

A Software Tool for Whole Genome Synteny Comparisons Olivier Gingras<sup>(1)</sup>, Yannick Gingras<sup>(1,2)</sup>, André Levasseur<sup>(1)</sup>, Anne Bergeron<sup>(1)</sup>, Cedric Chauve<sup>(1,2)</sup>

(1) Comparative Genomics Laboratory, (2) LaCIM, UQÀM, Montréal

email: {anne.bergeron,cedric.chauve}@uqam.ca

### Introduction

Genomes can be compared at many different levels of details: sequences, signals, domains, genes, or large synteny blocks such as those computed in [?]. When each marker has one – and only one – occurrence in each of the compared genomes, *common intervals* [?, ?] can be used to identify clusters [?], detect assembly inconsistencies [?], or propose evolution scenarios [?]. In this poster, we present Sequoia, a software tool that computes and displays the PQ-trees of the *strong* common intervals of two or more genomes. Sequoia implements the algorithms described in [?].

#### Definitions

## Sequoia

Sequoia accepts FASTA-like inputs. Multiple sequences are supplied with a comment line followed by a list of identifiers. The identifiers can be strings, numbers, or a combination of both. Positive signs are optional. Duplicate identifiers within a sequence, and identifiers that are missing in some sequences, are ignored.

## **Examples of input sequences**

> Homo sapiens Chromosome X

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

A signed permutation on n elements is a permutation on the set of integers  $\{1, 2, ..., n\}$  in which each element has a sign, positive or negative. An **interval** of a signed permutation is a segment of consecutive elements of the permutation. Two distinct intervals I and J are said to **commute** if either  $I \subset J$ , or  $J \subset I$ , or  $I \cap J = \emptyset$ . If intervals I and J do not commute, they are said to **overlap**.

A common interval of permutations P and Q is a set of one or more integers that is an interval in both P and Q. The singletons and the set  $\{1, 2, ..., n\}$  are always common intervals, and called **trivial** common intervals.

A common interval I of permutation P and Q is a **strong** if it commutes with every common interval of P and Q. Strong intervals are partially ordered by the inclusion relation. The inclusion order defines a tree whose ordered leaves spell out P or Q, and whose root is the whole permutation. A node of the strong intervals tree is:

**linear** if consecutive pairs of its children are common intervals, **prime**, otherwise.

#### **Strong intervals tree**

Example of strong intervals tree obtained by comparing the identity permutation and permutation P = $(1 \ 4 \ 2 \ 5 \ 3 \ 7 \ 8 \ 6 \ 9)$ . Prime and linear nodes are distinguished by their shape. There are two non-trivial linear nodes, the rectangular nodes: (7,8) is increasing and (7,8,6) is decreasing. There is only one prime node, the round node (4,2,3,5).



> Rattus norvegicus Chromosome X

-13 -4 5 -6 -12 -8 -7 2 1 -3 9 10 11 14 -15 16

Any set of identifiers can be used, such as pairs of orthologs in the following example.

#### > E. coli Genome, paired with S. typhimurium orthologs

yaaJ-yaaJ talB-talB dnaK-dnaK dnaJ-dnaJ nhaR-STM0014 ribF-fruR fkpB-slpA rihC-rihC caiE-caiE caiD-caiD caiC-caiC caiA-caiA caiT-caiT fixA-fixA fixB-fixB fixC-fixC fixX-fixX yaaU-yaaU yabF-yabF kefC-kefC pdxA-pdxA rluA-rluA araD-araD araC-araC yabI-yabI yabJ-yabJ tbpA-phnS yabN-yabN setA-STM0212 leuD-leuD leuC-leuC leuB-leuB leuO-leuO ilvI-ilvI fruR-STM0257 ftsI-ftsI murE-murE murF-murF murD-murD ftsW-ftsW murC-murC ddlB-ddlB mutT-STM0030 hofC-hofC ...

#### **Examples of output trees**

The following tree is a sub-tree of the strong intervals tree obtained by comparing 391 synteny blocks [?] obtained by the whole comparison of the human, mouse and rat genomes. Each block identifier is followed by its chromosomal location in each of the three species. It reveals interesting rearrangments of human chromosome 17, mouse chromosome 11, and rat chromosome 10.



#### ne

### Examples



Human =12345678910111213141516Rat = $\overline{13}$  $\overline{4}$ 5 $\overline{6}$  $\overline{12}$  $\overline{8}$  $\overline{7}$ 21 $\overline{3}$ 91011141516

**Comparing the Mouse and Rat Chromosomes X.** 

The following tree is a sub-tree of the strong intervals tree obtained by comparing 3000 pairs of orthologs *E. coli* and *S. typhimurium*. Sequoia can easily handle thousands of identifiers over dozens of species.



## URL : cgl.bioinfo.uqam.ca

#### References

- [1] Bérard, S., Bergeron, A., Chauve, C., Paul, C., Perfect sorting by reversals is not always difficult. To appear in WABI 2005 Proceedings, LNCS.
- [2] S. Bérard, A. Bergeron and C. Chauve. Conserved structures in evolution scenarios. In *Comparative Genomics, RECOMB 2004 International Workshop, RCG2004*, volume 3388 of *Lecture Notes in Bioin-formatics*, pages 1–15, Springer-Verlag, 2005.
- [3] Bergeron, A., Chauve, C., De Montgolfier, F., Raffinot, M., Computing common intervals of *K* permutations, with applications to modular decomposition of graphs. To appear in ESA 2005 Proceedings, LNCS.



# $Mouse = 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9 \quad 10 \quad 11 \quad 12 \quad 13 \quad 14 \quad 15 \quad 16$ $Rat = \overline{4} \quad \overline{3} \quad \overline{2} \quad 1 \quad \overline{13} \quad \overline{15} \quad 14 \quad \overline{16} \quad 8 \quad 9 \quad 10 \quad \overline{11} \quad 12 \quad 5 \quad 6 \quad 7$

- [4] G. Bourque, P. A. Pevzner and G. Tesler. Reconstructing the genomic architecture of ancestral mammals: Lessons from human, mouse, and rat genomes. *Genome Res.*, 14(4):507–516, 2004.
- [5] S. Heber and J. Stoye. Finding all common intervals of *k* permutations. In *Combinatorial Pattern Matching, 12th Annual Symposium, CPM 2001*, volume 2089 of *Lecture Notes in Comput. Sci.*, pages 207–218, Springer-Verlag, 2001.
- [6] G.M. Landau, L. Parida and O. Weimann. Using PQ trees for comparative genomics. To appear in *Combinatorial Pattern Matching, 16th Annual Symposium, CPM 2005*, 2005.
- [7] T. Uno and M. Yagiura. Fast algorithms to enumerate all common intervals of two permutations. *Algorithmica*, 26(2):290–309, 2000.