

Whole genome alignment in High Performance Computing environments

Julio César García Vizcaíno Directores: Antonio Espinosa, Juan Carlos Moure



Computer Architecture & Operating Systems Department
Universitat Autònoma de Barcelona

18 de enero de 2013

Contents

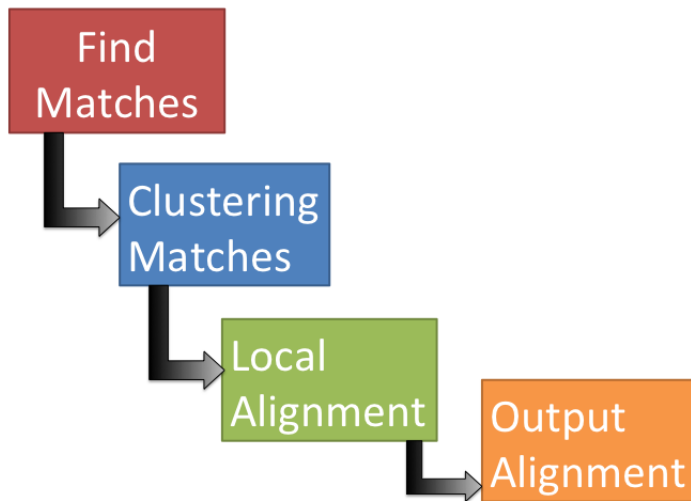
1 *Problem definition*

2 *Objectives*

3 *Parallel search of maximal unique matches*

4 *Conclusions*

Whole Genome Alignment in MUMmer



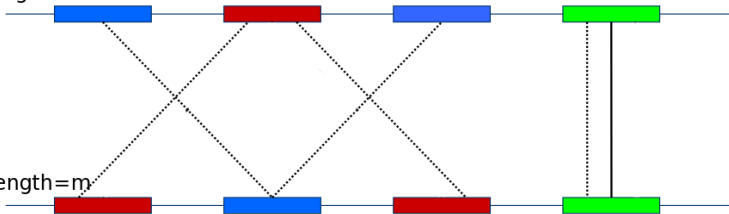
Search of Maximal Unique Matches

MUM: Maximal Unique Match

MEM: Maximal Exact Match

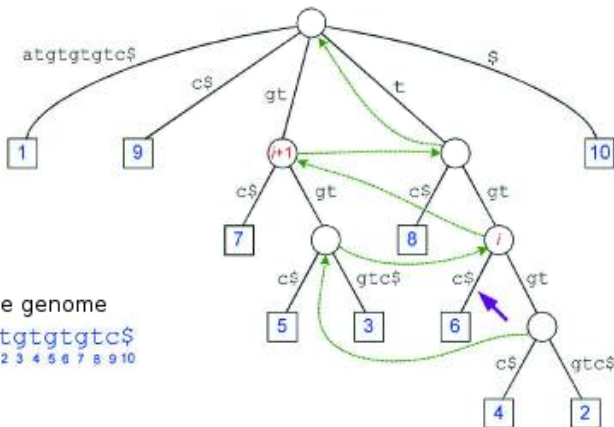
R length = n

Q length = m



Traversal of suffix tree

Query genometgtcc...



Suffix tree of reference genome

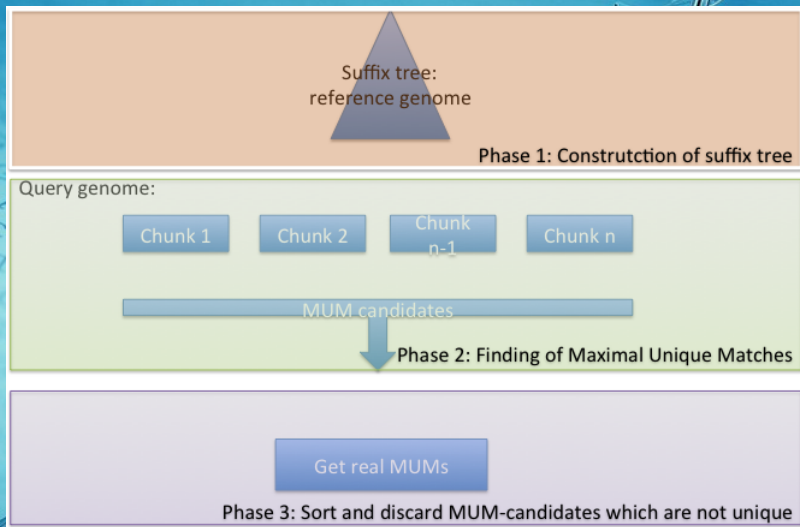
atgtgtgttc\$
1 2 3 4 5 6 7 8 9 10

General objective

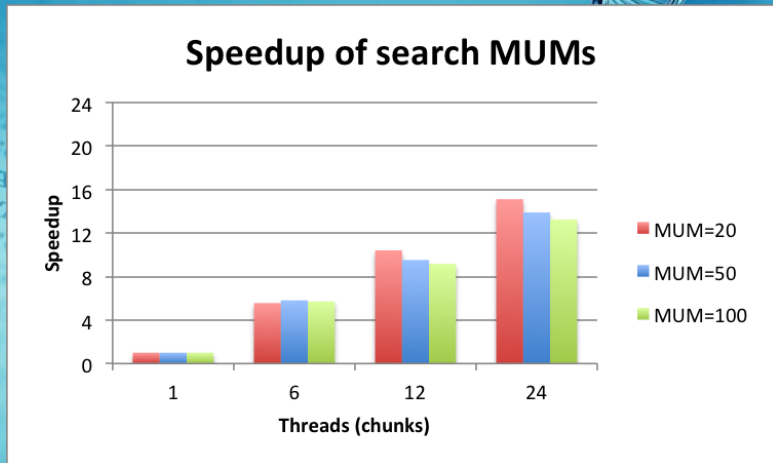
General objective

Speed up the search of exact matches (distributed) considering the use of computer and memory resources; and adapt it to application MUMmer for its execution in HPC cluster multicore environments.

Search of MUMs in Multicore architectures



Parallel search of maximal unique matches



Conclusions

Current work

- Evaluation of performance to search MUMs of a query and reference genome in multi-core architectures with OpenMP.
- Results shows that the heaviest section of searching MUMs in a suffix tree is improved with the use of a multi-core architecture.
- Bottleneck is in suffix tree: traverse a suffix tree in multi-core architectures.

Future work

Implement Distributed Suffix Tree.

Perform massive searches of maximal matches: design and test a parallel and distributed algorithm to perform the search of maximal matches in Distributed Suffix Tree for HPC multicore environments.



Thanks!

Whole genome alignment in High Performance Computing environments

Julio César García Vizcaíno Directores: Antonio Espinosa, Juan Carlos Moure



Computer Architecture & Operating Systems Department
Universitat Autònoma de Barcelona

18 de enero de 2013