrinking" (around 1500 genes have already "disappeared" – but 165 are "new")

A summarized view of the quality of the results obtained by us, and by other methods

The nice points: we are (slightly) ahead in most cases and we can treat (sometimes much) bigger datasets

The provocative points:

everybody is reasonably good at detecting what the human eye would (almost) be capable of detecting without any help

everybody is as bad as the others at detecting subtle things, certainly at getting near the original goal

What are some of our perspectives ?

Biological questions (the goal nevertheless remains)

- obtain a catalog, as complete as possible, of all the elements involved in the expression and regulation of genes in any type of organism
- study the evolution of the regulation system and its relation with the structure of a genome

Algorithmical and statistical issues

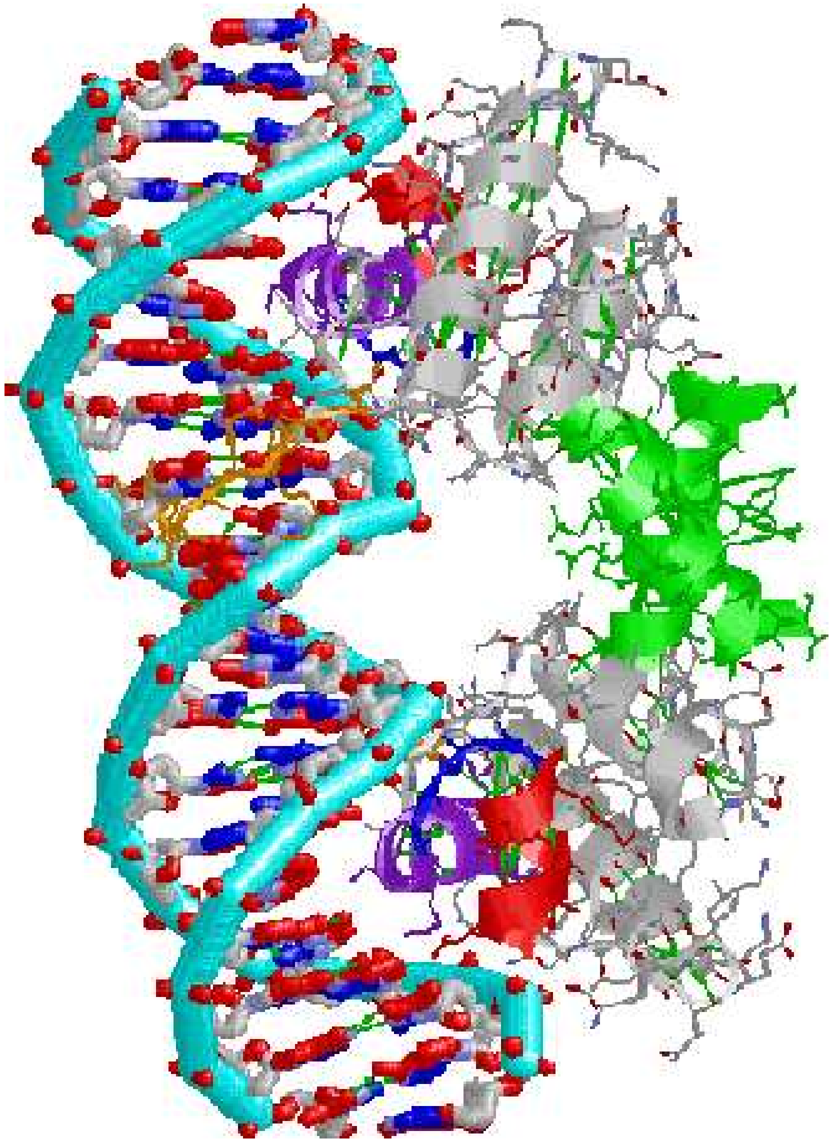
- there are many, some exceedingly beautiful ones
- curious ? $\rightarrow$ come talk to us !!

But just to give a few, hopefully seducing, hints

conservation measure which uses also structural information when known or when this may be inferred

Between the form of Life and Life
The difference is as big
As liquor at the Lip between
And Liquor in the Jug

from poem #1101, Emily Dickinson

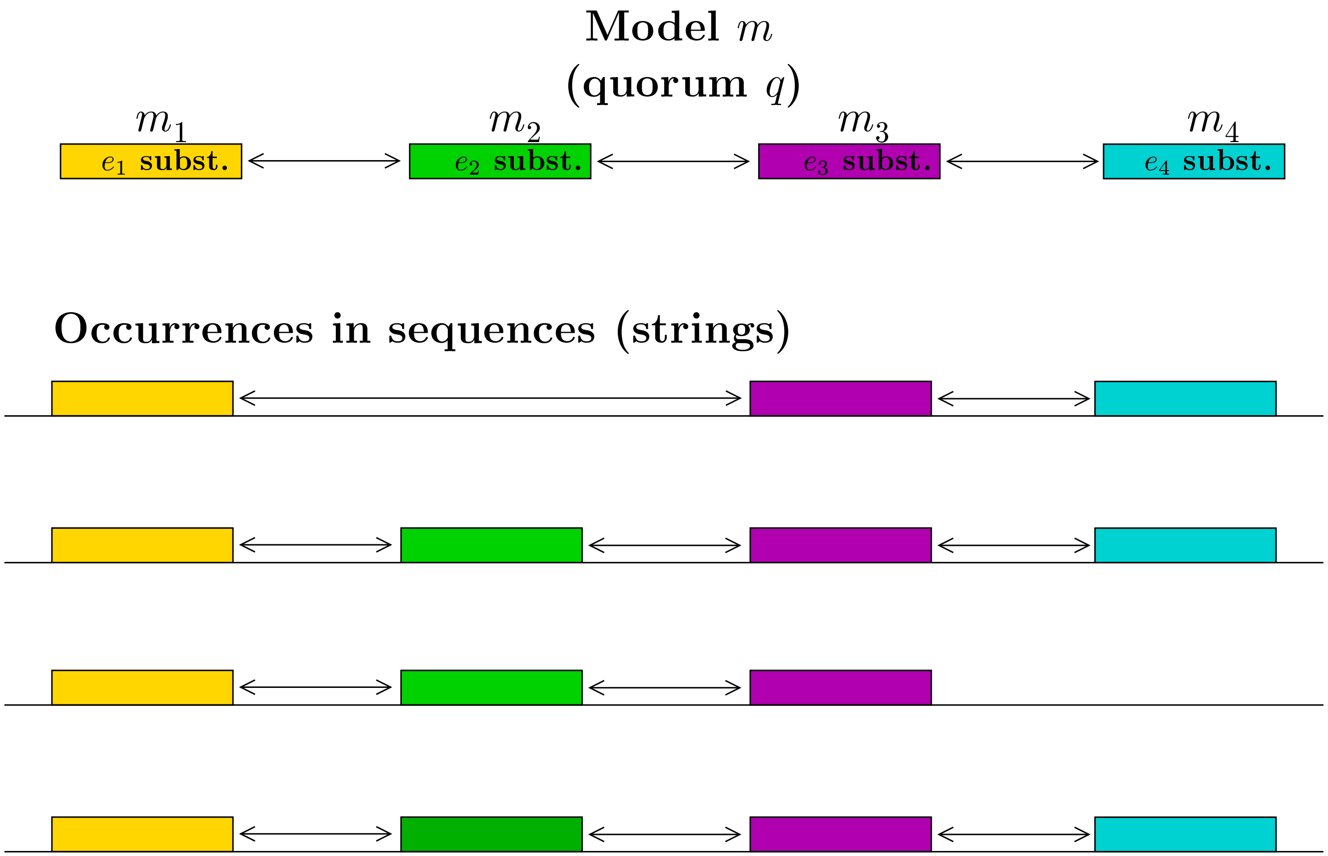


Hints (continuation)

conservation measure which uses also structural information when known or when this may be inferred

meta-conservation

Meta-conservation



Hints (continuation)

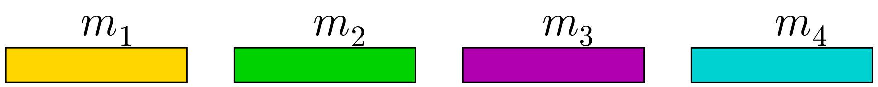
conservation measure which uses also structural information when known or when this may be inferred

meta-conservation

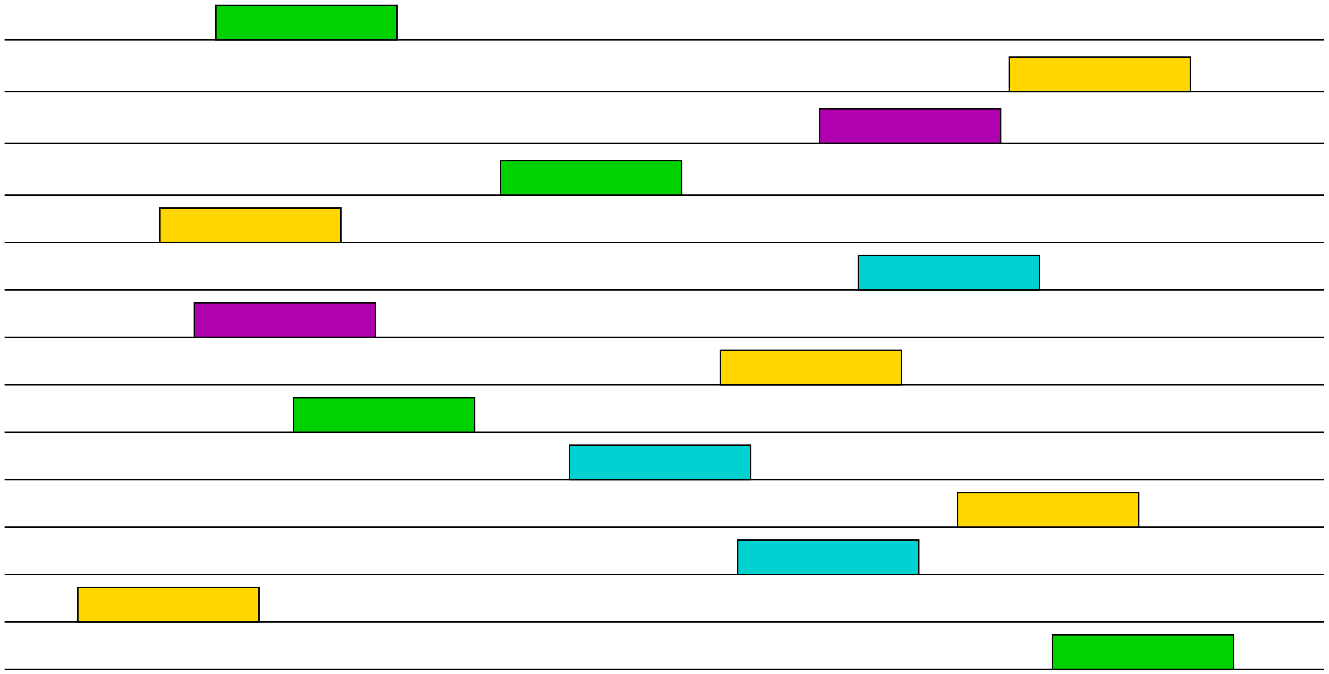
covering motifs

Covering motifs

Minumum set of models (one box) satisfying the properties

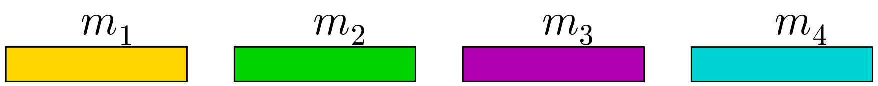


that "covers exactly" a set of sequences (strings)

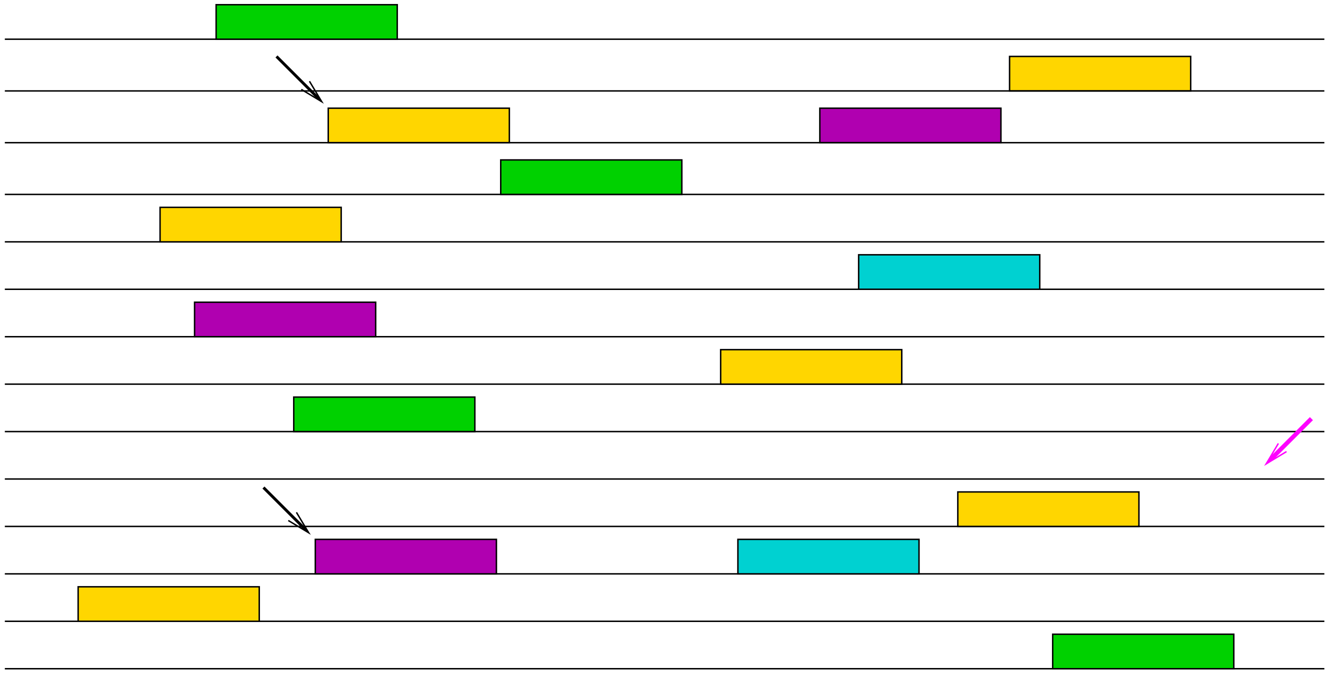


Covering motifs (variant)

Minumum set of models (one box) satisfying the properties



that "covers approximately" a set of sequences (strings)



Hints (continuation)

conservation measure which uses also structural information when known or when this may be inferred

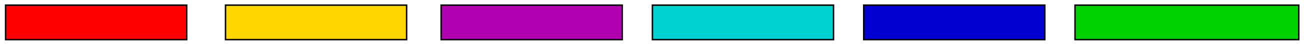
meta-conservation

covering motifs

(a more technical but biologically no less important issue) basis of motifs

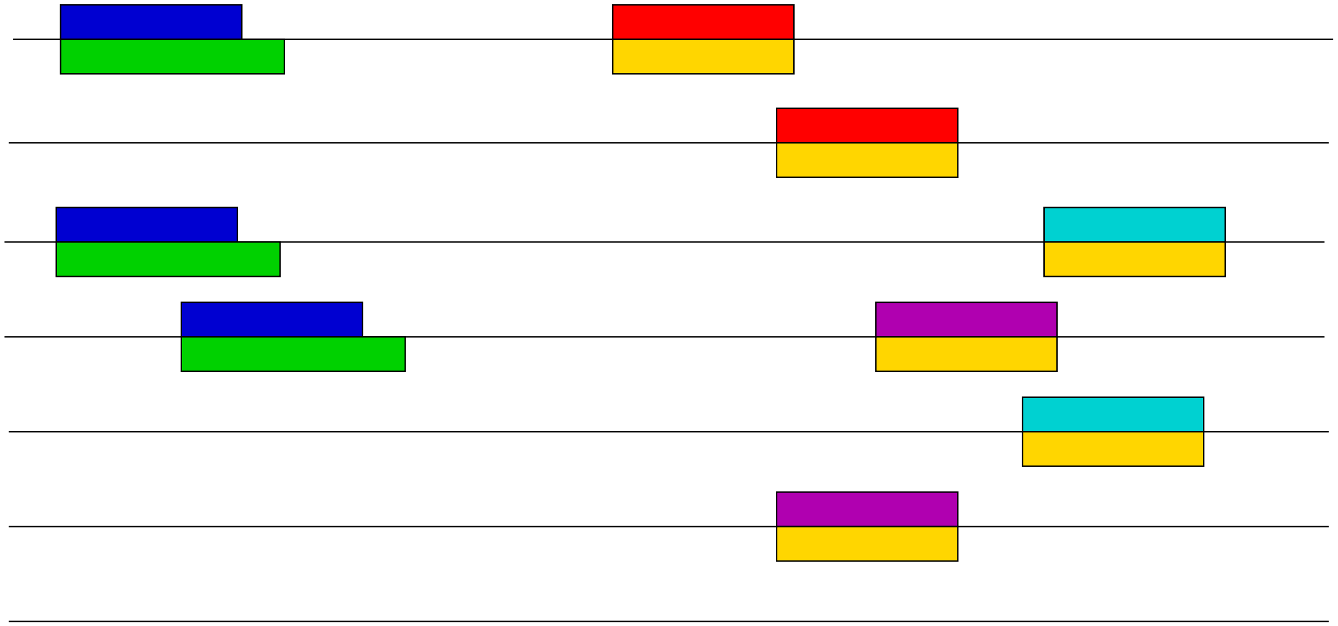
Basis of motifs

Models (one box) found with no differences and written over alphabet DNA/protein plus wild card



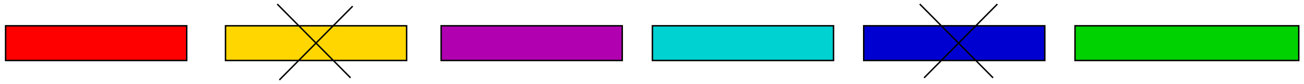
Occurrences in sequences (strings)

keep only maximal motifs | **keep only irredundant motifs**



“Darling, I shrinked the motifs”

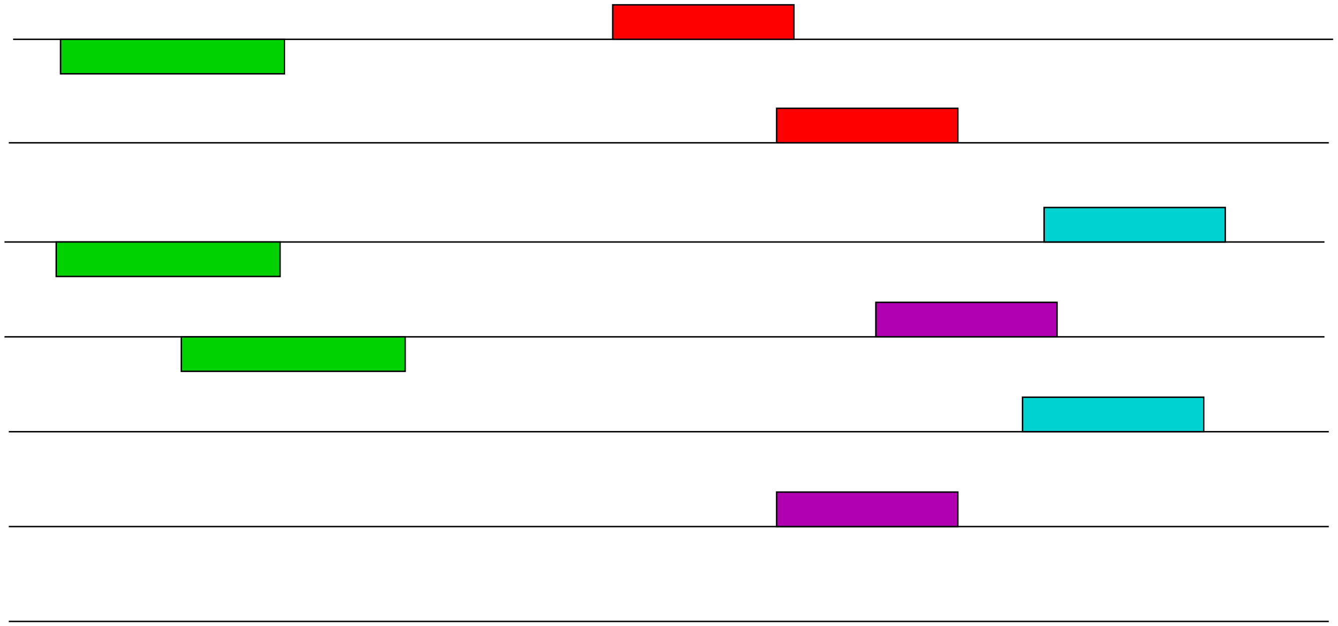
Models (one box) found with no differences and written over alphabet DNA/protein plus wild card



Occurrences in sequences (strings)

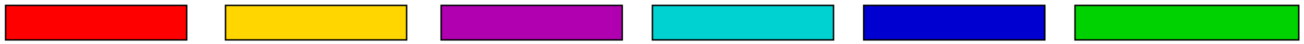
keep only maximal motifs

keep only irredundant motifs



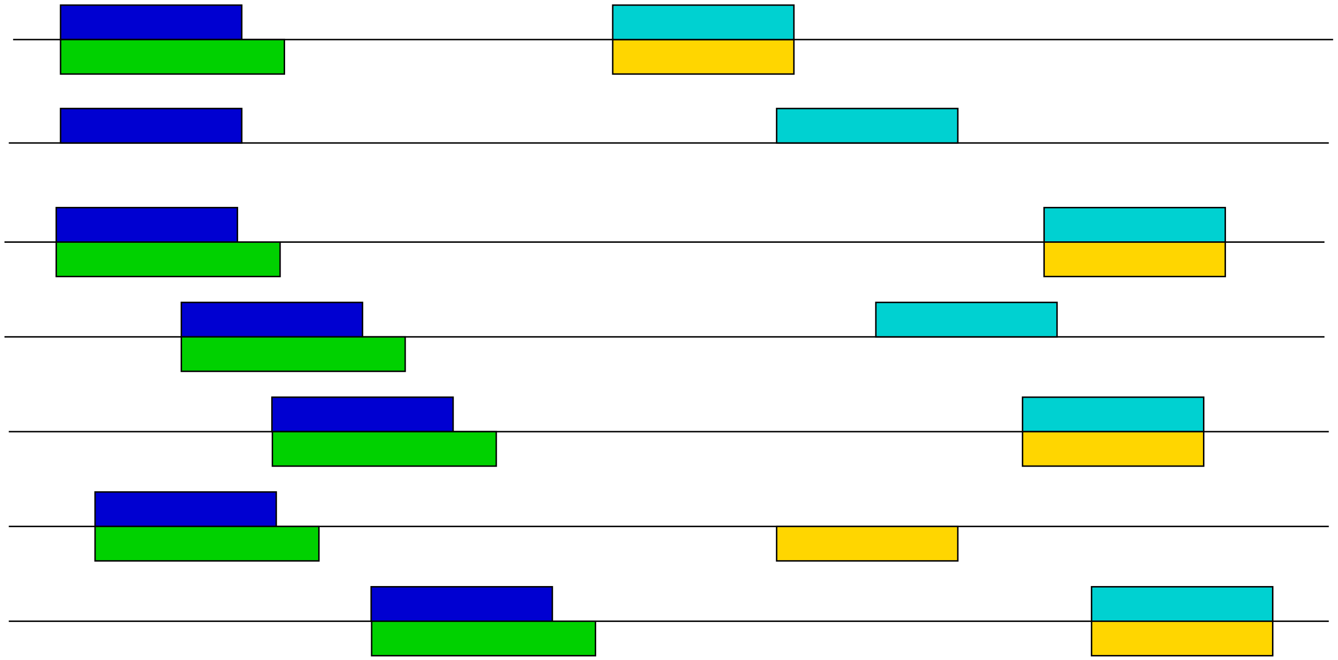
Less clear-cut situations

Models (one box) found with differences and/or representing network expressions (DNA or protein)



Occurrences in sequences (strings)

Two examples of less clear-cut situations



Second area of research:

GENOME STRUCTURE

what is meant by “genome structure” ?

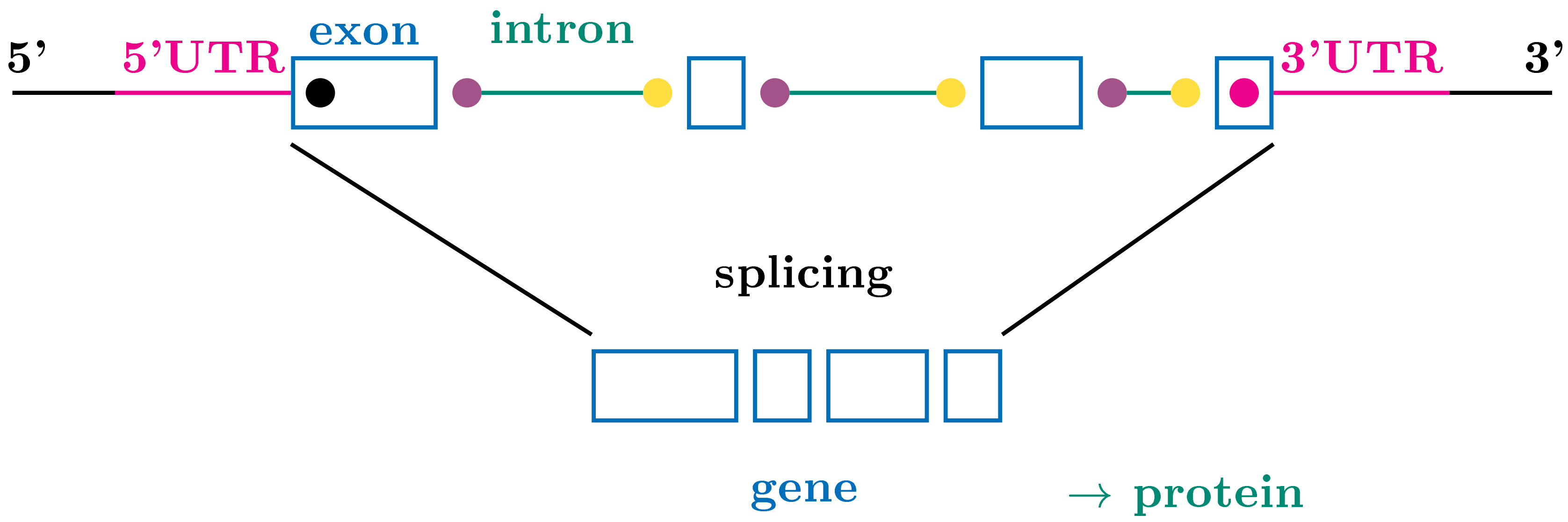
a division of the genome into well-identified (big) units

genes in eukaryotes (such as vertebrates)

A very elementary view of an eukaryotic gene

start codon **GT (donor site)** **AG (acceptor site)** **stop codon**

3 nucleotides (codon) → 1 amino acid



5'UTR and 3'UTR: transcribed into RNA but not translated

what is meant by "genome structure" ?

a division of the genome into well-identified (big) units

genes in eukaryotes (such as vertebrates)

repeated elements (dispersed or in "organized fashion", in particular transposons, satellites and protein domains)

Satellites informally put

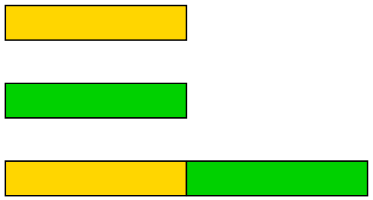
(this represents a continuous segment of a DNA sequence)

```
gaaagctcttctgctccagtaccaactccatccagctctactact
gaaagctcttctgctccagt   aa     ccagctctacCact
gaaagctcttctgctccagtaccaactccatccagctctactact
gaaagctcttctgctccagCTccaacCccatcAagctcCacCact
gaaagctcttctgctccagt   aa     ccagctctacCact
gaaagctcttctgctccagtaccaactccatccagctctactact
gaaagctcttctactccagt   aa     ccagctccacCact
gaaagctcttctgctccagtGccaactccatcAagctctactact
gaaagctcttctgctccagtaccaactccatcAagctctactact
gaaagctcttctgctccagCTccaacCccatcAagctctactact
gaaagctcttctgctccagt   aa     ccagctctacCact
gaaagctcttctgctccagtaccaactccatccagctctactact
gaaagctcttctgctccagtaccaacCccatccagctctactact
gaaagctcttctgctccagtaccaactccatccagctctactact
```

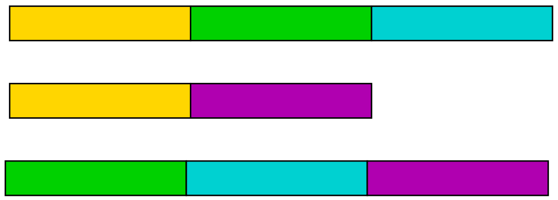
A schematic view of protein domains and their combinatorics

Proteins from same or different organisms

Fused domains



The combinatorics may be higher



what is meant by "genome structure" ?

a division of the genome into well-identified (big) units

genes in eukaryotes (such as vertebrates)

repeated elements (dispersed or in "organized fashion", in particular transposons, satellites and protein domains)

(of course, there may be other units)

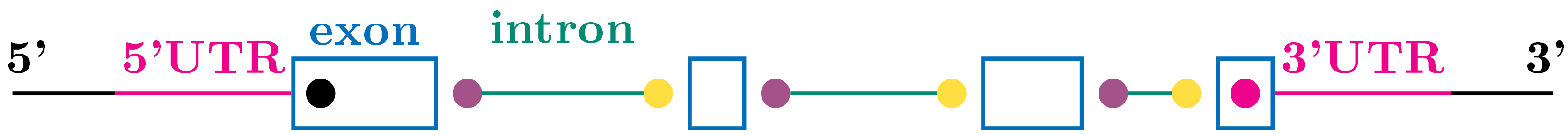
Sorry, but all this part will need to be skipped

it has many links with the previous one though

Genes may be seen as (long) motifs in variable number and with different properties (satisfying "gene model")

start codon **GT (donor site)** **AG (acceptor site)** **stop codon**

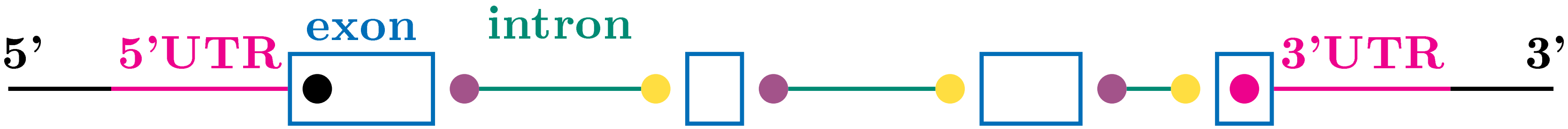
3 nucleotides (codon) $\rightarrow$ 1 amino acid



Genes may be seen as (long) motifs in variable number and with different properties (satisfying "gene model")

start codon **GT (donor site)** **AG (acceptor site)** **stop codon**

3 nucleotides (codon) → 1 amino acid



Two perspectives

introduce a multiple (-gene and -organism) approach

put together gene detection and motif inference

In the same way, repeats are, possibly long, motifs with, sometimes, further constraints

repeats appearing contiguously (satellites)
repeats appearing in different contexts (protein domains)
repeats with an evolutionary history

In the same way, repeats are, possibly long, motifs with, sometimes, further constraints

repeats appearing contiguously (satellites)
repeats appearing in different contexts (protein domains)
repeats with an evolutionary history

Some biological perspectives (work of R. Tavares)

is there a relation between the distribution of repeats (tandem, palindromic or dispersed) along a genome and the gene expression level ?

can we detect and use the combinatorics of protein domains in order to infer functional links ?

Our work up to now in this area comprises:

Utopia (P. Blayo *et al.*, TCS, in press)

“satellites” (Sagot and Myers, JCB, 5:539-554, 1998)

Third area of research:

MOLECULES IN SPACE: Starting to get things out of the jug/paper

what is covered by this (very broad) title ?

longer range interactions which may or may not be related, directly or indirectly, to a structure

(how may this be detected ? by trying to detect "correlated mutations")

the structure itself of a macromolecule

using known or predicted information concerning it

(to, for instance, improve the prediction of motifs)

inferring it

The problems we have been adressing for now

obtaining a represention for comparing RNA secondary (tertiary) structures

exploring tree and graph (local) distances

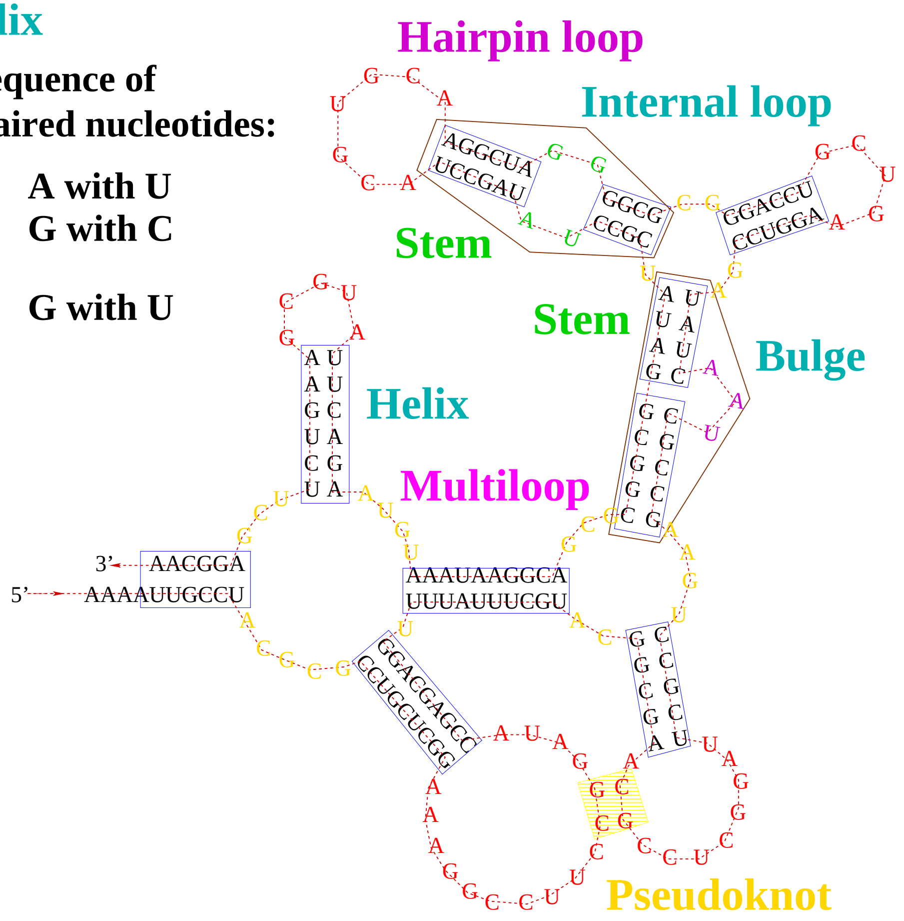
A simplified view of an RNA secondary structure

Helix

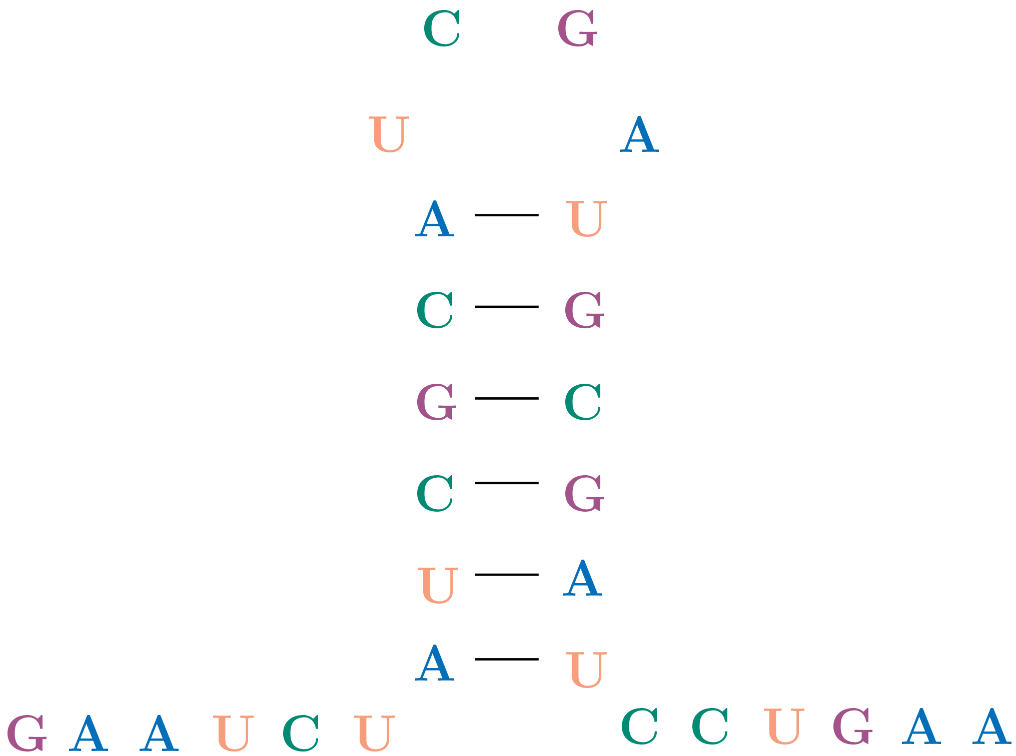
sequence of paired nucleotides:

A with U
G with C

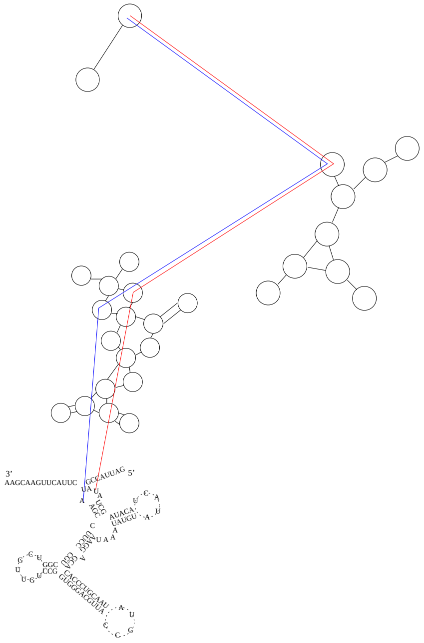
G with U



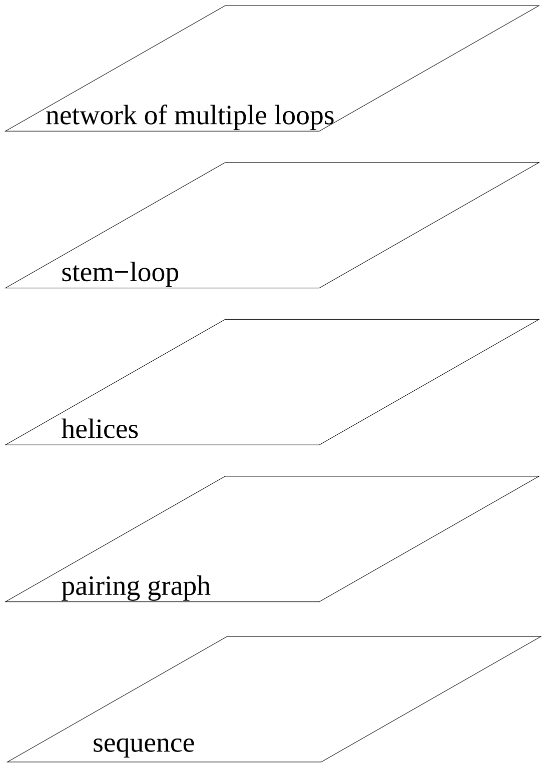
An helix



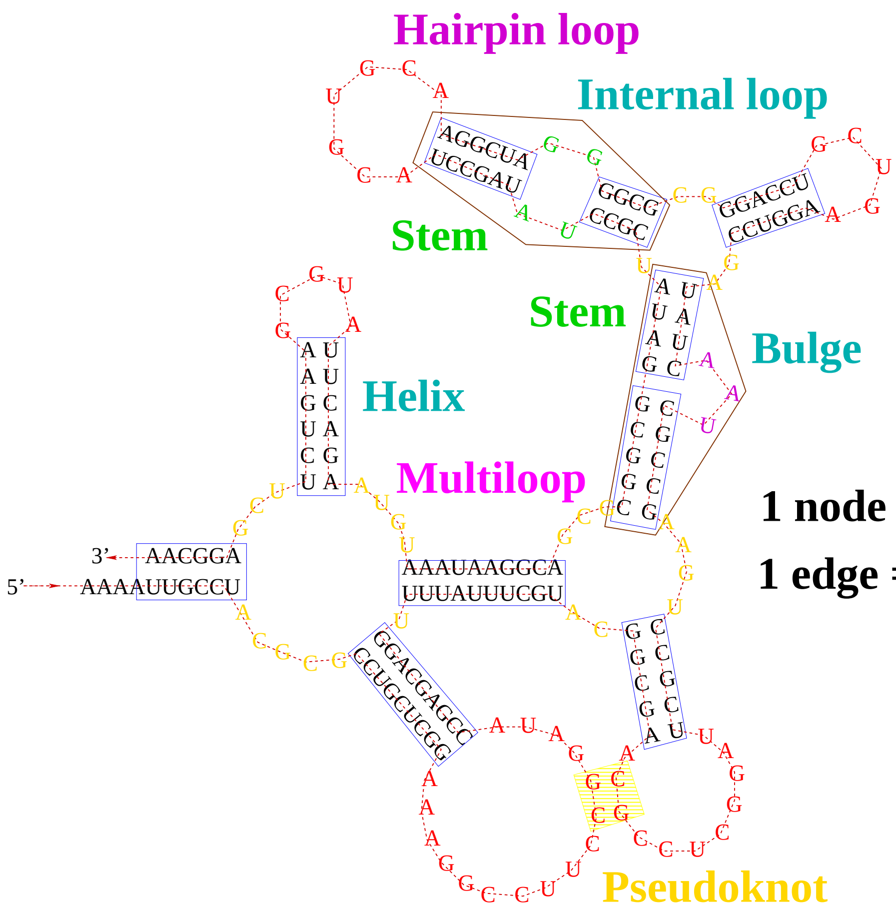
The representation we have obtained: MIGAL (“MultIple GrAph Layer”)



ACGCCUGCAAUGCAC.........



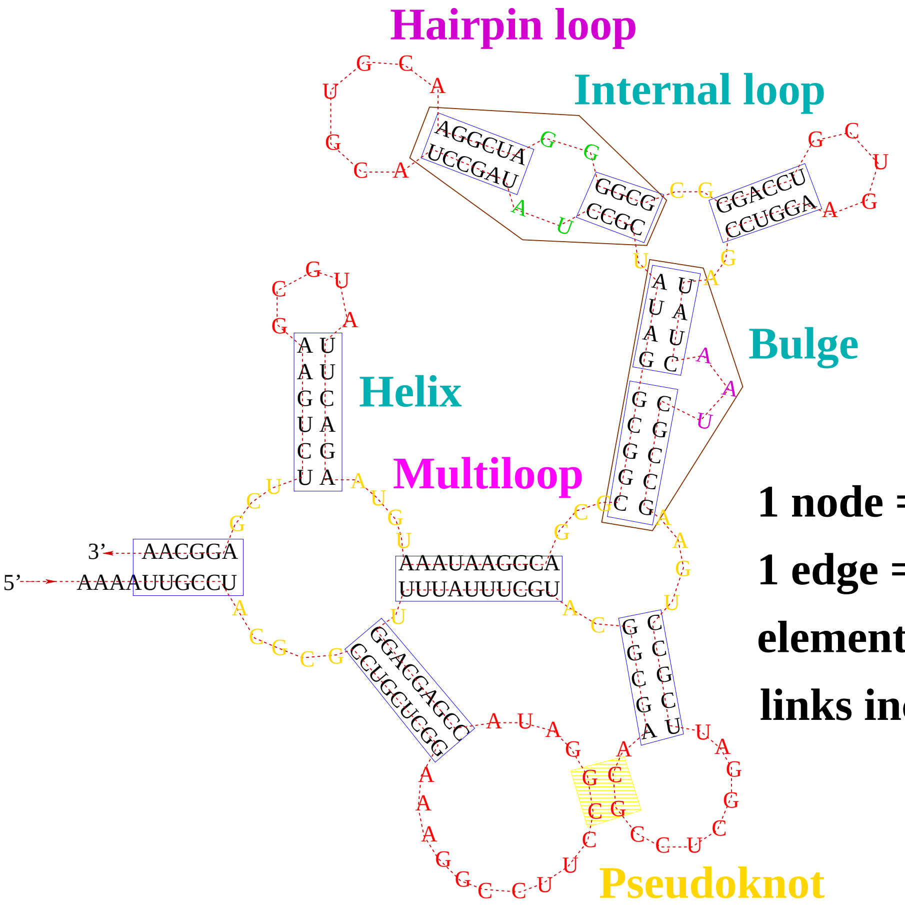
Pairing graph



1 node = 1 nucleotide

1 edge = 2 paired nucleotides

Helix graph



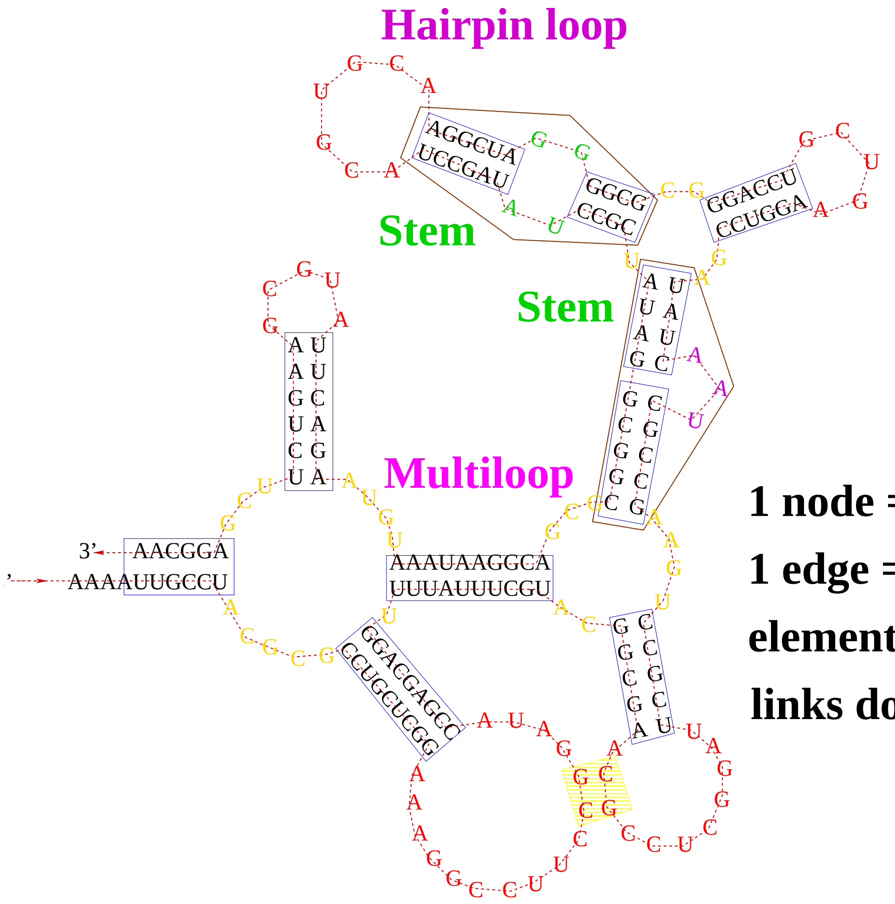
1 node = 1 structural element

1 edge = 2 linked elements

elements = helices, bulges and loops

links include pseudoknots

Stem tree



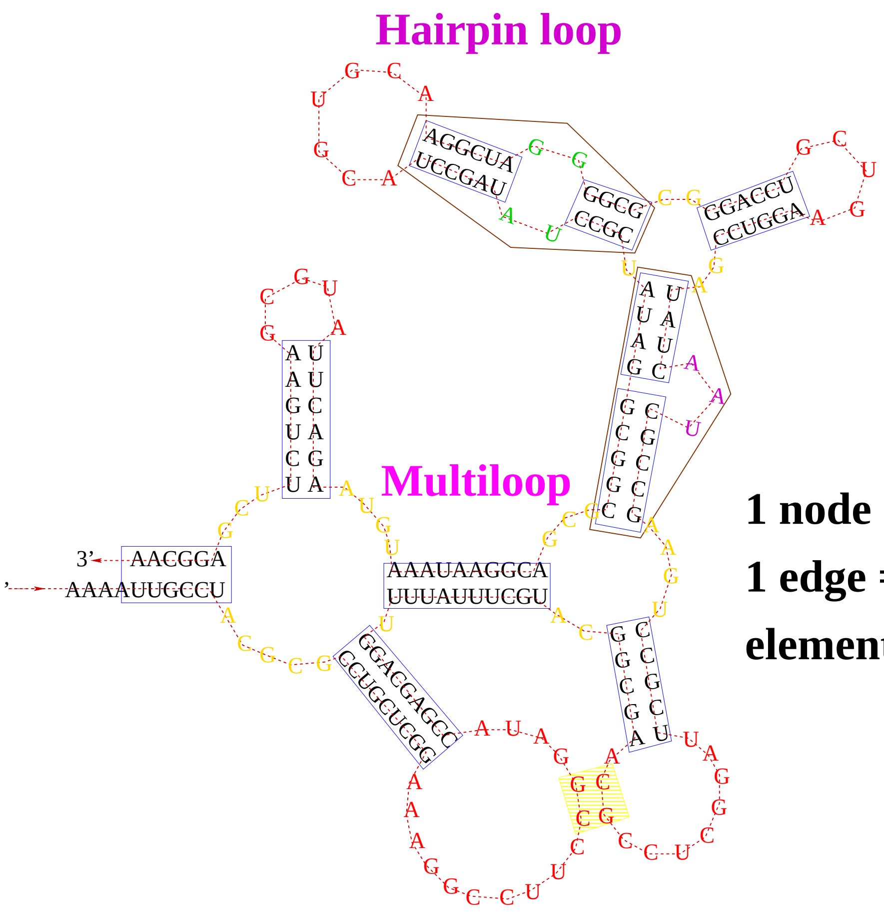
1 node = 1 structural element

1 edge = 2 linked elements

elements = stems, hairpins, multiloops

links do NOT include pseudoknots

Multiloop network (tree)

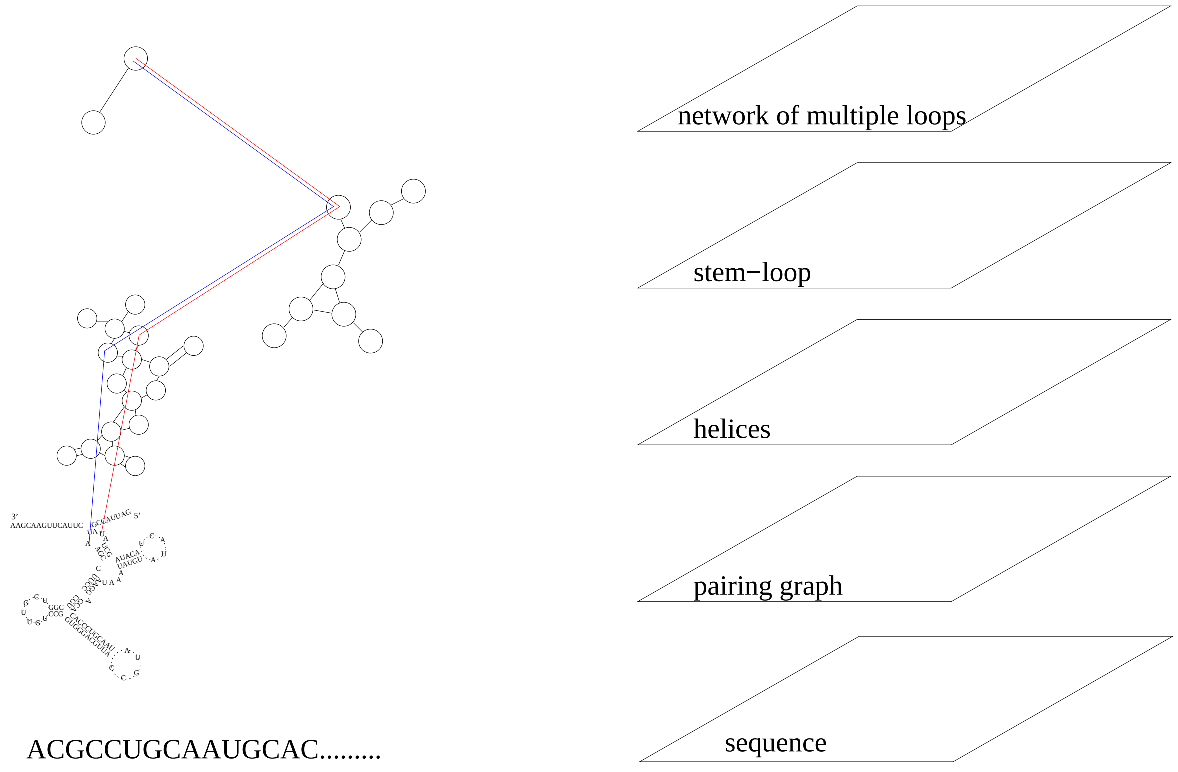


1 node = 1 structural element

1 edge = 2 linked elements

elements = multiloops (internal nodes)
and hairpin loops (leaves)

The representation we have obtained: MIGAL (“MultIple GrAph Layer”)



An implementation exists (J. Allali, mémoire de DEA)

What are some of our perspectives ?

Biological questions

- given two or more structures, calculate a local distance between them and find a model which represents what they have in common
- given a (possibly vague) description of a structure, find all sequences which may fold according to it
- given a set of sequences, find the element(s) of structure they may have in common

Algorithmical issues

- there is a huge number, some exceedingly beautiful ones
- in fact, two complete areas of research to develop
- curious ? → come talk to us !!

But just to give a, hopefully seducing, hint

Introduce notions of "allowed" differences that may represent two very distinct concepts

biological variability

errors (in general man-made) in the data

into tree and graph theory and algorithms

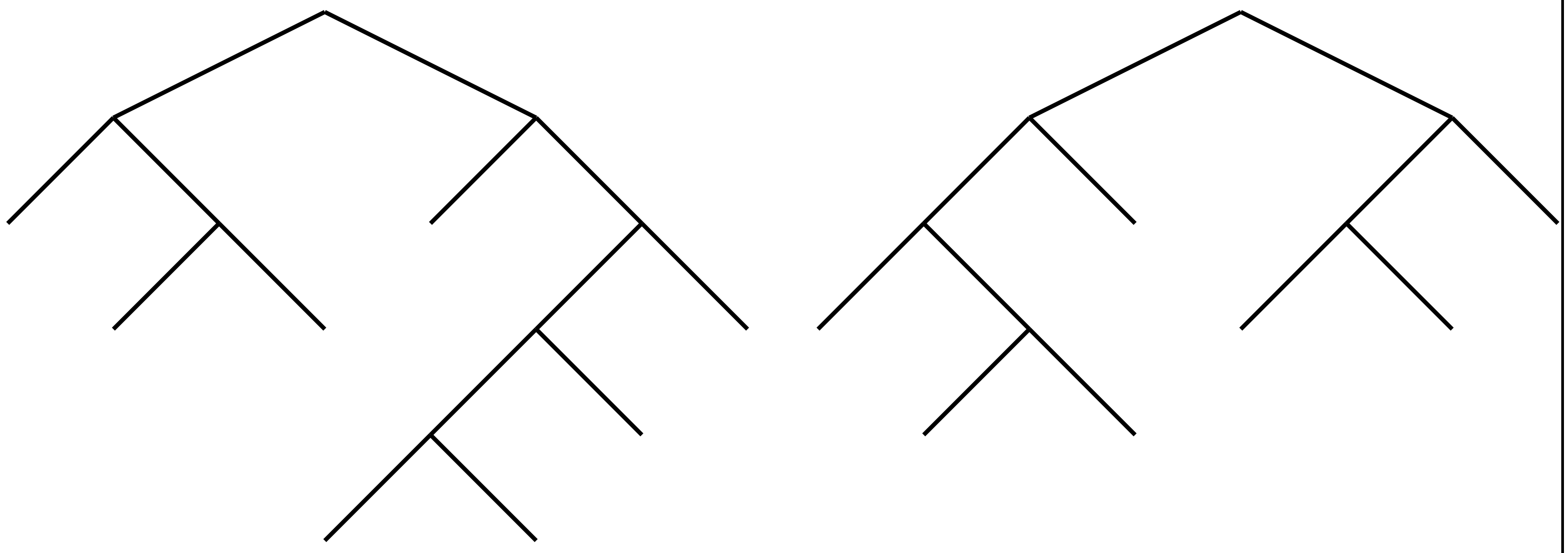
Leads to problems of identification of Maximum Agreement SubTree (MAST)

that is, of maximum homeomorphic subtree common to two, or more than two, trees

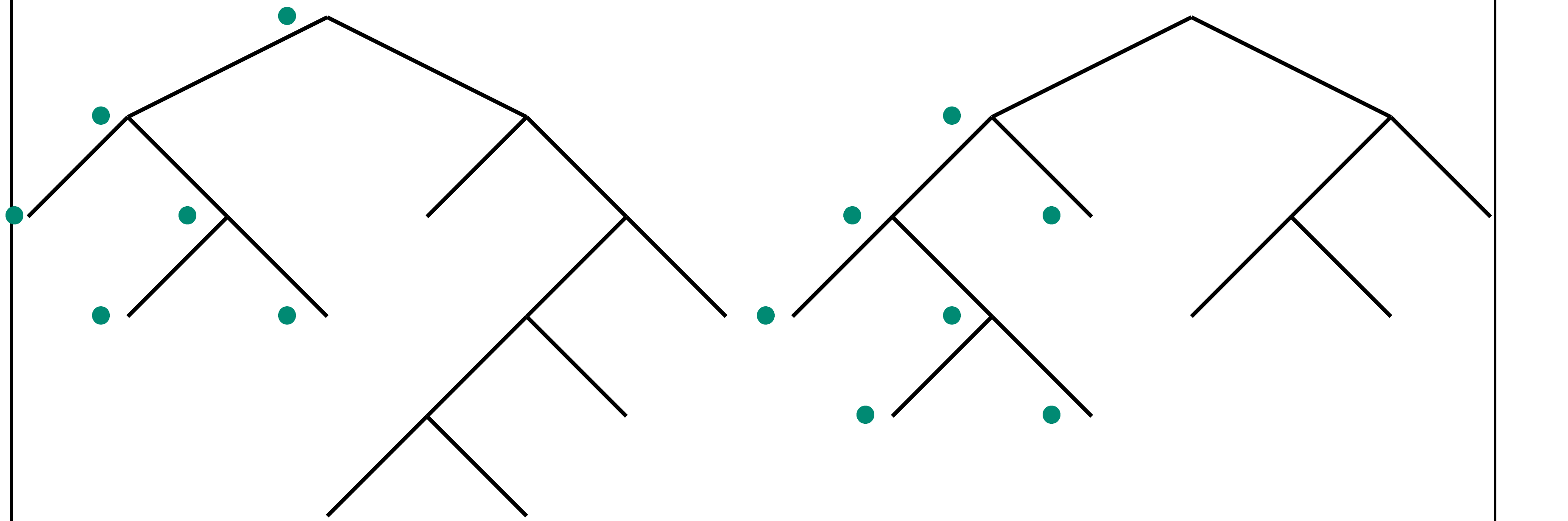
where the trees, in this case, are unrooted and oriented

and are valued at both the nodes and edges

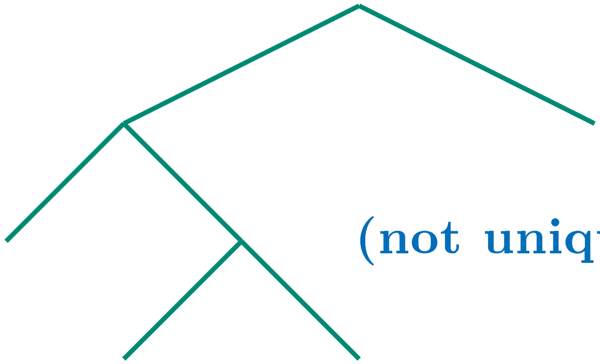
Maximum Agreement SubTree (MAST)



Maximum Agreement SubTree (MAST)

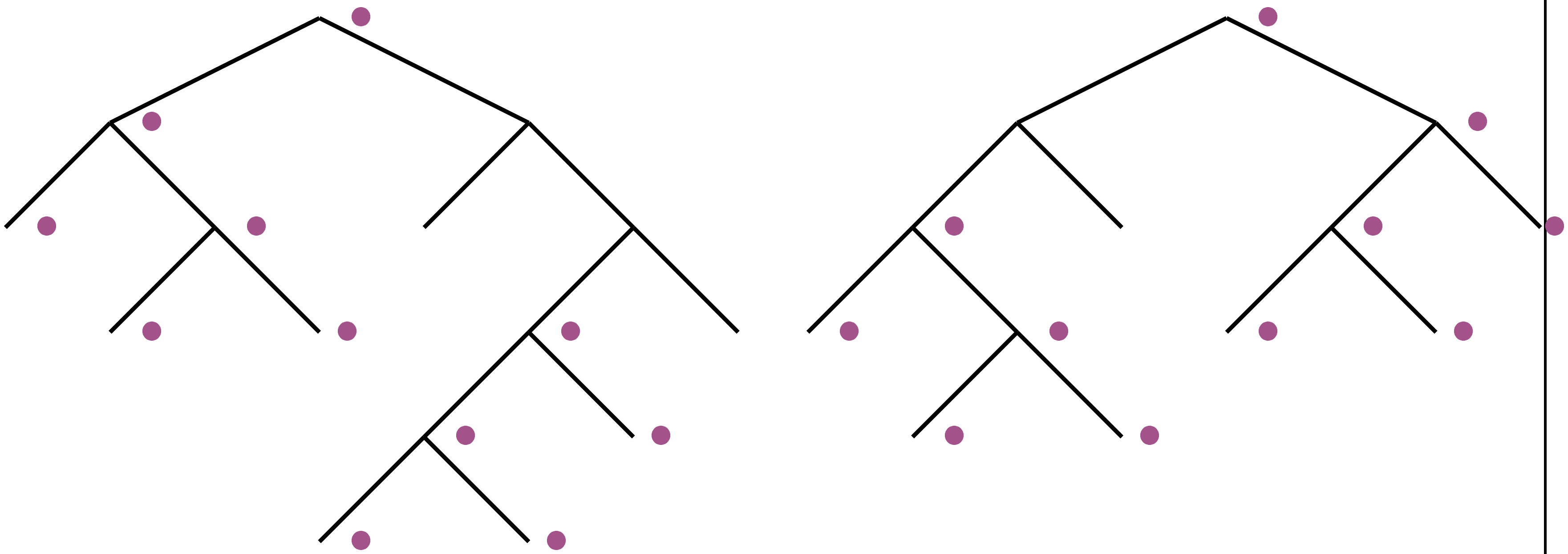


Max. isomorphic subtree

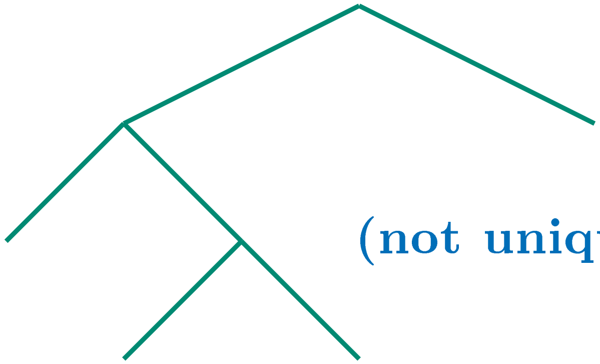


(not unique)

Maximum Agreement SubTree (MAST)

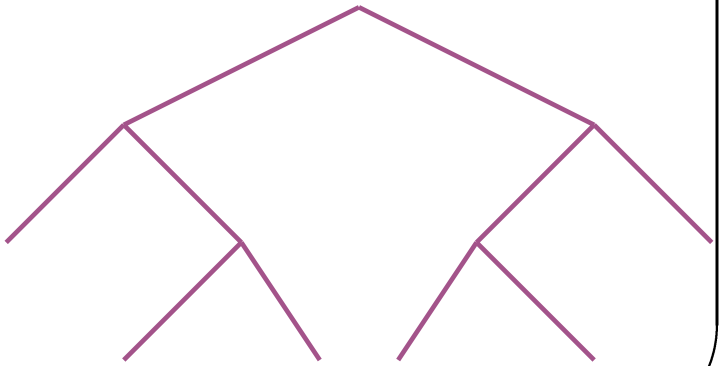


Max. isomorphic subtree



(not unique)

Max. homeomorphic subtree



Equivalent problems will have to be considered for graphs

The future

Another problem within "MOLECULES IN SPACE" which was already mentioned

introducing structural considerations into the protein motif inference problem

for which an algorithm is, however, available

CoSamp (for Combinatorial Sampling – Créno *et al.*, not published yet)

which introduces not (yet) structural but physico-chemical considerations

Fourth area of research:

GENOME DYNAMICS: Starting to add movement, and so life, into the story

Listen ! a frog

Jumping in the stillness

Of an ancient pond !

Basho (Spring days)

Listen ! a frog

Jumping in the stillness

Of an ancient pond !

Basho (Spring days)

Listen ! a piece of *E. coli*

Jumping in the stillness

Of my genome !

slight modification of
a poem by Basho

Our concern will therefore be, not with frogs jumping about and into ponds, but with

> **genomic fragments "jumping" (sometimes making copies of themselves) intra or inter-strains (species), or "disappearing"**
>
> > **how to identify with a fair degree of confidence and accuracy the genetic transfers that took place ?**
> >
> > **how to use rearrangement events to calculate a distance and infer a phylogeny between organisms ?**

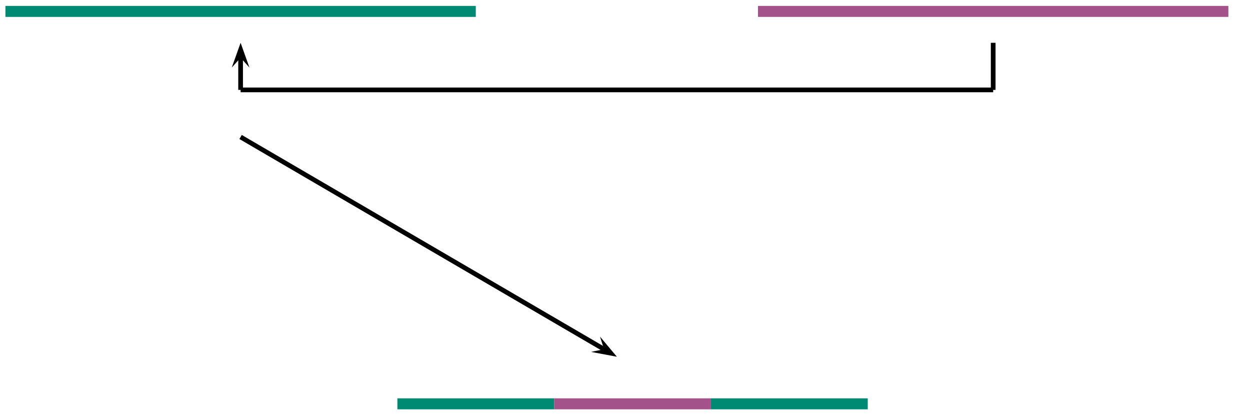
First sub-area of GENOME DYNAMICS:

Genetic transfer detection and localization (at a fine level)

A rapid word on genetic transfer

Horizontal transfer

species or strain $\mathbf{X}_1$ **species or strain $\mathbf{X}_2$**



A rapid word on genetic transfer (continuation)

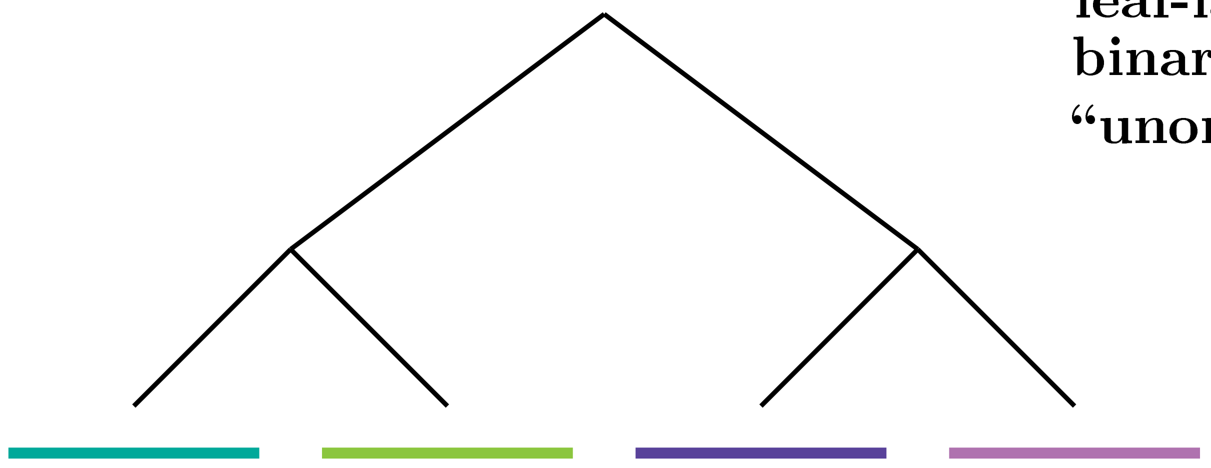
Conversion

species X

Effect of transfer on the history of the organisms

Reminder: Evolutionary history represented by means of a so-called phylogenetic tree

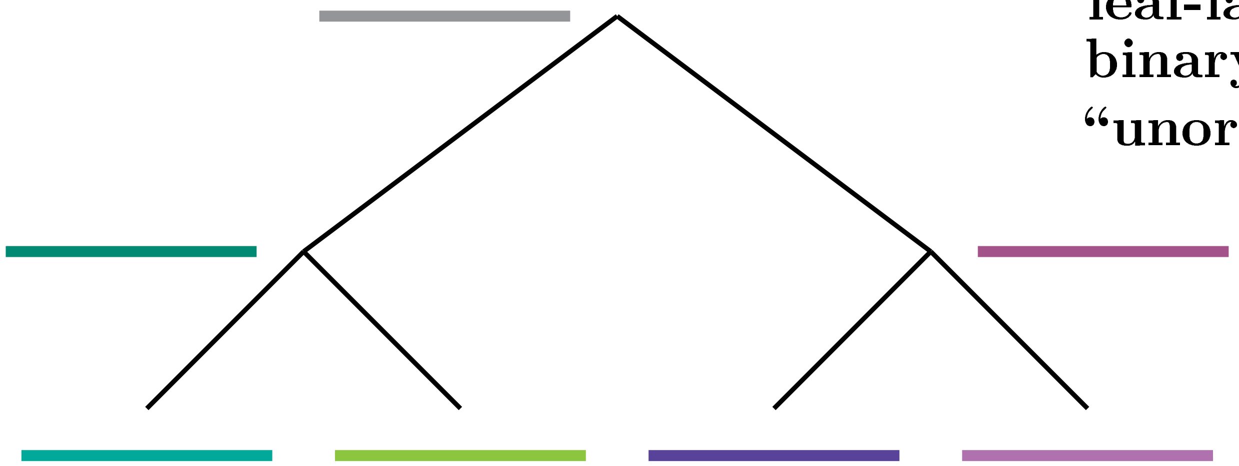
rooted
leaf-labeled
binary
“unordered”



Effect of transfer on the history of the organisms

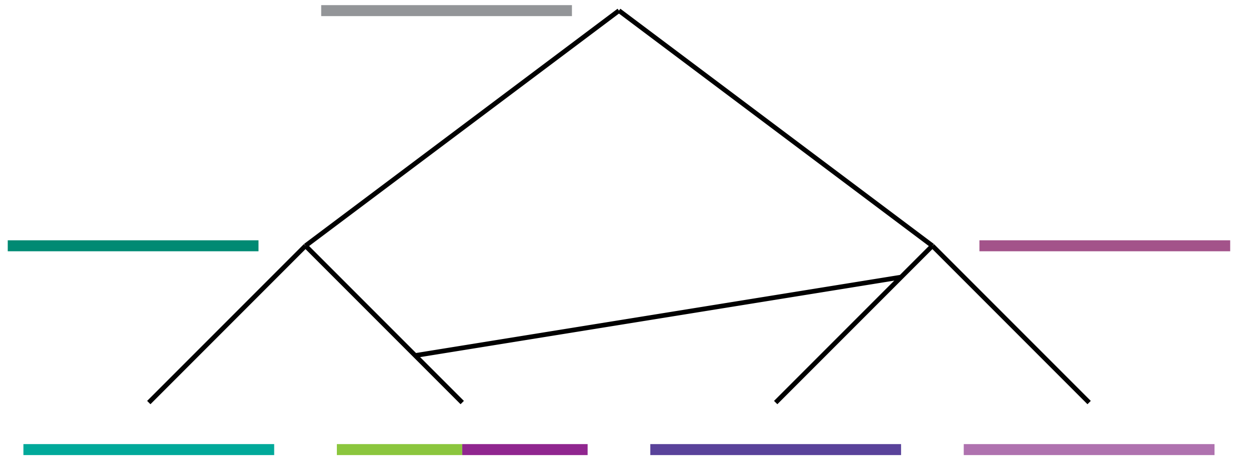
Reminder: Evolutionary history represented by means of a so-called phylogenetic tree

rooted
leaf-labeled
binary
"unordered"



Effect of transfer on the history of the organisms

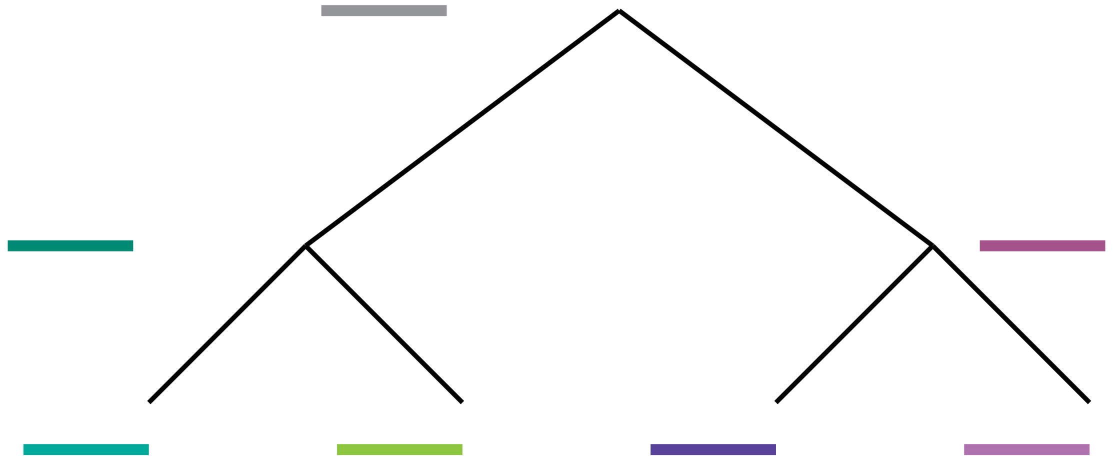
Not a single tree anymore



Effect of transfer on the history of the organisms

Not a single tree anymore **but two ! (or more)**

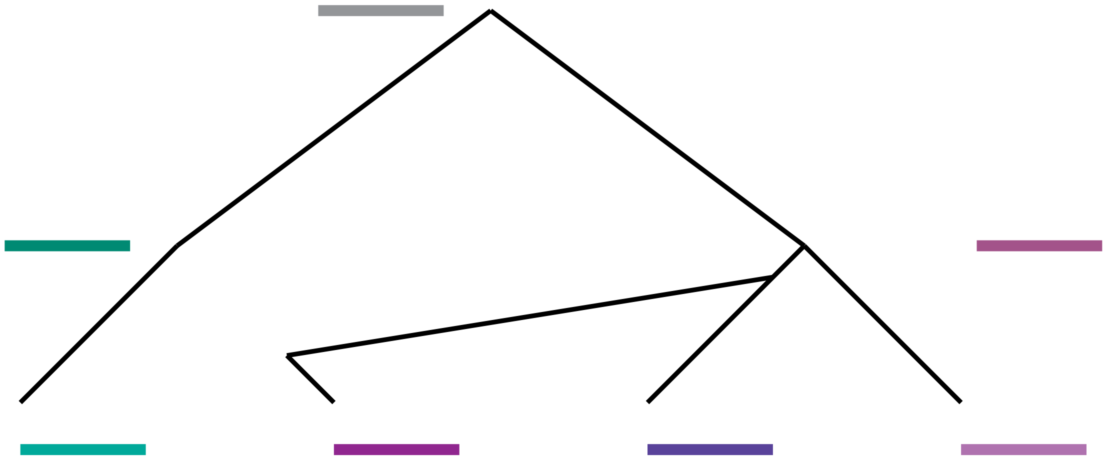
the first:



Effect of transfer on the history of the organisms

Not a single tree anymore but two ! (or more)

the second:



Genetic transfer detection and localization

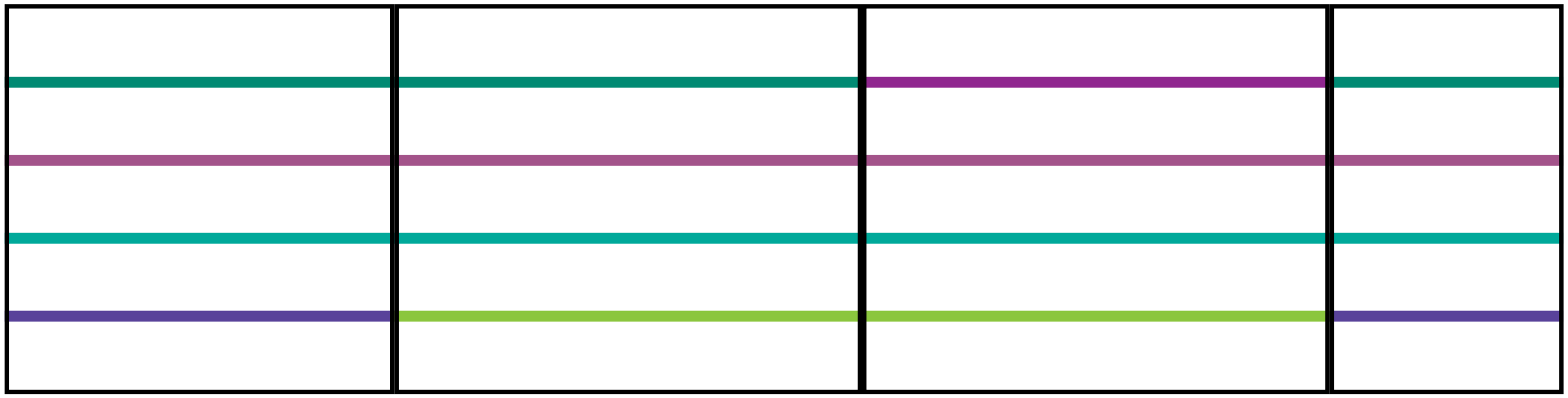
Data

a set of phylogenetically "correctly" aligned sequences from closely related organisms

Genetic transfer detection and localization as an optimisation problem

The (ideal ?) result

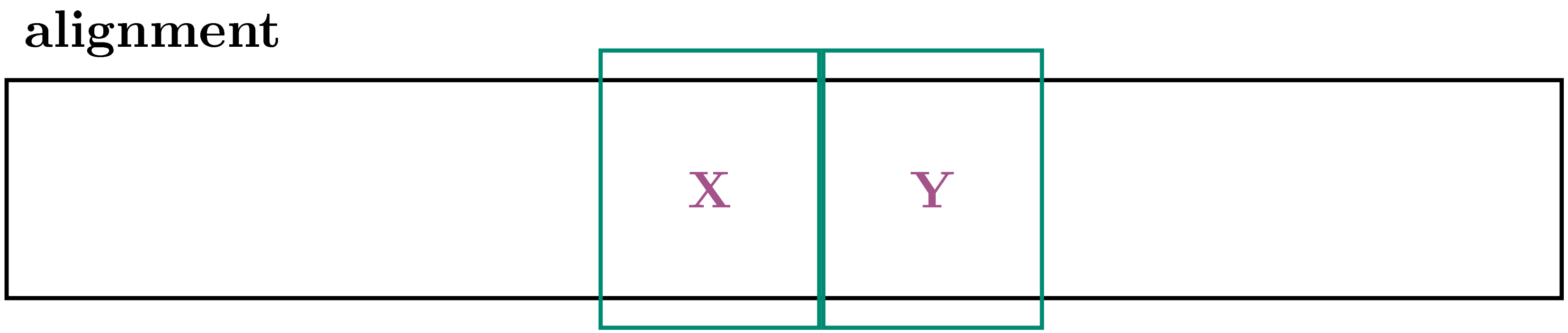
alignement



tree 1 **tree 2** **tree 3** **tree 4**

"optimal" partition: sum of lengths of trees + penalty for genetic transfers is MINIMUM

Main (crucial) questions



How to calculate the distance between two trees in this specific case ?

- length (number of mutations) difference
- genetic transfer (inter-species or inter-species rearrangement) distance

How to determine whether the obtained distance is statistically significant ?

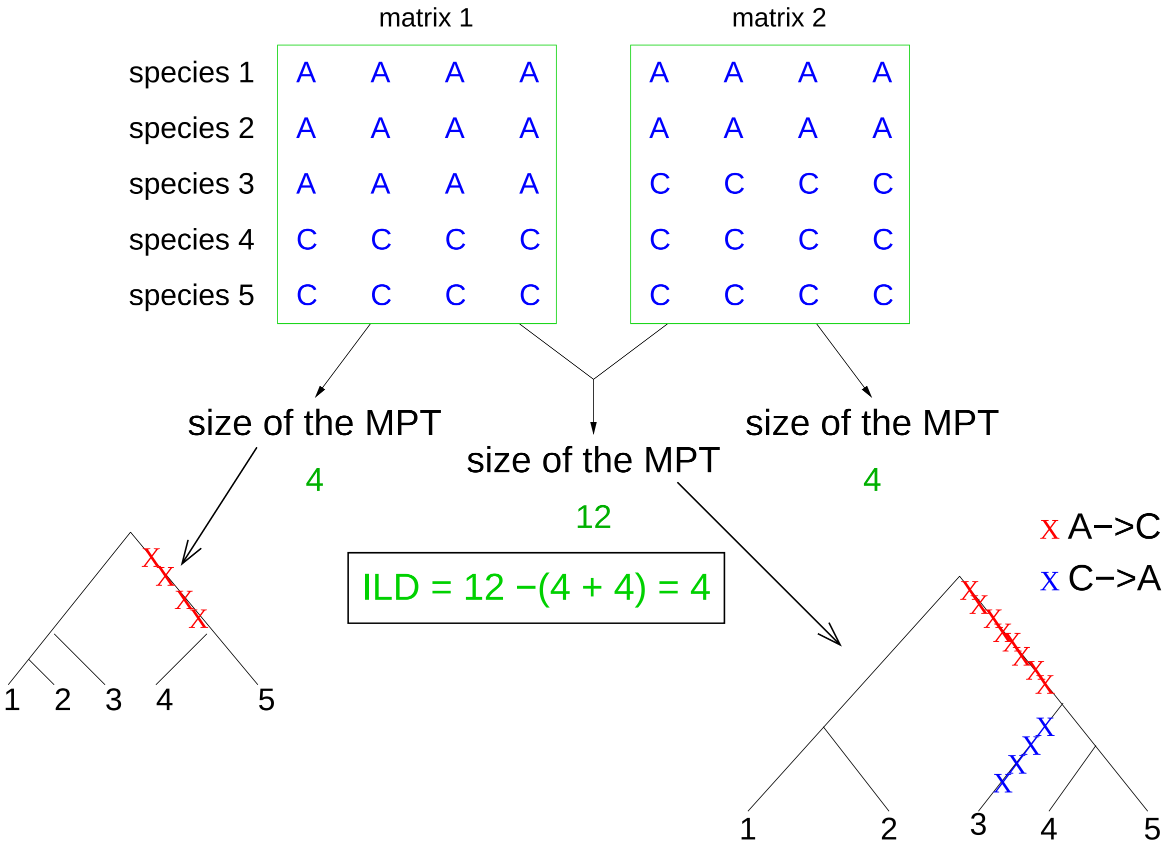
Genetic transfer detection and localization: a first pragmatic approach

Slide a double (small) window along the alignment

species \ sites	1	2	3	4	5	6	7	8	9	10	11	12	...	n
1	A	T	G	G	A	G	C	T	A	C	T	C	...	T
2	A	T	A	G	A	G	A	T	A	T	A	C	...	T
3	A	C	A	T	C	C	A	T	A	T	A	C	...	T
4	C	C	T	T	C	C	A	G	C	C	T	G	...	G
5	C	C	T	T	A	C	C	G	C	C	T	G	...	G
				matrix X					matrix Y					

Genetic transfer detection and localization: a first pragmatic approach

Use an Incongruence Length Difference (ILD) test

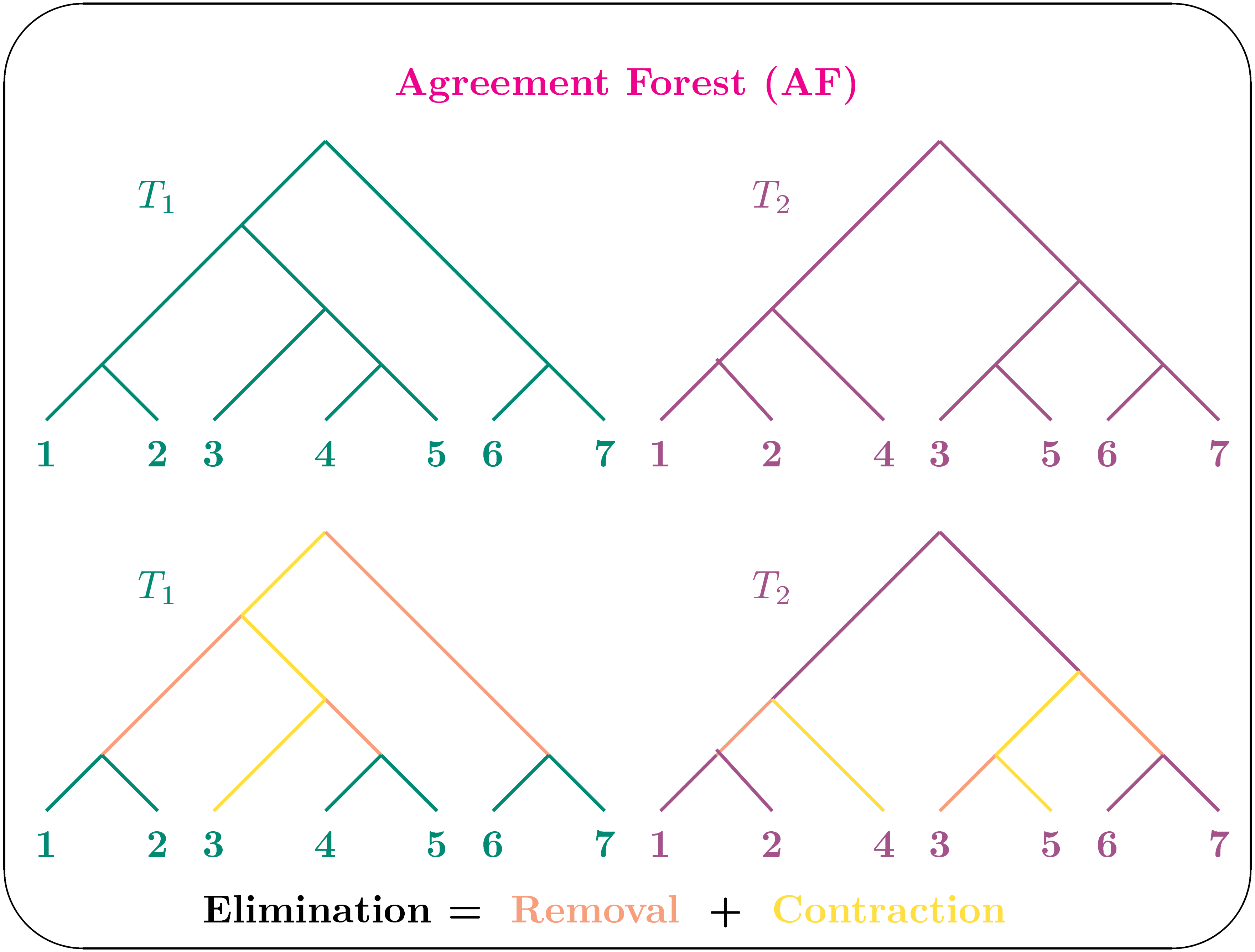


DRUID (M. Zelwer *et al.* – accessible at Pasteur web site)

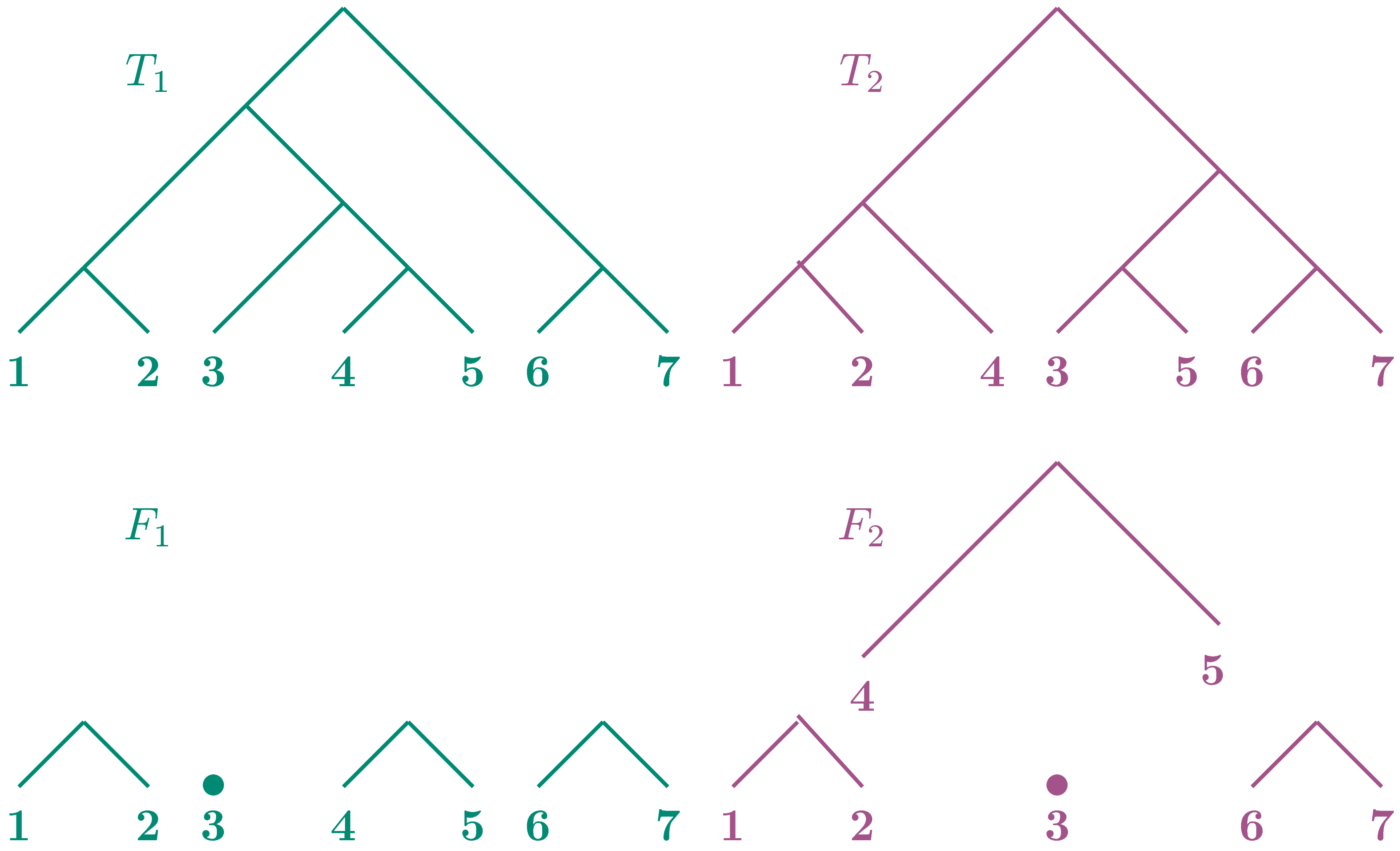
One possibly more appropriate and sophisticated genetic transfer distance between two phylogenetic trees

Maximum Agreement Forest of two trees

MAFALDA (E. M. Rodrigues *et al*, Approx'2001, LNCS, vol. 2129, pages 159-169, Springer-Verlag)

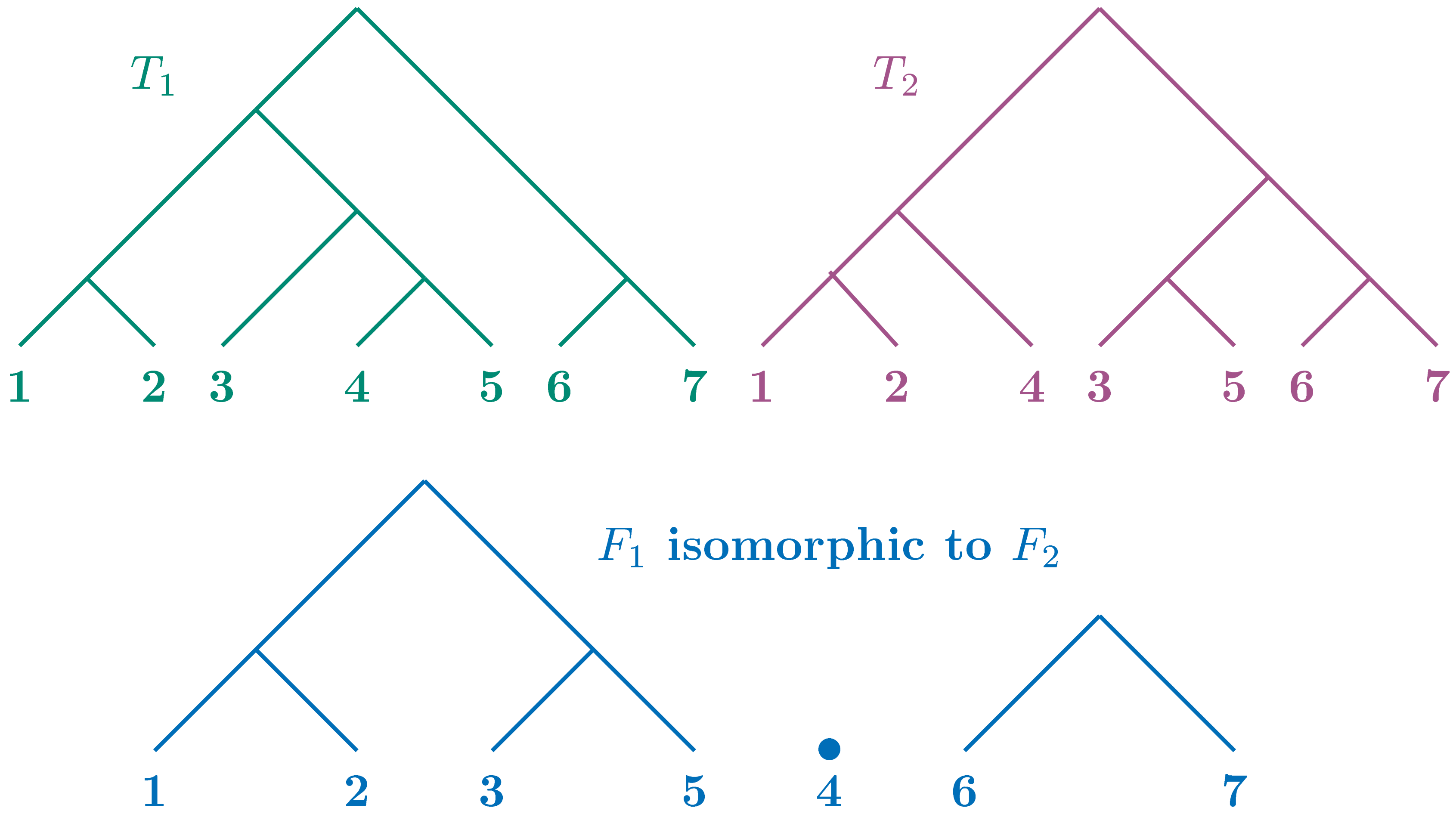

Agreement Forest (AF)
T_1
1 2 3 4 5 6 7
T_2
1 2 4 3 5 6 7
T_1
1 2 3 4 5 6 7
T_2
1 2 4 3 5 6 7
Elimination = Removal + Contraction

Agreement Forest (AF)



size (*i.e.* number of components) = 4

Maximum Agreement Forest (MAF)



size of MAF = 3

Maximum Agreement Forest problem for binary trees

Shown to be NP-Hard by Hein *et al.*, 1996

(Suggested to be MAX-SNP-Hard)

Who also provided a 3-approximation algorithm for the problem

Maximum Agreement Forest problem for binary trees

Shown to be NP-Hard by Hein *et al.*, 1996

(Suggested to be MAX-SNP-Hard)

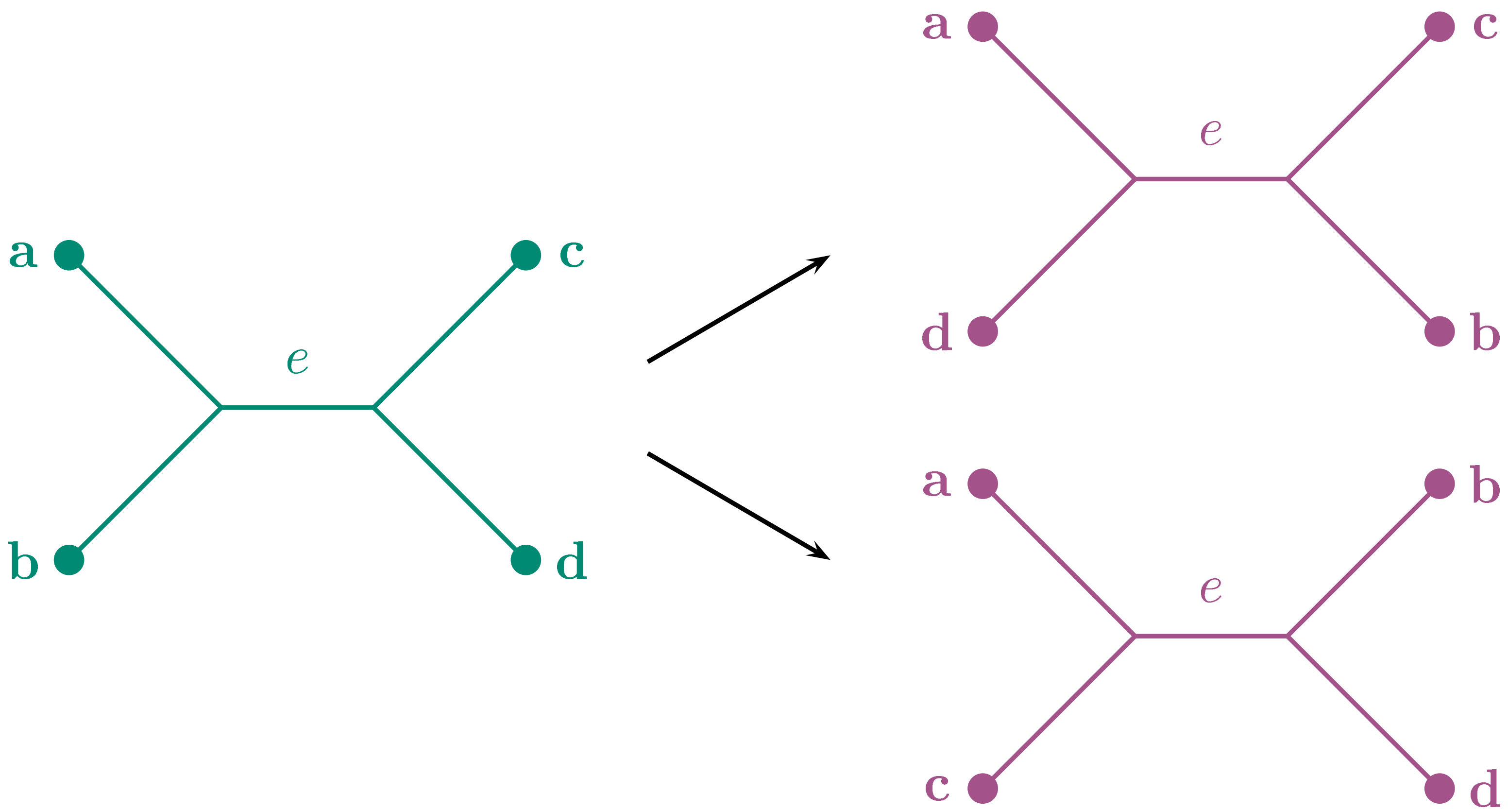
Who also provided a 3-approximation algorithm for the problem

Which, in fact, we show is a 4-approximation

We introduce an algorithm which is a 3-approximation

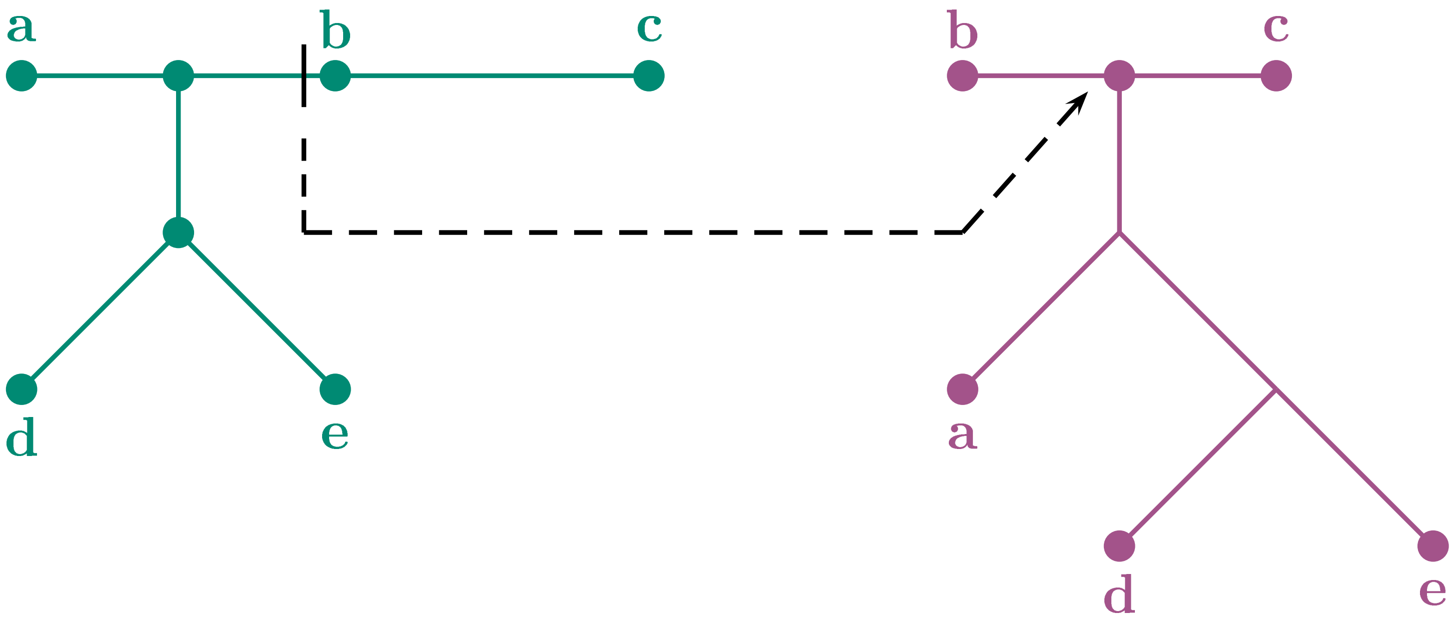
There are other genetic transfer distances between two trees

Nearest Neighbour Interchange (NNI)



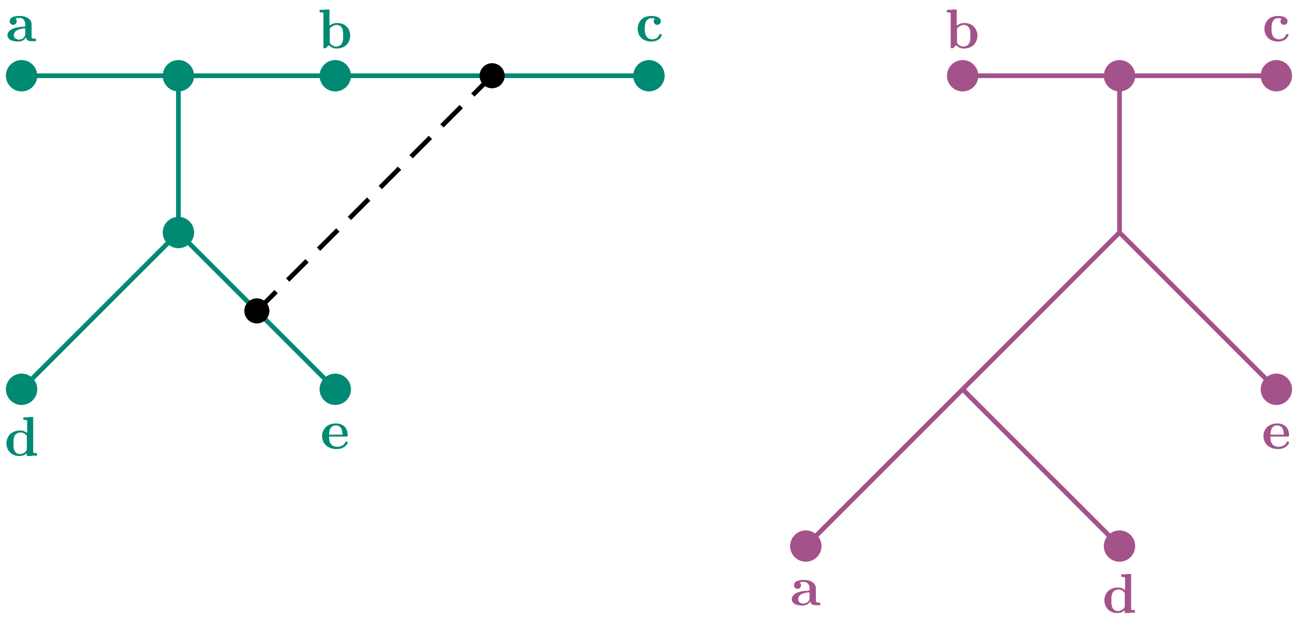
There are other genetic transfer distances (continuation)

Subtree Prune and Regraft (SPR)



There are other genetic transfer distances (continuation)

Tree Bisection and Reconnection (TBR)



a SPR operation is always a TBR but not conversely

Relation between the various distances

For any two binary trees T_1 and T_2 labelled at the leaves:

$$D_{TBR}(T_1, T_2) \leq D_{SPR}(T_1, T_2) \leq D_{NNI}(T_1, T_2)$$

$$D_{SPR}(T_1, T_2) \leq 2 \times D_{TBR}(T_1, T_2)$$

$$MAF(T_1, T_2) - 1 = D_{TBR}(T_1, T_2)$$

Our conjecture: $D_{SPR}(T_1, T_2) = \ MAF(T_1, T_2)$

or $MAF(T_1, T_2) - 1$

What are some of our perspectives ?

Prove the conjecture and improve the bounds

Obtain a better than 3-approximation for the MAF

Explore relation between MAF and MAST

Examine the "fixed parameter tractability" of each distance

> **Ex: When the maximum distance between the two trees is (small) d, can we devise an algorithm whose complexity depends on d and only linearly on the number of leaves ?**

Examine further what is (perhaps ?) the most interesting genetic transfer distance: weighted SPR

The most difficult ? go back to where we came from

Genetic transfer detection and localization as an optimisation problem

The (ideal ?) result

alignement

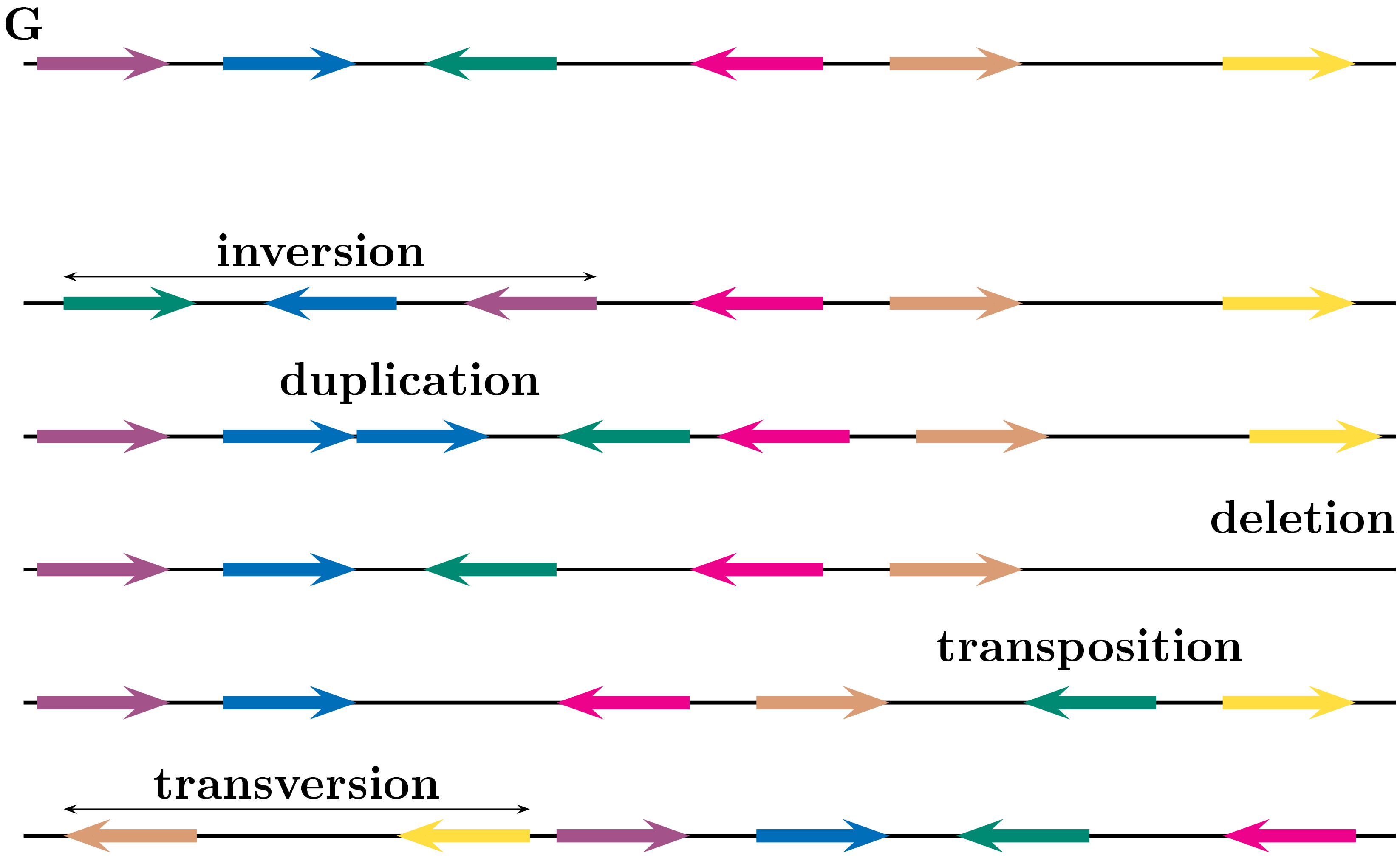
“optimal” partition: sum of lengths of trees + penalty for genetic transfers is **MINIMUM**

Second sub-area of GENOME DYNAMICS:

Calculate a rearrangement distance between genomes (and detect "conserved segments")

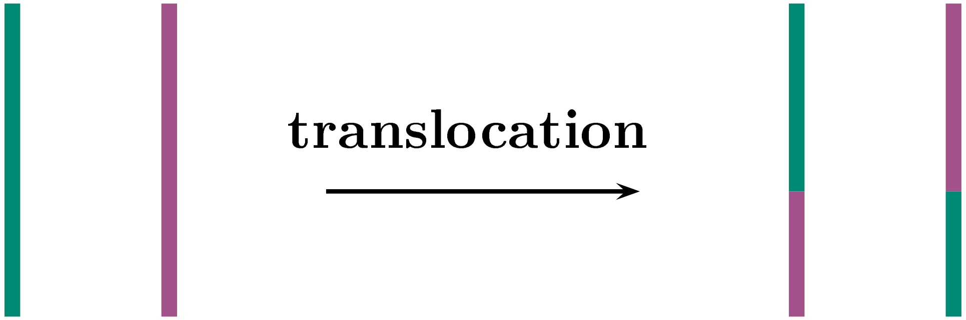
A rapid word on rearrangements

here, “considered unit” = gene



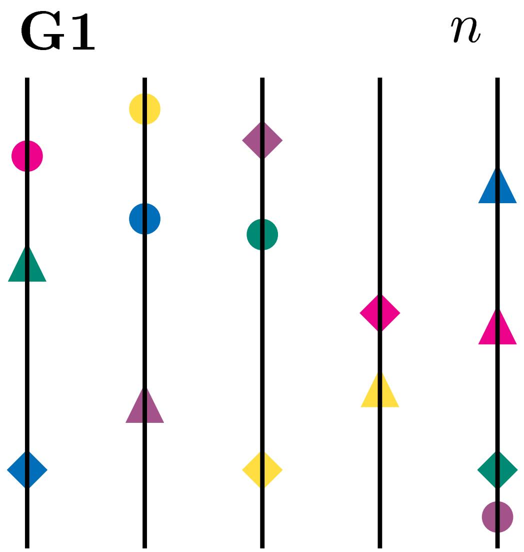
And, in the case of multichromosomal genomes

here, "considered unit" = gene



Rearrangement distance between two genomes

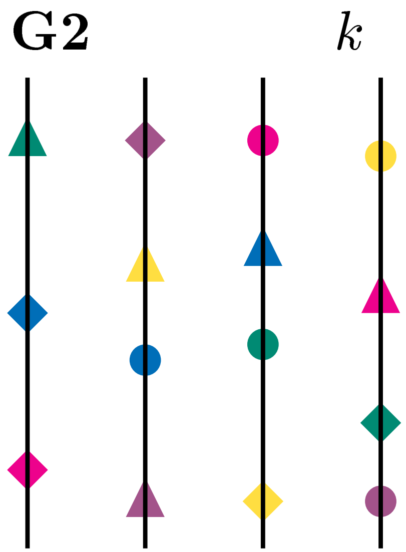
(a very special case of rearrangement distance)



genome

=

set of
bags of genes



distance = minimum number of 3 types of operations
to transform one genome into another

fusion

fission

translocation

(N. Pisanti *et al.*, accepted *Algorithmica*)

What are some of our persectives ?

Biological issues

- consider more realistic distances between genomes
- review and expand models and identification of "conserved segments"
  - *i.e.* segments which are subject to no or to "only local" rearrangements which do not affect function
- address the problem of phylogenetic reconstruction in presence of rearrangements (F. Chataigner, M. Zelwer)

Algorithmical issues

- there is a huge number, some exceedingly beautiful ones
- curious ? $\rightarrow$ come talk to us !!

Fifth area of research:

CELL MODELLING: Putting all this back inside the cell

The future (with S. Provencher)

Recalling the main participants

Inria - Lyon (HELIX)	Various - Paris (Jan. 2002)
Nadia Pisanti	Julien Allal (IGM)i
Sébastien Provencher	Philippe Blayo (IGM)
Marie-France Sagot	Anne Cariou (IGM)
Marina Zelwer	Frédéric Chataigner (LIAFA)
USP - Brazil	Sophie Créno (IGM)
Estela M. Rodrigues	Laurent Marsan (IGM)
	Raquel Tavares (Pasteur)

plus some others soon, in France, Brazil and etc

Mentioning the main collaborators past and present

S. Aubourg	Y. D'Aubenton	J. Berstel
S. Cole	M. Crochemore	P. Darlu
J.-J. Daudin	H. Geiselmann	J. Houmard
C. Iliopoulos	H. de Jong	A. Marcel
C. Mathé	A. Morgat	L. Mouchard
G. Myers	J. Pothier	M. Raffinot
M. Régnier	S. Robin	P. Rouzé
S. Schbath	T. Schiex	H. Soldano
C. Thermes	A. Viari	Y. Wakabayashi

N. Tandeau de Marsac

plus others soon, among whom perhaps some of you !