Example: Microsatellite data set

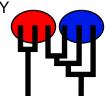
POPULATION SIZE, MIGRATION, DIVERGENCE, ASSIGNMENT, HISTORY

Bayesian inference using the structured coalescent

Migrate-n version 5.0.7 [May-01-2025]

Program started at Sat May 3 12:24:21 2025

Program finished at Sat May 3 12:24:21 2025 [Runtime:0000:00:00:00]



Options

Datatype: Microsatellite data [Brownian motion]
Missing data: not included

Inheritance multipliers in use for Thetas:

All loci use an inheritance multiplier of 1.0

Data set was subsampled: used a random sample of size:

5 and seed 13

Random number seed: (from parmfile) 310705631

Start parameters:

Theta values were generated Using a percent value of the prior

M values were generated Using a percent value of the prior

Connection matrix:

m = average (average over a group of Thetas or M,

s = symmetric migration M, S = symmetric 4Nm,

0 = zero, and not estimated,

* = migration free to vary, Thetas are on diagonal

d = row population split off column population, D = split and then migration

Population 1 2 1 population_numb * 0 2 population_numb D *

1 population_numb growing/shrinking

						•		
2 population_numb	o co	onstant size						
Order of paramete	rs:							
-	Θ_1		<	displayed>				
	$\Theta_2^{^1}$			displayed>				
	$M^{2}_{1->2}$		<	displayed>				
4	$\Lambda_{1\rightarrow 2}$		<	displayed>				
5 (5 _{1->2}		<	<displayed></displayed>				
6 (9 1		<	displayed>				
Mutation rate amor	na loci:				ľ	Mutation rate i	s consta	nt for all loci
	.5							
Analysis strategy:							Bayesi	an inference
-Population size e	stimatio	on:				Ехр	-	l Distribution
-Population growth	n estim	ation: Grow	th [Exponential]					
-Geneflow estimat	ion:					Exp	onentia	l Distribution
-Divergence time	estimat	ion:			Normal Dist	tribution (mea	n and sta	andard dev.)
Proposal distribution	ns for	parameter						
Parameter			Proposa					
Theta			Slice sampling	I				
				_				
M		Mat	Slice sampling					
Divergence Divergence Spread			ropolis sampling ropolis sampling					
Growth			ropolis sampling					
Genealogy			ropolis sampling					
Conculogy		Wiet	ropolio riadiliige	,				
Prior distribution fo	r paran	neter						
Parameter	•		ior Minimum	Mean*	Maximum	Delta	Bins	UpdateFreq
1 Theta	* *	Uniform	0.000000	10.00	20.00	2.000000	1500	0.08333
2 Theta	* *	Uniform	0.000000	10.00	20.00	2.000000	1500	0.08333
3 M	* *	Uniform	0.000000	10.00	20.00	2.000000	1500	0.08333
4 Splittime mean	* *	Uniform	0.000000	10.00	20.00	2.000000	1500	0.08333
5 Splittime std	* *	Uniform	0.000000	10.00	20.00	2.000000	1500	0.08333
6 Growth	* *	Uniform	-100.00000	0.000	100.0	20.000000	1500	0.08333
[* * means priors w	ere set	globally]						

Posterior distribution:
Parameter values were

Parameter values were collected using MCMC, these values

were then used to generate the posterior histograms using KERNEL SMOOTHING (window=41)

and subsequent SAVITZKY-GOLAY SMOOTHING (window=41) for combination over loci

Markov chain settings:Long chainNumber of chains1Recorded steps [a]200Increment (record every x step [b]1Number of concurrent chains (replicates) [c]2Visited (sampled) parameter values [a*b*c]400Number of discard trees per replicate (burn-in * b)50

Multiple Markov chains:

Static heating scheme 4 chains with temperatures

1000000.00 3.00 1.50 1.00

Swapping interval is 1

Print options:

Data file: infile.msat

parmfile.msat

Haplotyping is turned on:

Output file: outfile_msat

Posterior distribution raw histogram file: bayesfile_msat

Raw data from the MCMC run: bayesallfile_msat.gz

Print data: No

Print genealogies [only some for some data type]:

Data summary

Data file:

Datatype:

Microsatellite data [Brownian]

[Data was used as repeat-length information]

Number of loci: 10

Mutationmodel:

Locus Sublocus	Mutationmodel	Mutationmodel parameters
1 1	Brownian Motion	[none]
2 1	Brownian Motion	[none]
3 1	Brownian Motion	[none]
4 1	Brownian Motion	[none]
5 1	Brownian Motion	[none]
6 1	Brownian Motion	[none]
7 1	Brownian Motion	[none]
8 1	Brownian Motion	[none]
9 1	Brownian Motion	[none]
10 1	Brownian Motion	[none]

Data set was subsampled: used a random sample of size:

5

Population	Locus	Gene co	pies
		data	(missing)
1 population_number0	1	50	(0)
	2	50	(0)
	3	50	(0)
	4	50	(0)
	5	50	(0)
	6	50	(0)
	7	50	(0)
	8	50	(0)
	9	50	(0)
	10	50	(0)
2 population_number1	1	42	(0)
	2	42	(0)
	3	42	(0)
	4	42	(0)
	5	42	(0)
	6	42	(0)
	7	42	(0)

	E	cample: Microsatel	iite data set -
	8	42	(0)
	9	42	(0)
	10	42	(0)
Total of all populations	1	92	(0)
	2	92	(0)
	3	92	(0)
	4	92	(0)
	5	92	(0)
	6	92	(0)
	7	92	(0)
	8	92	(0)
	9	92	(0)
	10	92	(0)
			, ,

Subsampled dataset

Data set was subsampled randomly per population: 5 samples taken

Locus Population Individuals

1	population	0BAC_0BBU 0BBQ_0BBD 0BBO_0BBJ 0BAV_0BAO 0BAX_0BCA
	population	1BBN_1BAX 1BAQ_1BAN 1BAL_1BAV 1BBK_1BAS 1BAP_1BBP
2	population	0BAF_0BBV 0BAL_0BBL 0BAK_0BAD 0BBA_0BAI 0BAC_0BBU
	population	1BAH_1BAT 1BAE_1BAI 1BBF_1BAO 1BBM_1BAM 1BBJ_1BBC
3	population	0BAK_0BAD 0BBK_0BAH 0BBR_0BAW 0BBF_0BAE 0BAN_0BAT
	population	1BBE_1BAJ 1BAL_1BAV 1BAE_1BAI 1BBM_1BAM 1BBD_1BAG
4	population	0BBB_0BBX 0BAU_0BAP 0BAV_0BAO 0BAS_0BBT 0BBF_0BAE
	population	1BAE_1BAI 1BAQ_1BAN 1BBD_1BAG 1BAL_1BAV 1BBF_1BAO
5	population	0BAJ_0BAB 0BAU_0BAP 0BBF_0BAE 0BAC_0BBU 0BAG_0BBI
	population	1BBB_1BBR 1BAC_1BAB 1BAR_1BBA 1BBN_1BAX 1BBK_1BAS
6	population	0BAL_0BBL 0BCB_0BAM 0BAR_0BBW 0BBA_0BAI 0BAJ_0BAB
	population	1BBH_1BBO 1BAR_1BBA 1BBQ_1BBI 1BBL_1BAK 1BBG_1BAU
7	population	0BAU_0BAP 0BAV_0BAO 0BAG_0BBI 0BBF_0BAE 0BBB_0BBX
	population	1BBK_1BAS 1BBQ_1BBI 1BAA_1BAF 1BAE_1BAI 1BBL_1BAK
8	population	0BAF_0BBV 0BBO_0BBJ 0BBP_0BBC 0BCB_0BAM 0BAJ_0BAB
	population	1BBE_1BAJ 1BBH_1BBO 1BAP_1BBP 1BAD_1BAW 1BBL_1BAK
9	population	0BAU_0BAP 0BBR_0BAW 0BBH_0BAA 0BAG_0BBI 0BAS_0BBT
	population	1BAA_1BAF 1BAP_1BBP 1BAC_1BAB 1BBB_1BBR 1BAE_1BAI
10	population	0BBA_0BAI 0BAX_0BCA 0BCB_0BAM 0BAU_0BAP 0BBH_0BAA
	population	1BAP_1BBP 1BAR_1BBA 1BAQ_1BAN 1BBQ_1BBI 1BAD_1BAW

Allele frequency spectra

Locus 1						
Allele	Pop1	Pop2	All			
16	0.220	0.167	0.196			
19	0.040	0.071	0.054			
18	0.060	0.119	0.087			
15	0.220	0.024	0.130			
21	0.020	0.167	0.087			
23	0.020	0.119	0.065			
17	0.280	0.095	0.196			
22	0.060	0.119	0.087			
25	0.060	0.024	0.043			
24	0.020	-	0.043			
26	-	0.024	0.011			
27	_	0.048	0.022			
29	_	0.040	0.022			
Alleles	10	12	13			
Samplesize	50	42	92			
H _{exp}	0.811	0.883	0.874			
' 'exp	0.011	0.000	0.07 1			
Locus 2						
Allele	Pop1	Pop2	All			
16	0.520	0.571	0.543			
19	0.040	-	0.022			
18	0.220	0.119	0.174			
17	0.160	0.167	0.163			
15	0.020	-	0.011			
21	0.020	0.071	0.043			
20	0.020	0.024	0.022			
22	-	0.048	0.022			
Alleles	7	6	8			
Samplesize	50	42	92			
$H_{\rm exp}$	0.653	0.624	0.644			
Locus 3			• • •			
Allele	Pop1	Pop2	All			
19	0.240	0.262	0.250			
20	0.280	0.476	0.370			
	0.200	0.770	0.070			

Allele	Pop1	Pop2	All	
18	0.080	0.095	0.087	
21	0.280	0.119	0.207	
22	0.120	0.048	0.087	
Alleles	5	5	5	
Samplesize	50	42	92	
H _{exp}	0.765	0.679	0.743	
Locus 4				
Allele	Pop1	Pop2	All	
16	0.080	0.071	0.076	
24	0.180	0.024	0.109	
15	0.020	0.048	0.033	
25	0.160	0.167	0.163	
14	0.020	0.048	0.033	
19	0.100	0.143	0.120	
12	0.060	-	0.033	
20	0.080	0.190	0.130	
23	0.060	0.119	0.087	
28	0.020	-	0.011	
22	0.060	0.024	0.043	
21	0.160	0.119	0.141	
13	-	0.024	0.011	
26	-	0.024	0.011	
Alleles	12	12	14	
Samplesize	50	42	92	
H _{exp}	0.882	0.875	0.892	
Locus 5				
Allele	Pop1	Pop2	All	
20	0.400	0.524	0.457	
21	0.420	0.357	0.391	
19	0.180	0.119	0.152	
Alleles	3	3	3	
Samplesize	50	42	92	
H _{exp}	0.631	0.584	0.615	
Locus 6				
Allele	Pop1	Pop2	All	
19	0.060	-	0.033	
20	0.100	0.024	0.065	

Allele	Pop1	Pop2	All
40	0.000	0.044	0.004
18	0.300	0.214	0.261
22	0.200	0.119	0.163
21	0.120	0.476	0.283
16	0.060	-	0.033
24 17	0.160	0.048	0.109
	- 7	0.119	0.054
Alleles	7	6	8
Samplesize	50	42	92
H _{exp}	0.813	0.696	0.804
Locus 7			
Allele	Pop1	Pop2	All
23	0.040	0.238	0.130
20	0.660	0.143	0.424
22	0.180	0.190	0.185
21	0.100	0.333	0.207
19	0.020	0.095	0.054
Alleles	5	5	5
Samplesize	50	42	92
H _{exp}	0.520	0.766	0.724
Locus 8			
Allele	Pop1	Pop2	All
19	0.520	0.524	0.522
17	0.040	0.048	0.043
18	0.100	0.071	0.087
20	0.140	0.190	0.163
16	0.080	-	0.043
22	0.100	0.048	0.076
15	0.020	0.048	0.033
23	-	0.071	0.033
Alleles	7	7	8
Samplesize	50	42	92
H _{exp}	0.682	0.672	0.682
Locus 9 Allele	Pop1	Pop2	All
	rupi	rupz	All
24	0.000	0.024	0.054
47	0.080	0.024	0.00 1
19	0.080	0.429	0.359

Allele Pop1 Pop2 All
23 0.180 0.143 0.163
22 0.080 0.024 0.054
18 0.020 0.071 0.043
21 0.040 0.095 0.065
25 - 0.048 0.022
Alleles 7 8 8
Samplesize 50 42 92
H _{exp} 0.773 0.751 0.775
ogua 10
Locus 10 Allele Pop1 Pop2 All
Allele Top1 Top2 All
22 0.100 0.214 0.152
20 0.440 0.214 0.337
23 0.080 0.167 0.120
24 0.020 - 0.011
19 0.160 0.167 0.163
21 0.060 0.048 0.054
18 0.080 - 0.043
15 0.020 0.071 0.043
17 0.040 0.048 0.043
25 - 0.071 0.033
Alleles 9 8 10
Samplesize 50 42 92 H _{exp} 0.752 0.838 0.813
H _{exp} 0.752 0.838 0.813
Average expected heterozygosity
Pop1 Pop2 All
H _{exp} 0.728 0.737 0.757

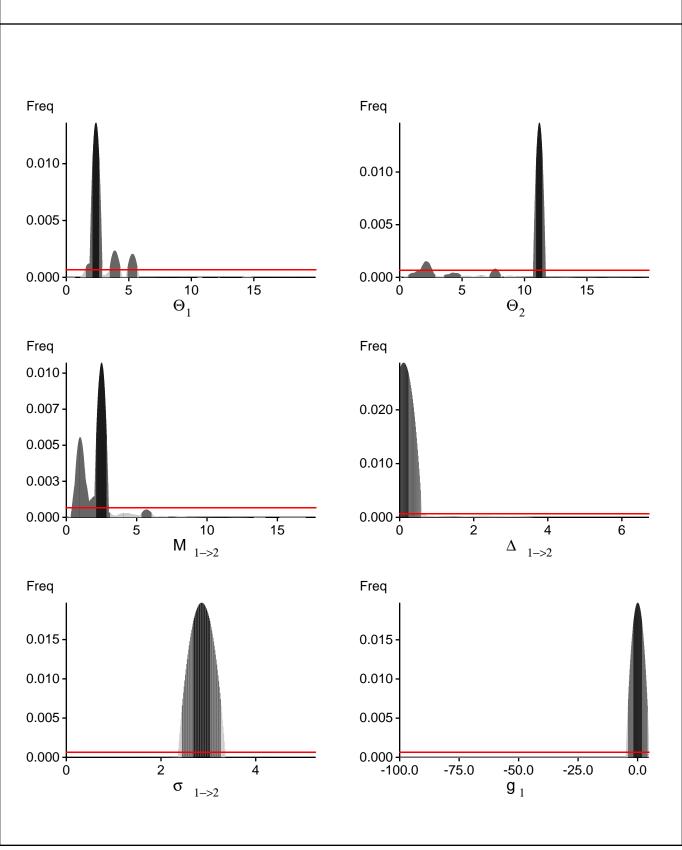
Bayesian Analysis: Posterior distribution table

Loove	Doromata	0.50/	25.00/	Mada	7F 00/	07.50/	Modian	Maaa
Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	Θ_1	2.06667	2.72000	3.74000	4.21333	4.44000	6.03333	8.37430
1	Θ_2	3.33333	3.52000	3.88667	4.22667	4.44000	3.75333	4.08255
1	M _{1->2}	0.747	5.120	5.553	6.040	6.693	4.633	4.730
1	D _{1->2}	0.00000	0.12000	0.50000	0.98667	2.57333	1.36667	2.04551
1	S _{1->2}	3.14667	3.21333	3.56667	3.89333	3.93333	2.43333	2.19997
1	g ₁	-3.60000	-1.20000	0.73333	2.53333	4.93333	0.86667	0.67156
2	Θ_1	1.25333	2.01333	2.55333	3.02667	4.05333	2.75333	3.02211
2	Θ_2	1.84000	2.21333	2.78000	3.73333	3.98667	9.95333	9.21780
2	M _{1->2}	7.347	7.573	7.980	9.107	10.240	7.540	7.108
2	D _{1->2}	1.58667	1.88000	2.22000	2.57333	2.84000	2.00667	1.84222
2	S _{1->2}	0.84000	1.38667	1.79333	2.05333	2.66667	1.76667	1.76160
2	g ₁	-4.53333	-2.00000	-0.06667	1.73333	4.26667	0.06667	-0.08046
3	Θ_1	1.46667	1.66667	2.16667	2.81333	3.66667	3.38000	6.71034
3	Θ_2	0.36000	1.16000	1.86000	2.44000	3.00000	4.26000	5.75152
3	M _{1->2}	0.080	0.453	0.887	1.293	4.347	2.300	2.389
3	D _{1->2}	0.24000	1.06667	1.68667	2.08000	3.38667	1.72667	1.76439
3	S _{1->2}	0.05333	0.20000	0.59333	0.96000	2.25333	0.86000	1.00955
3	g_1	-4.00000	-1.60000	0.33333	2.13333	4.53333	0.46667	0.30360
4	Θ_1	2.37333	2.50667	2.98000	4.14667	8.04000	5.86000	7.00615
4	Θ_2	12.70667	12.92000	13.43333	13.93333	14.16000	7.34000	8.04551
4	M _{1->2}	0.133	1.653	2.687	3.053	3.520	2.113	2.083
4	D _{1->2}	0.00000	0.61333	0.98000	1.41333	2.94667	2.24667	2.90958
4	S _{1->2}	2.94667	3.36000	3.90000	4.50667	5.34667	4.19333	4.62447
4	g ₁	-4.13333	-1.73333	0.06667	1.86667	4.26667	0.20000	0.12602
5	Θ_1	1.24000	1.62667	1.96667	2.57333	3.93333	2.43333	2.64593
5	Θ_2	0.36000	4.04000	4.66000	5.09333	6.14667	4.40667	4.92240
5	M _{1->2}	5.240	5.333	5.793	6.173	7.253	5.460	5.245
5	D _{1->2}	1.26667	2.62667	2.75333	3.05333	5.04000	2.92667	2.86121
5	S _{1->2}	0.50667	1.44000	1.88667	2.40000	3.72000	2.00667	2.06014
5	g ₁	-4.00000	-1.60000	0.33333	2.13333	4.53333	0.46667	0.25501
6	Θ_1	3.06667	3.98667	4.52667	5.10667	5.81333	5.47333	9.13478

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
6	Θ_2	3.34667	4.14667	4.56667	4.90667	5.17333	4.16667	5.36807
6	M _{1->2}	0.000	0.813	1.060	1.507	2.947	1.527	1.540
6	D _{1->2}	0.00000	0.00000	0.31333	0.60000	1.70667	1.63333	1.64083
6	S _{1->2}	1.45333	1.88000	2.06000	2.22667	3.66667	2.59333	2.58037
6	g ₁	-5.06667	-2.13333	-0.20000	1.86667	4.80000	0.06667	-0.21713
7	Θ_1	1.64000	4.40000	4.94000	5.38667	7.24000	4.98000	5.35538
7	Θ_2	6.22667	6.36000	7.48667	7.94667	8.10667	7.59333	8.93198
7	M _{1->2}	0.267	0.440	0.820	1.173	2.827	3.207	3.462
7	D _{1->2}	0.77333	2.49333	2.96667	3.40000	3.62667	2.55333	2.28157
7	S _{1->2}	1.41333	1.82667	2.40667	2.84000	3.94667	2.56667	2.72622
7	g_1	-3.86667	-1.33333	0.60000	2.40000	4.93333	0.73333	0.61233
8	Θ_1	8.73333	8.81333	9.30000	9.72000	9.90667	4.71333	5.62873
8	Θ_2	0.40000	2.10667	4.16667	4.68000	4.94667	4.15333	6.66395
8	M _{1->2}	0.173	0.360	0.780	1.693	2.440	4.100	3.943
8	D _{1->2}	0.00000	0.05333	0.43333	0.78667	2.68000	1.18000	1.31636
8	S _{1->2}	0.65333	0.88000	1.28667	1.64000	4.66667	2.51333	2.57079
8	g_1	-4.13333	-1.73333	0.20000	2.00000	4.40000	0.33333	0.23158
9	Θ_1	1.74667	3.32000	3.60667	3.92000	4.89333	3.48667	3.49764
9	Θ_2	0.36000	0.81333	1.39333	1.88000	2.88000	4.76667	6.64425
9	M _{1->2}	0.520	1.733	2.647	3.213	3.320	4.020	5.907
9	D _{1->2}	1.20000	1.70667	2.15333	2.62667	3.48000	2.11333	2.01224
9	S _{1->2}	0.00000	0.13333	0.54000	1.05333	1.73333	2.38000	2.52680
9	g ₁	-4.26667	-1.86667	0.06667	1.86667	4.13333	0.20000	0.02860
10	Θ_1	1.41333	1.82667	2.30000	2.73333	2.88000	3.64667	3.70170
10	Θ_2	0.30667	1.88000	2.43333	2.96000	4.29333	3.36667	6.92783
10	M _{1->2}	2.440	2.653	3.060	3.853	8.107	5.007	4.969
10	D _{1->2}	0.00000	0.00000	0.19333	0.54667	2.85333	1.04667	1.01289
10	S _{1->2}	1.72000	3.82667	4.23333	4.62667	4.84000	3.76667	3.58304
10	g_1	-4.66667	-2.13333	-0.20000	1.60000	4.13333	-0.06667	-0.21024
All	Θ_1	1.56000	2.09333	2.38000	2.64000	2.86667	2.48667	2.90454
All	Θ_2	10.68000	10.93333	11.19333	11.44000	11.69333	11.10000	9.41134
All	M _{1->2}	0.333	2.120	2.500	2.867	3.120	2.327	2.201
All	D _{1->2}	0.00000	0.00000	0.11333	0.24000	0.57333	0.24667	0.15867
All	S _{1->2}	2.42667	2.66667	2.86000	3.04000	3.26667	2.87333	2.85932
All	g ₁	-4.13333	-1.86667	0.06667	1.86667	4.26667	0.20000	0.06667

Citation suggestions:
Beerli P., 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. Bioinformatics 22:341-345 Beerli, P., H. Ashki, S. Mashayekhi, and M. Palczewski, 2022. Population divergence time estimation
using individual lineage label switching. G3 Genesâ Genomesâ Genetics, 12(4), 02 2022. Beerli P., 2009. How to use MIGRATE or why are Markov chain Monte Carlo programs difficult to use? In Population Genetics for Animal Conservation, G. Bertorelle, M. W. Bruford, H. C. Hauffe, A. Rizzoli, and C. Vernesi, eds., vol. 17 of Conservation Biology, Cambridge University Press, Cambridge UK, pp. 42-79.
Beerli, P., S. Mashayekhi, M. Sadeghi, M. Khodaei, and K. Shaw, 2019. Population genetic inference with migrate Current Protocols in Bioinformatics, 68(1):e87.

Bayesian Analysis: Posterior distribution over all loci



Example: Microsatellite data set 15				

Log-Probability of the data given the model (marginal likelihood)

Use this value for Bayes factor calculations:

BF = Exp[In(Prob(D | thisModel) - In(Prob(D | otherModel) or as LBF = 2 (In(Prob(D | thisModel) - In(Prob(D | otherModel)) shows the support for thisModel]

Locus	TI(1a)	BTI(1b)	HS(3)	
1	-140112.04	-22568.95	-490.85	
2	-13782.42	-2248.73	-45.89	
3	-319.61	-94.68	-71.19	
4	-522.87	-334.77	-1857.85	
5	-53.41	-31.80	-26.43	
6	-561.49	-189.99	-348.54	
7	-1554.27	-279.29	-79.10	
8	-234.04	-94.19	-126.01	
9	-208.75	-119.91	-247.52	
10	-239.78	-85.34	-207.56	
All	-157533.48	-25992.47	-3445.74	

- (1a) TI: Thermodynamic integration: log(Prob(D|Model)): Good approximation with many temperatures
- (1b) BTI: Bezier-approximated Thermodynamic integration: when using few temperatures USE THIS!
- (3) HS: Harmonic mean approximation: Overestimates the marginal likelihood, poor variance [Scaling factor = 55.195736]

Citation suggestions:

Beerli P. and M. Palczewski, 2010. Unified framework to evaluate panmixia and migration direction among multiple sampling locations, Genetics, 185: 313-326.

Palczewski M. and P. Beerli, 2014. Population model comparison using multi-locus datasets.

In M.-H. Chen, L. Kuo, and P. O. Lewis, editors, Bayesian Phylogenetics: Methods, Algorithms, and Applications, pages 187-200. CRC Press, 2014.

Acceptance ratios for all parameters and the genealogies

Parameter	Accepted changes	Ratio
Θ_1	351/351	1.00000
Θ_2	330/330	1.00000
$M_{1\rightarrow 2}$	317/317	1.00000
Δ $1\rightarrow 2$	333/353	0.94334
$\sigma_{1\rightarrow 2}$	304/327	0.92966
g_1	25/317	0.07886
Genealogies	943/2005	0.47032

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sampe Size		
Θ_1	0.93458	136.36		
Θ_2	0.92261	161.90		
$M_{1\rightarrow 2}$	0.91389	180.65		
Δ 1->2	0.96660	67.96		
$\sigma_{1\rightarrow 2}$	0.95335	95.67		
g_1	-0.00003	3980.20		
Genealogies	0.92028	175.17		

Average temperatures during the run

Chain	Temperatures					
1	1.00000					
2	0.66667					
3	0.33333					
4	0.00000					

Example: Microsatellite data set -- 20

Potential Problems

This section reports potential problems with your run, but such reporting is often not very accurate. With many parameters in a multilegus analysis, it is common that some parameters for some loci will not be informative

These parameters then trigger suggestions to increase the prior range that are not sensible. Do not blindly follow the suggestions given. If some parameters are flagged, inspect the tables carefully and judge whether an action is required. Suppose you run a Bayesian inference with sequence data for macroscopic species. In that case, there is rarely the need to increase the prior for Theta beyond 0.1. If you use microsatellites data, it is rather common that your prior distribution for Theta should have a range from 0.0 to 100 or more. With many populations (>3), it is also very common that some migration routes are estimated poorly because the data contains little or no information for that route. Increasing the prior range will not help in such situations, but reducing the number of parameters may help.	
No warning was recorded during the run	