

Raxml repeats benchmarks

April 4, 2017

1 Execution time

We run raxml for each dataset with and without the repeats option, and compare the execution times.

Speedup is calculated as follow :

$$speedup = \frac{time_{tipinner}}{time_{repeats}}$$

When we don't indicate times, the cluster cancelled the jobs before the end (after 24h). We then compare the speedup at the last reached step.

Raxml commands used:

```
$ ./raxml-ng-mpi --seed=42 --msa data.phy --model data.part  
--simd AVX --threads 16 --redo --repeats on  
$ ./raxml-ng-mpi --seed=42 --msa data.phy --model data.part  
--simd AVX --threads 16 --redo
```

dataset	taxas	sites	partitions	type	<i>time_{repeats}</i>	<i>time_{tipinner}</i>	speedup
404	404	7444	11	DNA	766s	1131s	1.47
1kite.science2013	144	371434	50	DNA	11400s	20414s	1.8
1kite.hyme	174	2248590	4116	DNA	cancelled	cancelled	1.53
Antl.1.1.nt	40	522173	658	DNA	1402s	2293s	1.63
Antl.1.1.aa	40	762438	659	prot	10161s	14500s	1.4
para.1.nt	193	1514275	3714	DNA	cancelled	cancelled	1.81

2 Load balancing

We define

$$busy\ ratio = (1 - \frac{waiting\ time\ in\ reduce\ operations}{elapsed\ time}) * 100$$

It directly gives the speedup we could get from a perfect load balancing.

dataset	taxas	sites	partitions	type	busy ratio repeats	busy ratio tipinner
1kite_science2013	144	371434	50	DNA	83%	83%
1kite_hyme	174	2248590	4116	DNA	93%	92%
Antl_1_1_aa	40	762438	659	prot	96%	89%
para_1_nt	193	1514275	3714	DNA	92%	90%