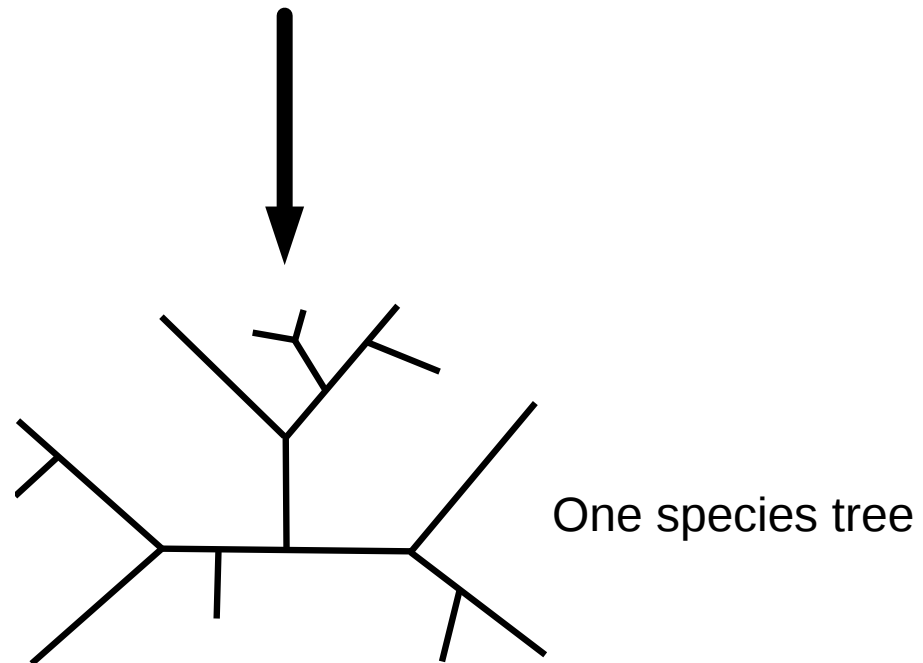
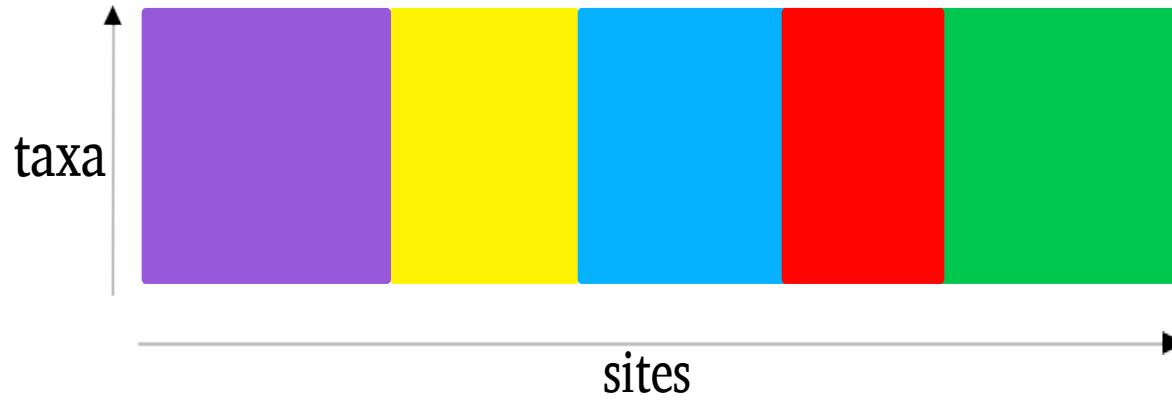


ParGenes

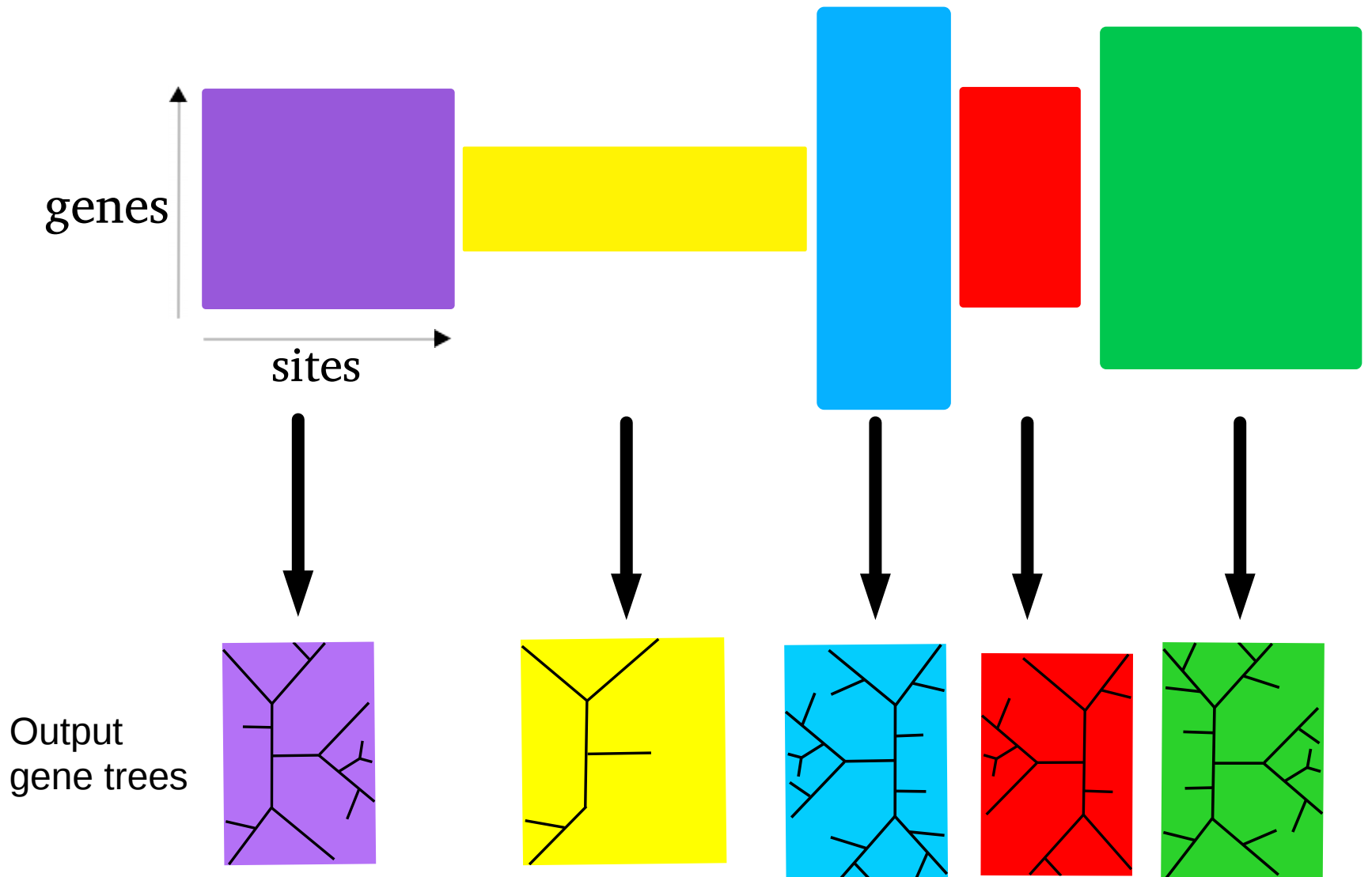
An integrated tool for model selection and maximum likelihood (ML) based phylogenetic inference on thousands of independent MSAs on clusters and supercomputers

Benoit Morel, Alexey Kozlov, Alexandros Stamatakis

Tree inference: Concatenated Methods

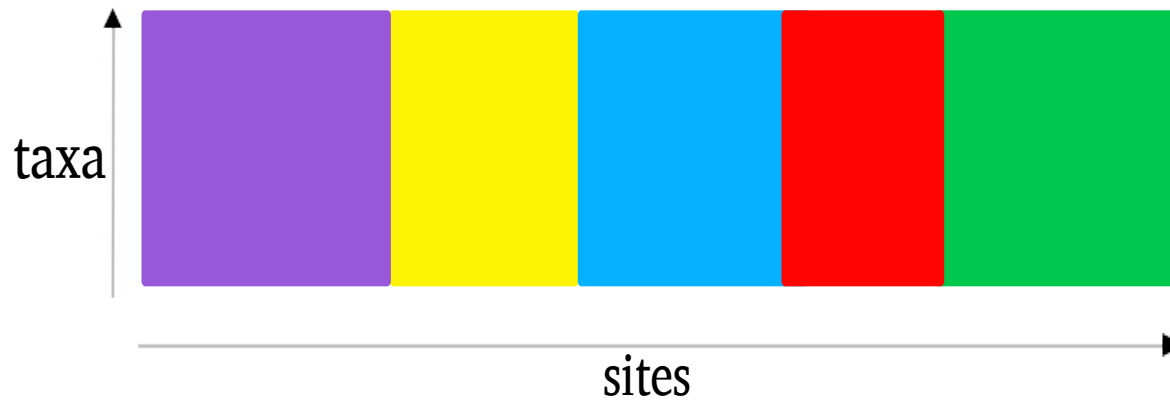


Gene trees methods



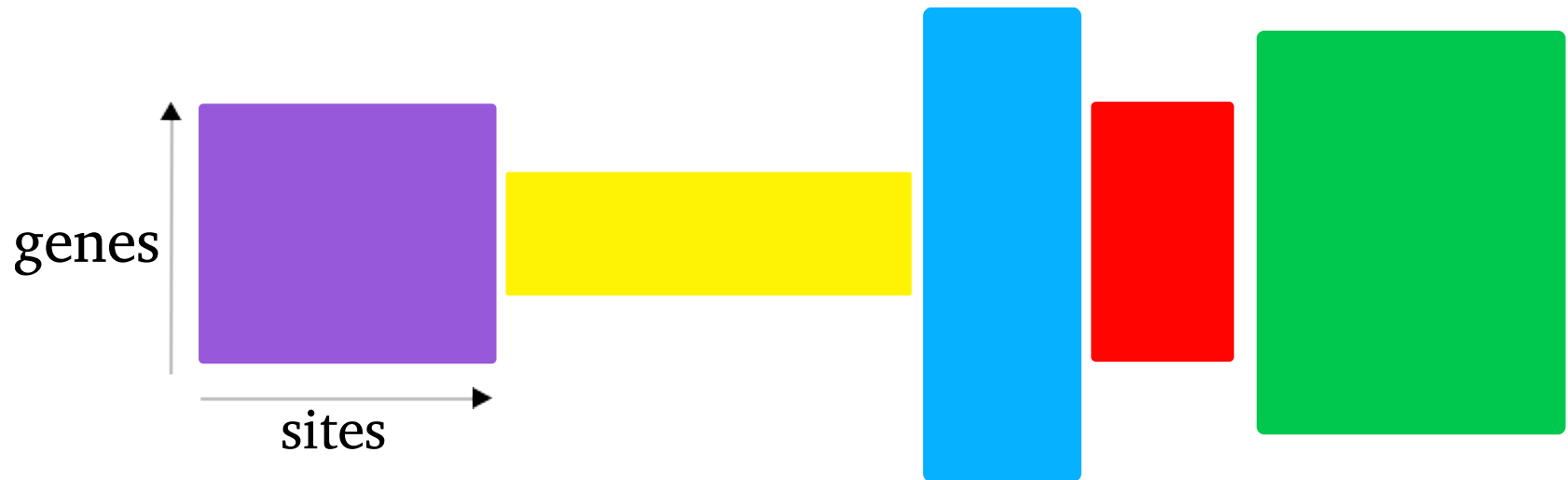
Existing tools

Existing ML inference tools are good at dealing with one single MSA...



Existing tools

... but they do not handle thousands of inputs of varying dimensions!

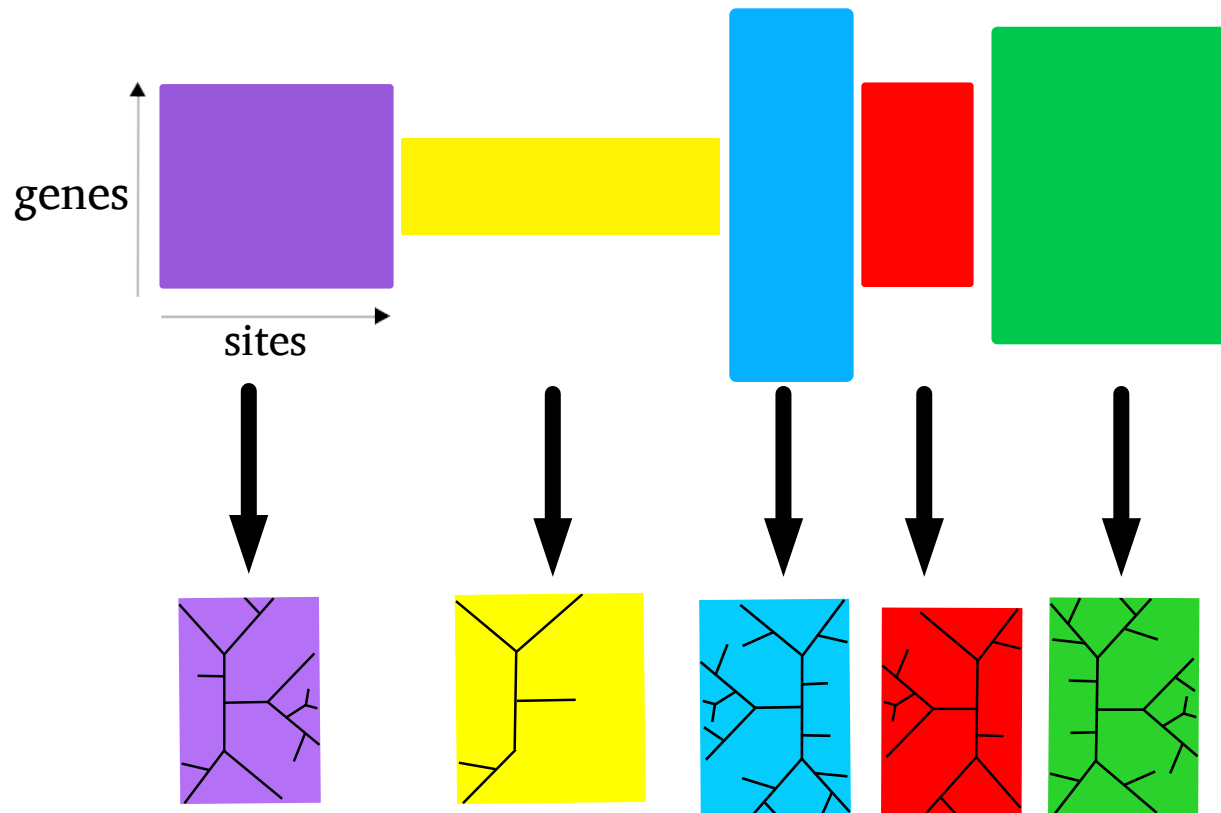


Building thousands of gene trees

- **Users have to write home-made scripts:** frustrating and time consuming
- **These scripts have to run on supercomputers.** Or users have to wait for a very long time.



Infer ML gene trees from independent MSAs on
supercomputers.



ParGenes interface

Folder containing
the alignments
-a msa_dir

Number of
Cores
-c 256

```
python pargenes.py -a msa_dir -o output_dir -m -s 10 -p 20 -b 100 -c 256
```


ParGenes interface

Folder containing
the alignments
-a msa_dir

Number of
Cores
-c 256

Apply model
selection?
-m

```
python pargenes.py -a msa_dir -o output_dir -m -s 10 -p 20 -b 100 -c 256
```

ParGenes interface

Folder containing
the alignments
-a msa_dir

Number of
Cores
-c 256

Apply model
selection?
-m

Number of
Starting trees
-p 20 -s 10

Number of BS
Replicates
-b 100

```
python pargenes.py -a msa_dir -o output_dir -m -s 10 -p 20 -b 100 -c 256
```

Additional features

- Checkpointing
- Invalid MSAs filtering
- Optimal number of cores evaluation

ParGenes:

- is a Python script
- relies on RaxML-NG [1] and ModelTest-NG [2]
- dynamically schedules jobs with MPI

[1] <https://github.com/amkozlov/raxml-ng/>

[2] <https://github.com/ddarriba/modeltest>

Thank you!

Heidelberg Institute for
Theoretical Studies



Available at <https://github.com/BenoitMorel/ParGenes>

Method and results: <https://doi.org/10.1101/373449>

User-support on the raxml google group

Suggestions are welcome!

Questions?