

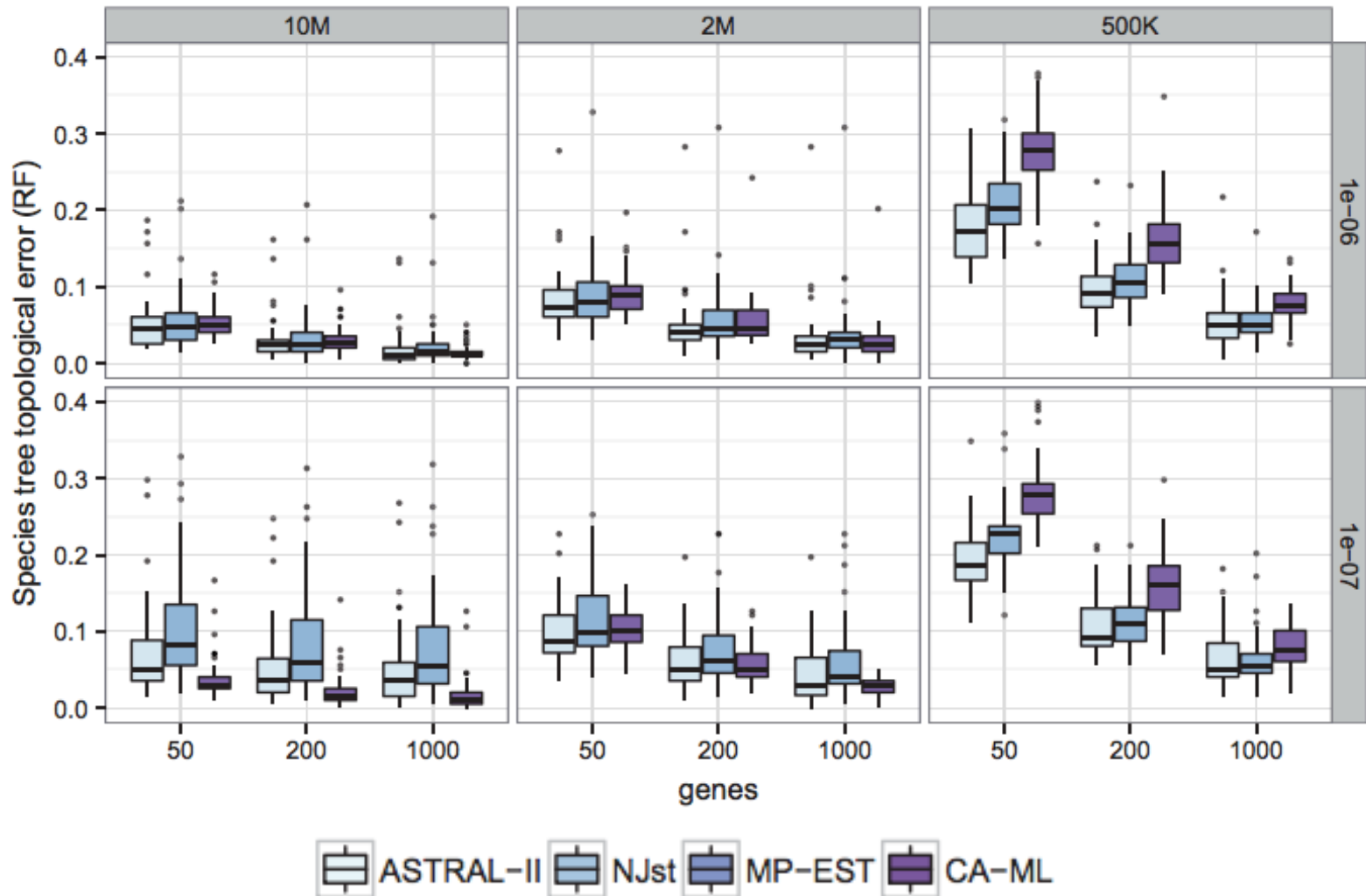
ASTRAL-II: coalescent-based species
tree estimation with many hundreds
of taxa and thousands of genes
Mirarab & Warnow, 2015

Review by Danielle Campbell

ASTRAL-II

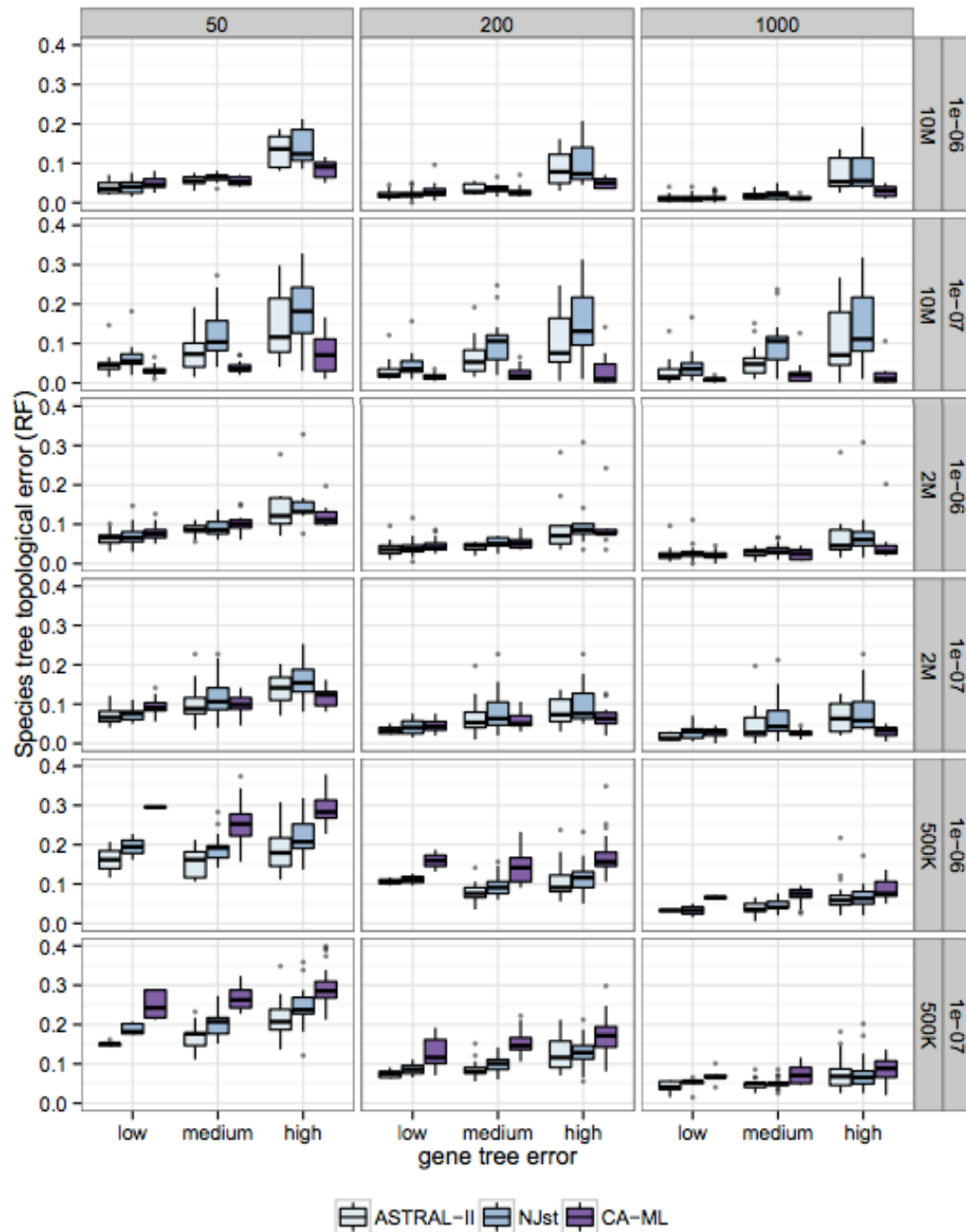
- A fast, coalescent-based method for species tree estimation on large datasets
 - Statistically consistent under the multispecies coalescent
 - Runs in polynomial time
- Takes gene trees as input
 - Uses a constrained search space defined by the bipartitions of the input gene trees

ASTRAL-II is more accurate than NJst



rate	height	genes	ASTRAL-II	NJst	CA-ML
1e-06	10M	50	5.2±0.5	5.6±0.6	5.4±0.3
1e-06	10M	200	3.1±0.4	3.4±0.5	3.1±0.3
1e-06	10M	1000	2.0±0.4	2.3±0.5	1.4±0.2
1e-06	2M	50	8.4±0.6	9.1±0.7	9.2±0.4
1e-06	2M	200	5.0±0.6	5.6±0.6	5.5±0.5
1e-06	2M	1000	3.4±0.6	3.9±0.6	2.8±0.4
1e-06	500K	50	17.6±0.7	20.9±0.7	27.9±0.7
1e-06	500K	200	9.6±0.5	11.0±0.5	16.2±0.7
1e-06	500K	1000	5.3±0.5	5.7±0.4	8.0±0.3
1e-07	10M	50	7.3±0.9	10.2±1.0	4.0±0.4
1e-07	10M	200	5.4±0.7	8.2±1.0	2.2±0.3
1e-07	10M	1000	5.0±0.8	8.0±1.0	1.8±0.3
1e-07	2M	50	10.2±0.6	11.7±0.7	10.3±0.3
1e-07	2M	200	6.0±0.5	7.5±0.7	5.7±0.3
1e-07	2M	1000	4.4±0.6	6.0±0.7	2.8±0.2
1e-07	500K	50	19.3±0.7	22.5±0.6	28.2±0.6
1e-07	500K	200	10.7±0.6	11.4±0.5	16.1±0.7
1e-07	500K	1000	6.3±0.5	6.3±0.5	8.0±0.4

- What causes CA-ML to sometimes be more accurate than ASTRAL-II?
- Gene tree error
- An artifact of using FastTree?



Summary & Critique

- ASTRAL-II is most advantageous over other coalescent-based methods when ILS is greatest (i.e. short tree lengths) and when speciation events are ancient
- ASTRAL-II is most advantageous over CA-ML when gene tree error is low
- There is now way to estimate gene tree error on real datasets; how can you gauge which method is better?