

Raxml repeats benchmarks

April 6, 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 (repeats, repeats with jemalloc, tipinner):

```
$ LD_PRELOAD=libjemalloc.so mpirun ./raxml-ng-mpi --seed=42 --msa data.ph  
--simd AVX --threads 1 --search --repeats on  
$ LD_PRELOAD=libjemalloc.so mpirun ./raxml-ng-mpi --seed=42 --msa data.ph  
--simd AVX --threads 1 --search
```

dataset	taxas	sites	partitions	type	<i>time_{repeats}</i>	<i>time_{tipinner}</i>	speedup
404	404	7444	11	DNA	457s	755s	1.65
1kite.science2013	144	371434	50	DNA	?	?	?
1kite.hyme	174	2248590	4116	DNA	?	?	?
Antl.1.1_nt	40	522173	658	DNA	1254s	1668s	1.3
Antl.1.1_aa	40	762438	659	prot	?	?	?
para.1_nt	193	1514275	3714	DNA	?	?	?

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.

I will recompute this table later with mpi.