

Supplementary Data for “Inferring Species Trees
Directly from Biallelic Genetic Markers:
Bypassing Gene Trees in a Full Coalescent
Analysis”

David Bryant

Department of Mathematics and Statistics, University of Otago, Dunedin, New Zealand, and the Allan Wilson Centre for Molecular Ecology and Evolution

Remco Bouckaert

Computational Evolution Group, University of Auckland, Auckland, New Zealand

Joseph Felsenstein

Department of Genome Sciences and Department of Biology, University of Washington, Box 355065, Seattle, WA 98195-5065

Noah A. Rosenberg

Department of Biology, Stanford University, Stanford, California, USA.

Arindam RoyChoudhury

Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, USA.

Corresponding Author: David Bryant, Department of Mathematics and Statistics, University of Otago, P.O. Box 56, Dunedin, New Zealand.
ph: +64 3 4797889. fax +64 3 479 8427.
email: david.bryant@otago.ac.nz

Here we report results of the simulations when dominant markers were generated and analysed using SNAPP, with and without the correction for dominance.

tree	4 taxa								8 taxa							
	Easy				Hard				Easy				Hard			
	c	i	c	i	c	i	c	i	c	i	c	i	c	i	c	i
100	1	1	1	1	3	2	2	2	1	1	1	1	6	3*	6	3*
200	1	1	1	1	3	3	3	3	1	1	1	1	15	6	16	6
300	1	1	1	1	3	3	3	3	1	1	1	1	36	34	37	34
400	1	1	1	1	3	3	3	3	1	1	1	1	13	7	14	10
500	1	1	1	1	3	3	3	3	1	1	1	1	7	5*	7	7
600	1	1	1	1	2	1	2	2	1	1	1	1	3	3	3	3
700	1	1	1	1	3	3	3	3	1	1	1	1	9	8	9	8
800	1	1	1	1	1	1	1	1	1	1	1	1	8	6	8	6
900	1	1	1	1	1	1	1	1	1	1	1	1	6	7*	6	7*
1000	1	1	1	1	3	3	3	3	1	1	1	1	8	7	8	6
10000	1	1	1	1	1	1	1	1	1	1	1	1	6	4	6	5
100000	1	1	1	1	1	1	1	1	1	1	1	1	2	3	2	2
1000000	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table S1: The size of the credibility sets in the first simulation when dominant markers were simulated but SNAPP was used *without* the dominant marker correction. Here, ‘tree’ indicates which of the trees in Fig. ?? was used to generate data. ‘c’ and ‘i’ indicate whether ‘correct’ or ‘incorrect’ priors were used on the θ values and on the speciation rate. Numbers 100 to 1000000 indicate the number of polymorphic sites generated. Values in the table are the numbers of trees in the credibility set. The seven instances where the true tree was not contained within this set are marked by an asterisk (*).

tree	4 taxa								8 taxa							
	Easy				Hard				Easy				Hard			
	c	i	c	i	c	i	c	i	c	i	c	i	c	i	c	i
θ -prior	c	i	c	i	c	i	c	i	c	i	c	i	c	i	c	i
t -prior	c	i	c	i	c	i	c	i	c	i	c	i	c	i	c	i
100	1	1	1	1	3	3	3	3	2	1	1	1	12	3	17	14
200	1	1	1	1	1	1	2	1	1	1	1	1	3	3	3	2
300	1	1	1	1	3	3	3	3	1	1	1	1	5	3	5	3
400	1	1	1	1	3	2	3	3	1	1	1	1	7	3	7	7
500	1	1	1	1	3	3	3	3	1	1	1	1	3*	3	3*	2*
600	1	1	1	1	3	3	3	3	1	1	1	1	6	3	5	5
700	1	1	1	1	3	3	3	3	1	1	1	1	4	2	4	6
800	1	1	1	1	1	1	1	1	1	1	1	1	7	3	7	6
900	1	1	1	1	3	2	3	3	1	1	1	1	7	1	7	6*
1000	1	1	1	1	3	2	3	2	1	1	1	1	6	2	6	4
10000	1	1	1	1	1	1	1	1	1	1	1	1	7	1	7	7
100000	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	2
1000000	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1

Table S2: The size of the credibility sets in the first simulation when dominant markers were simulated and SNAPP was used *with* the dominant marker correction. Here, ‘tree’ indicates which of the trees in Fig. ?? was used to generate data. ‘c’ and ‘i’ indicate whether ‘correct’ or ‘incorrect’ priors were used on the θ values and on the speciation rate. Numbers 100 to 1000000 indicate the number of polymorphic sites generated. Values in the table are the numbers of trees in the credibility set. The seven instances where the true tree was not contained within this set are marked by an asterisk (*). Note the similarity with the results in Table S1.