Results from scenario with 6 diploids, 7 tetraploids, 3 hybridizations

Graham Jones

2012-11-02

1   Scenario

Scenario shown in Fig 1. All genes have length 500. Population sizes are 100,000 at tips, and at rootward ends of branches, and 200,000 at tipward ends of internal branches and at the root. The root height is 0.035.

Strict clock branch rates assumed in simulation and BEAST.

HKY substitution model assumed in simulation and BEAST. In simulations, the Seq-Gen parameters are kappa=3, frequencies .3 A and T, .2 C and G. (Seq-Gen called with `-t3.0 -f0.3,0.2,0.2,0.3`. These are estimated in BEAST.

No site rate heterogeneity assumed within genes in simulation and BEAST.

Genes have the same mutation rate in simulations. Rates are estimated in BEAST.

Priors on population size scaling factor η, relative mutation rates of genes, and the 'event rate' (something like a diversification rate) λ for the diploid and tetraploid trees in the network, are all diffuse lognormals.

Number of genes G=5, individuals per species N=2, mutation rate per site per generation T=3.2e-7.

Root is $.035/(3.2 \times 10^{-7}) = 109375$ generations. The expected time for a coalesecence between two genes in the same population of size 100,000 (100,000 gene copies per diploid genome) is 100,000 generations, nearly the height of the tree. Since population sizes vary from 100,000 to 200,000, the coalescences are expected to be deep. In other words, incomplete lineage sorting is a huge problem in this scenario.

MCMC chains are 7,000,000 generations with first 2,000,000 discarded as burnin. Each runs takes about an hour.

2   Results

Figs 2 - 6. Simulated gene trees for replicate 1. These are the gene trees along which sequences are simulated. A label 02xB means individual 2 from species x, sequence (homeolog) B. Many of the coalescences are deeper than 0.035, and so are in the root of the MUL-tree. Many pairs of individuals from the same species do not coalesce with one another first.

Fig 7 shows how the MCMC chain explore different numbers of hybridizations. It spends about .88 of the time at 3 hybridizations, though the run is too short for a good estimate.

1

For comparison, Fig 8 shows a similar trace for a case with no data. In this case, priors on the population size scaling factor η and event rate λ are chosen to produce deep coalescences ($\eta\lambda \gg 1$), which keeps the single gene tree 'out of the way'. If this is not done mixing can be poor.

Figs 9 - 13. Maximum clade credibility MUL-trees for five replicates. In replicates 1-4, the tetraploid species z, y, x, w, v, u, t are correctly partitioned as $\{z, y, x, w\}, \{v\}, \{u, t\}$. In replicate 5, the partition is wrong, namely $\{z, y, x\}, \{w\}, \{v\}, \{u, t\}$. In all cases the MUL-tree is poorly resolved, with about 5 out of 19 nodes having posterior probability over 0.9, and about half the nodes having posterior probability below 0.5.

Fig 14 Maximum clade credibility MUL-tree for replicate 5 with posterior probabilities and 95% HPD node height bars.
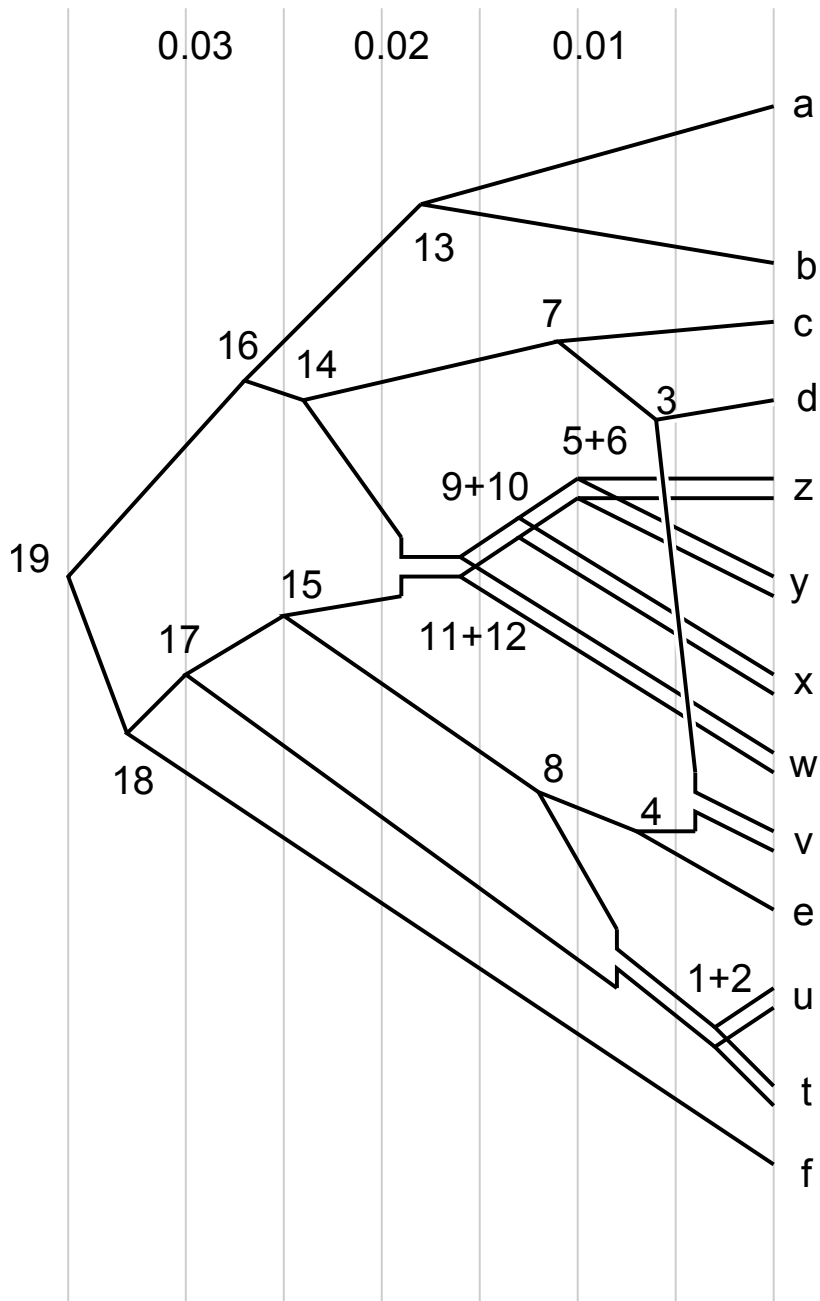
Figure 1: Scenario: the true MUL-tree
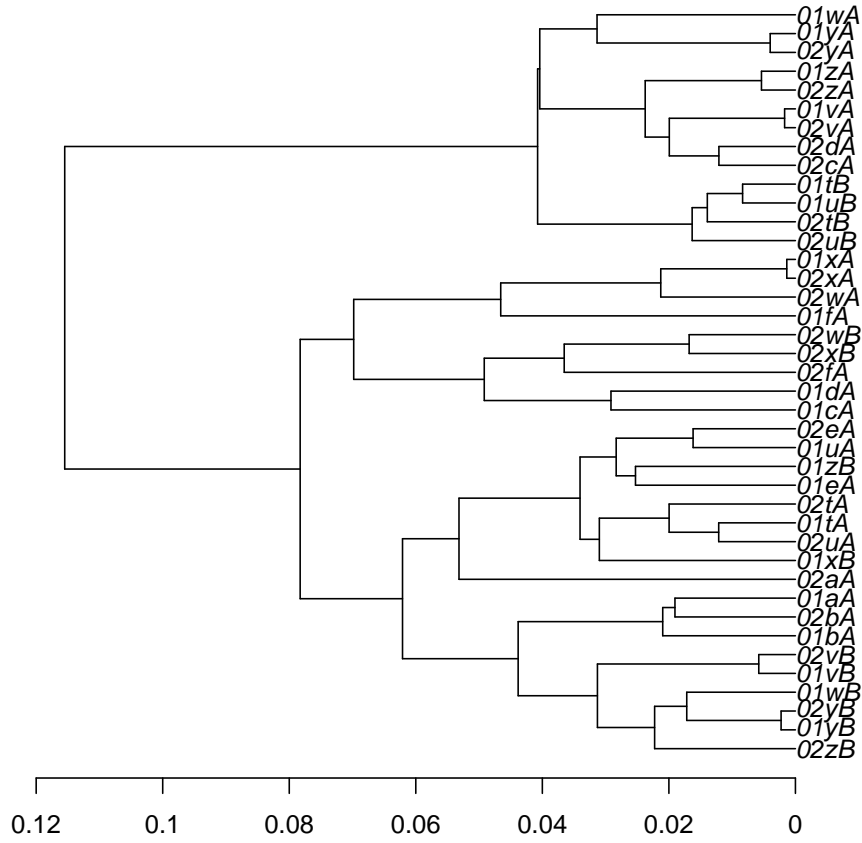
Simulated gene tree: rep 1 gene 1

01wA
01yA
02yA
01zA
02zA
01vA
02vA
02dA
02cA
01tB
01uB
02tB
02uB
01xA
02xA
02wA
01fA
02wB
02xB
02fA
01dA
01cA
02eA
01uA
01zB
01eA
02tA
01tA
02uA
01xB
02aA
01aA
02bA
01bA
02vB
01vB
01wB
02yB
01yB
02zB

| 0.12 | 0.1 | 0.08 | 0.06 | 0.04 | 0.02 | 0 |

Figure 2: Replicate 1, gene 1

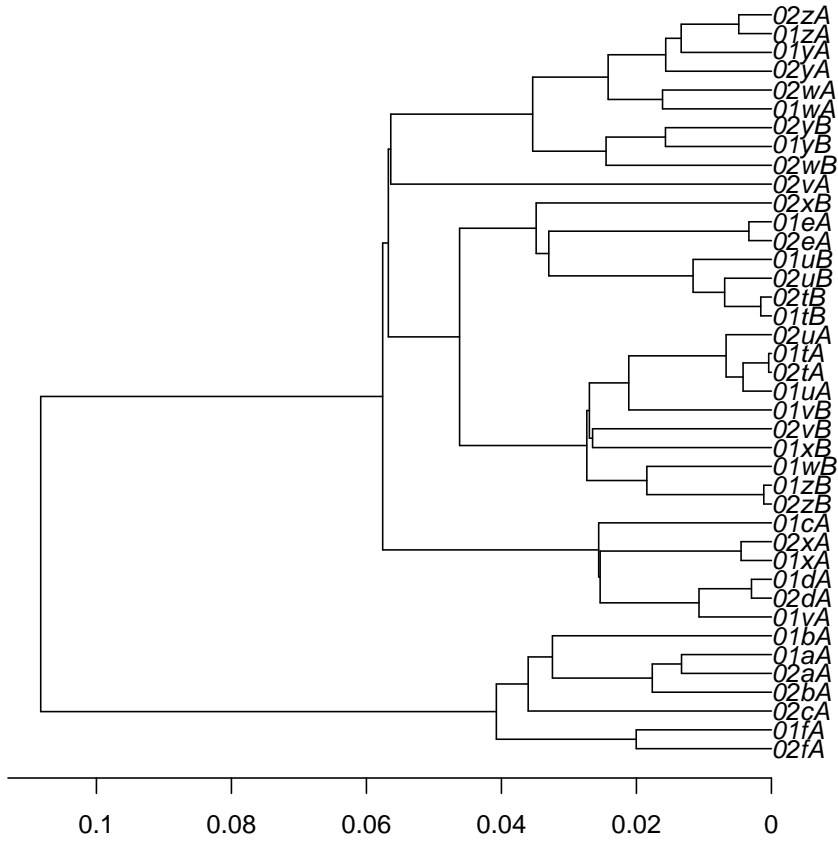**Simulated gene tree: rep 1 gene 2**

Figure 3: Replicate 1, gene 2
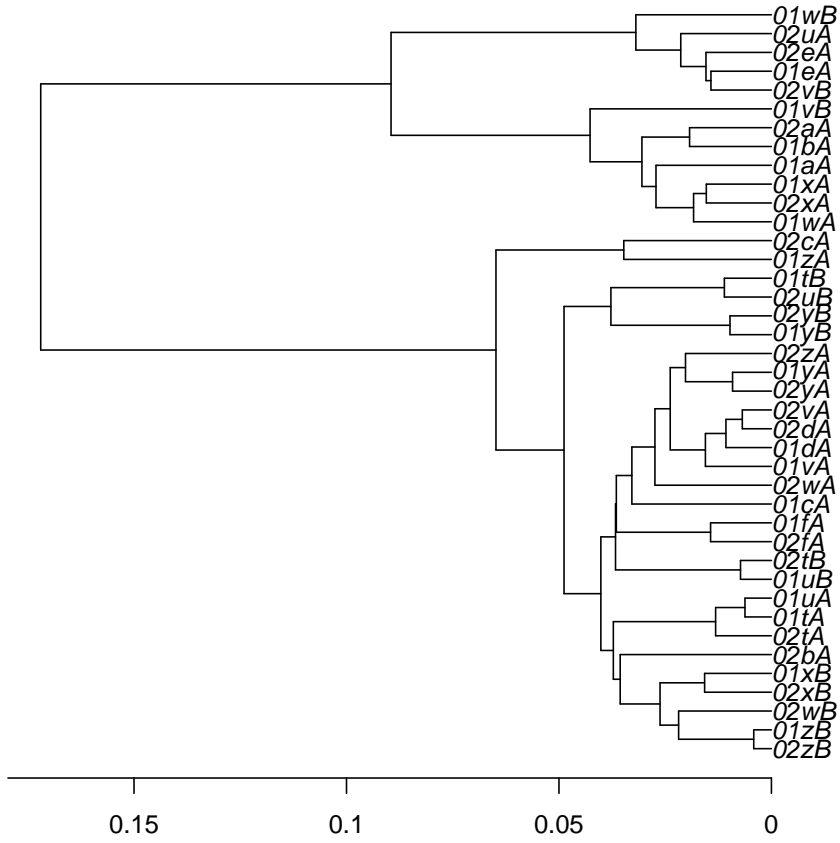
Simulated gene tree: rep 1 gene 3

Figure 4: Replicate 1, gene 3

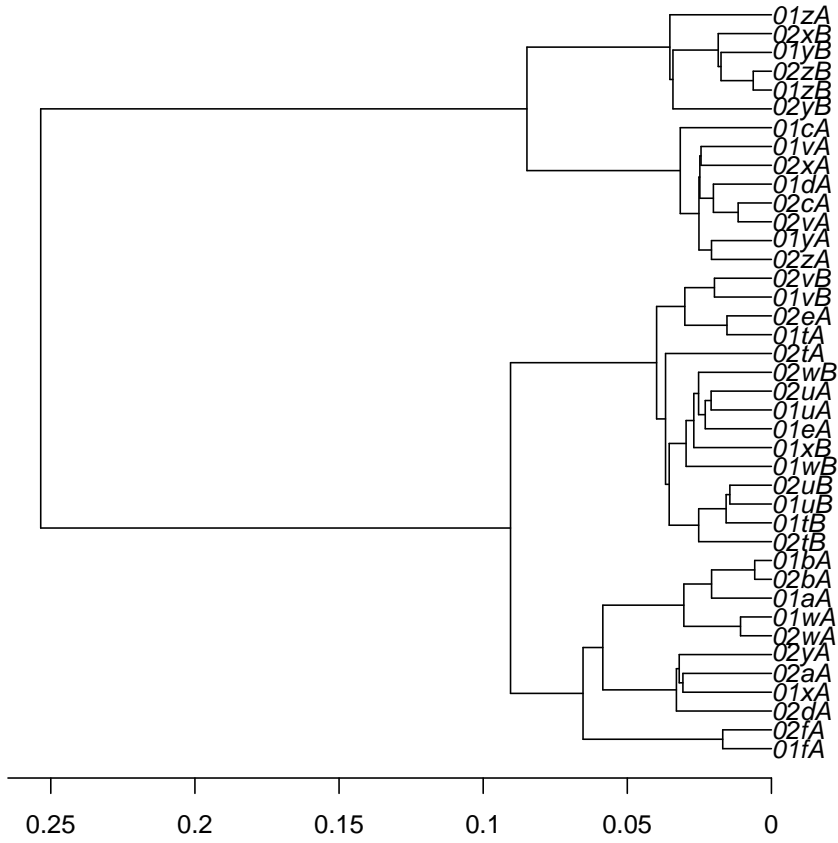**Simulated gene tree: rep 1 gene 4**

Figure 5: Replicate 1, gene 4

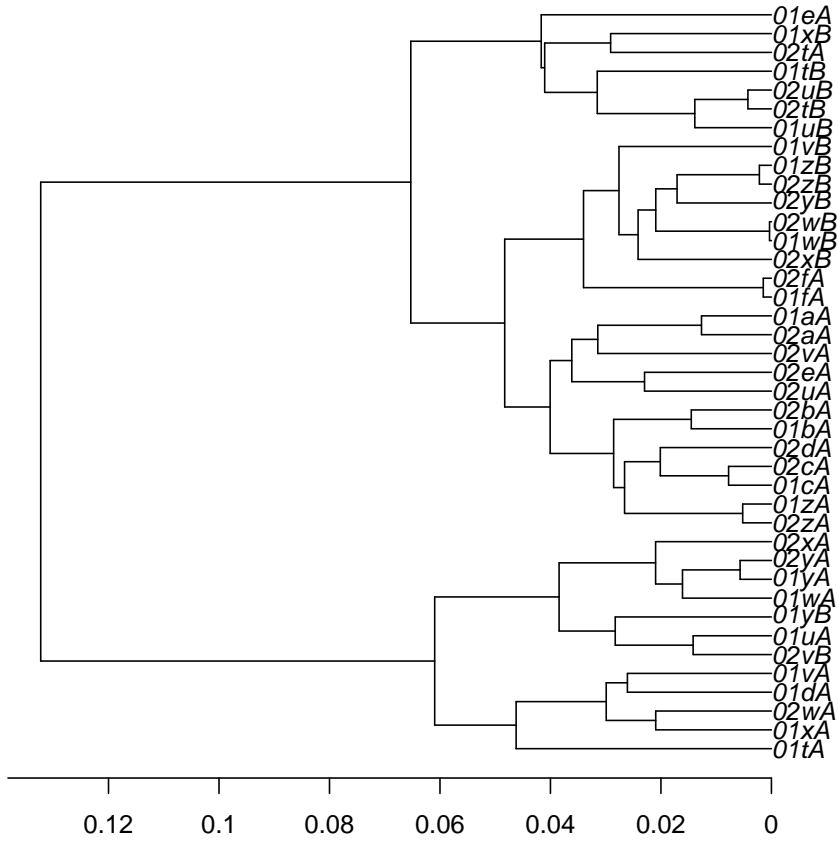**Simulated gene tree: rep 1 gene 5**

Figure 6: Replicate 1, gene 5
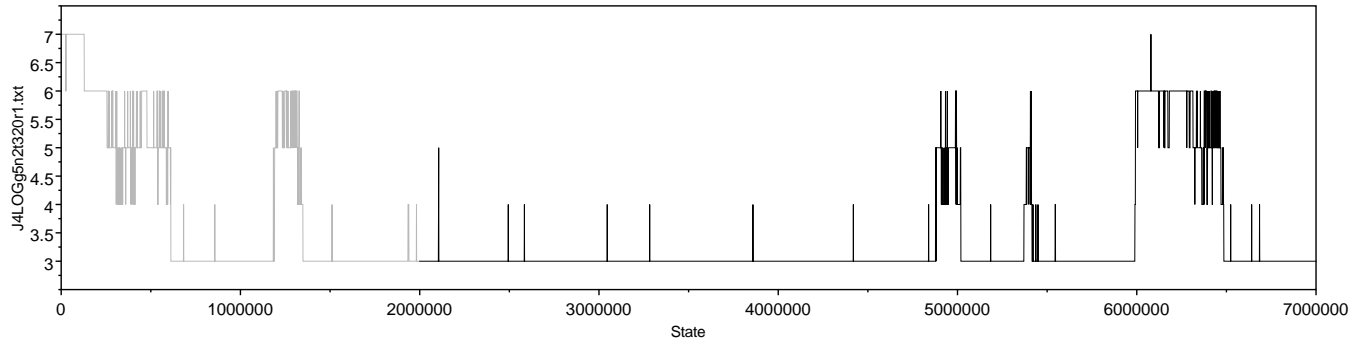
Figure 7: Trace for number of hybridizations for Replicate 1
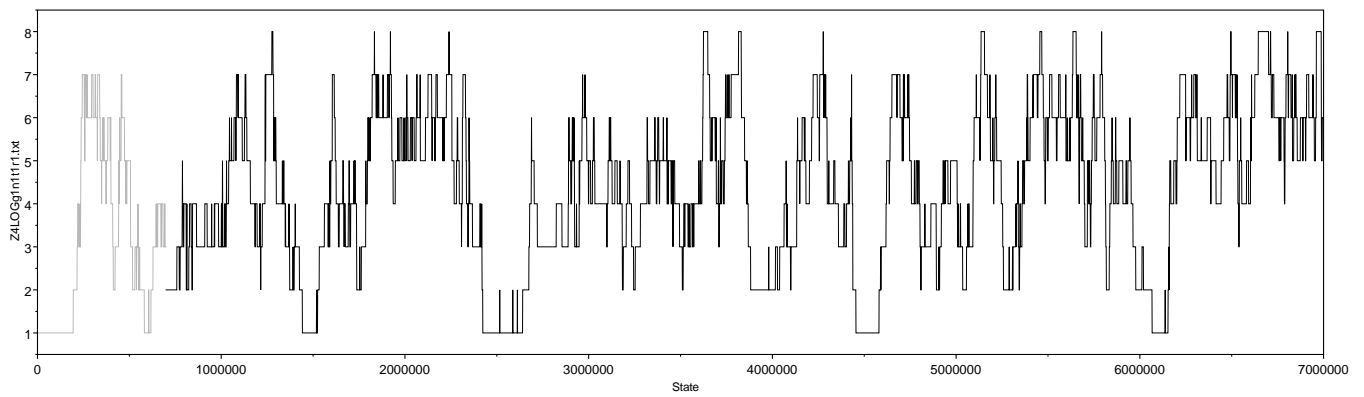
Figure 8: Trace for number of hybridizations for 8 diploids, 8 tetraploids, no data

9

Concensus MUL–tree: rep 1
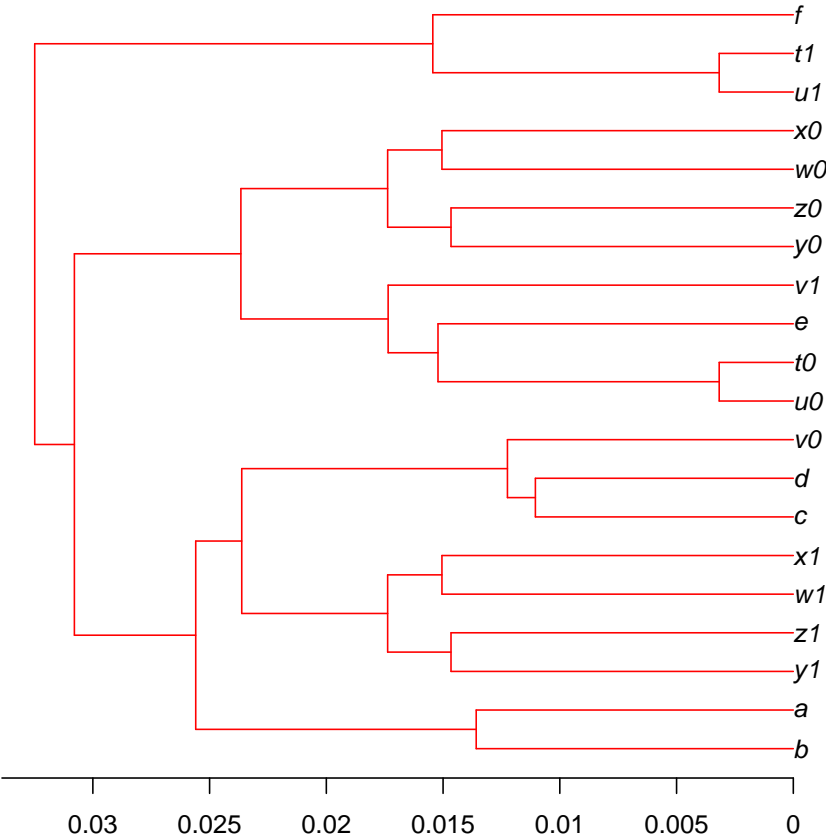
Figure 9: Replicate 1

Concensus MUL−tree: rep 2

b
a
t0
u0
f
c
v1
d
z0
y0
x0
w0
z1
y1
x1
w1
v0
e
t1
u1

0.03   0.02   0.01   0
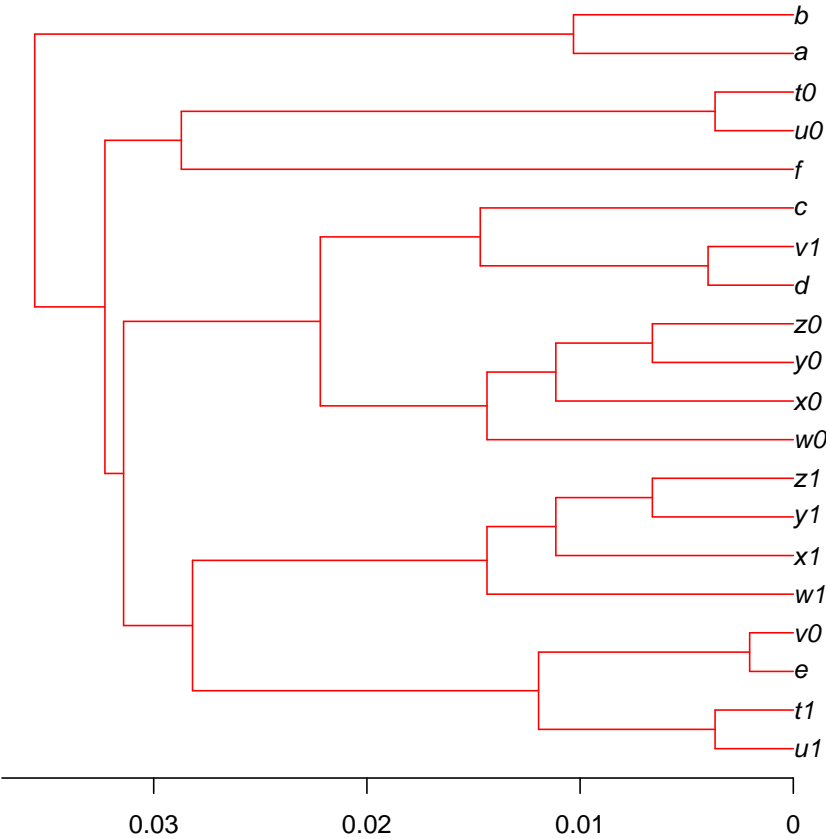
Figure 10: Replicate 2

Concensus MUL−tree: rep 3
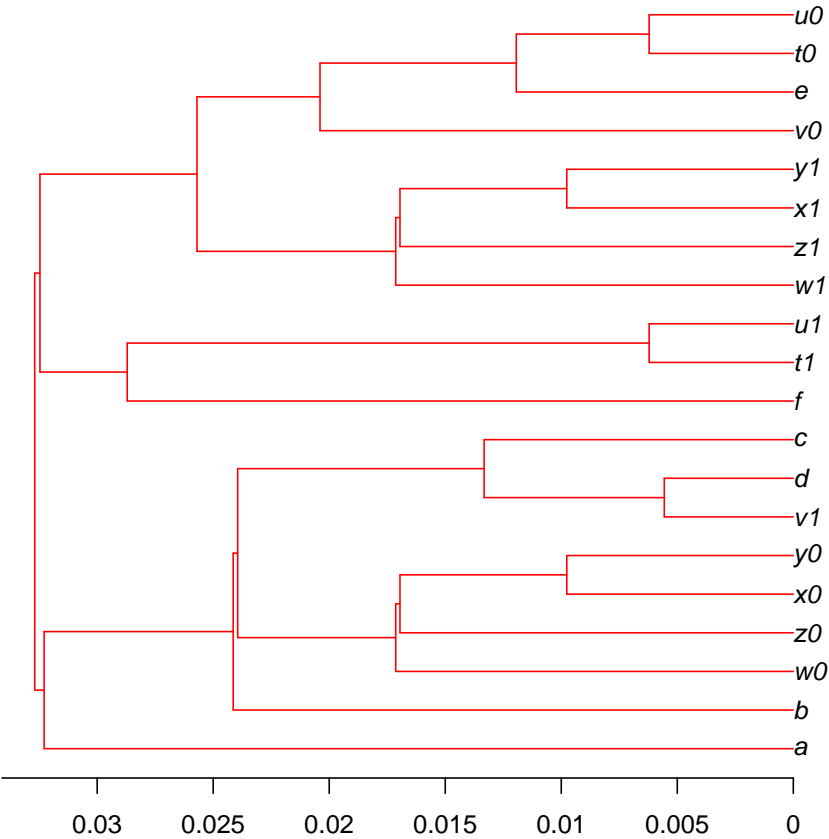
Figure 11: Replicate 3

Concensus MUL–tree: rep 4
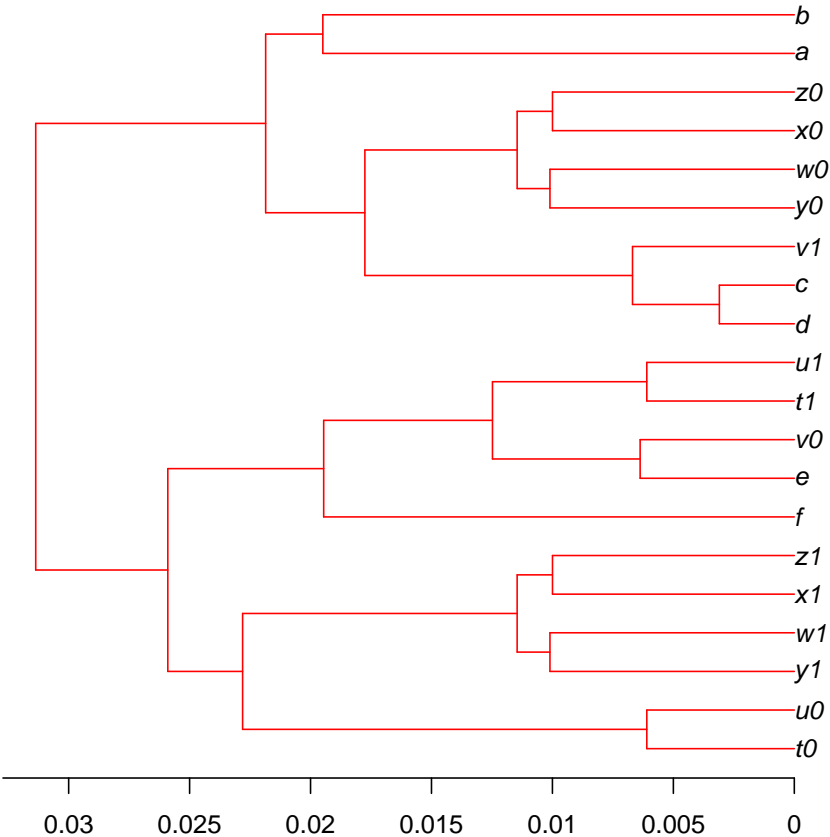
Figure 12: Replicate 4

13
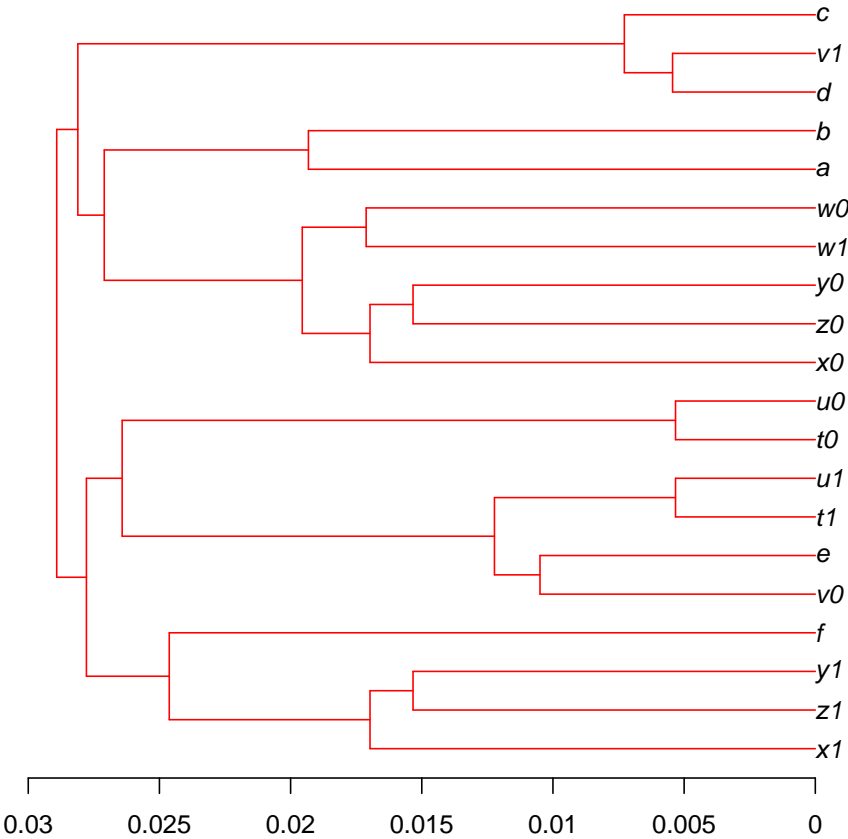
Figure 13: Replicate 5

14
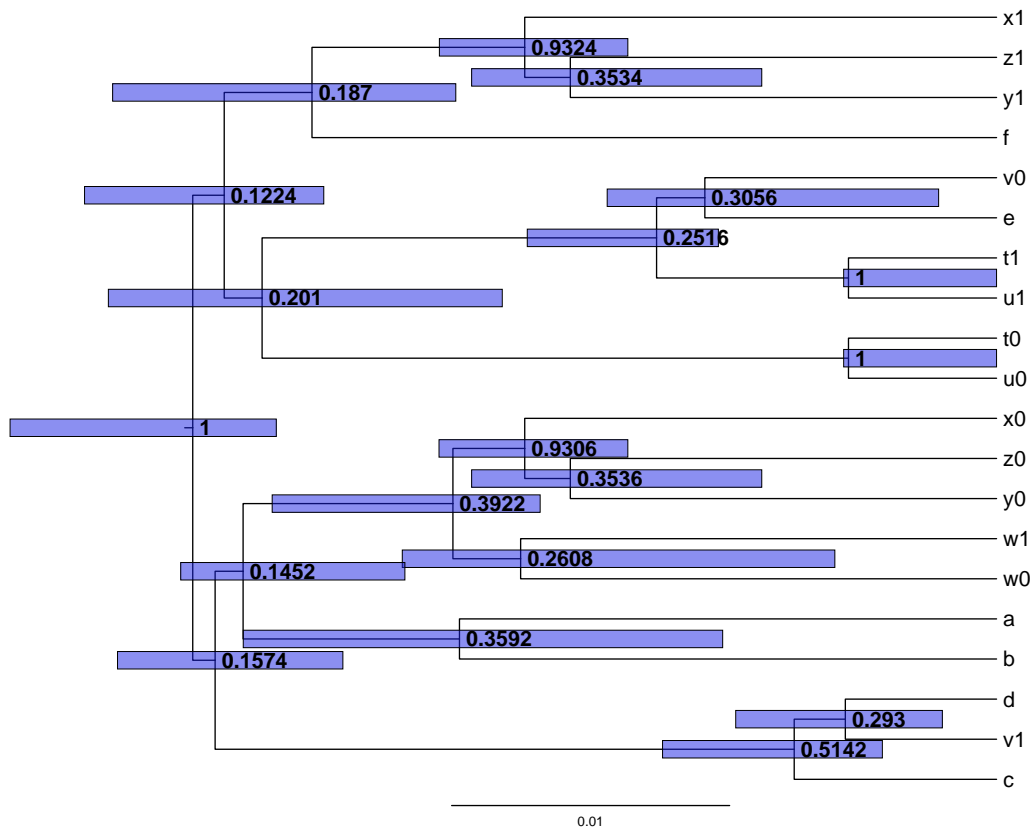
Figure 14: Replicate 5 again, with more detail

3   Configuration file

```
PROGRAMPARAMETERS
beastseed 42
beastchainlength 7000000
beastscreen.logevery 10000
beastparams.logevery 1000
beastgtrees.logevery 1000
beastmultree.logevery 1000
beastdbugtune.logevery 10000
treeannburnin 2001
SCENARIO
nofGNRT 1
5 2 5  3.2e-7
genelength 500
nofdips 6
noftets 7
aA                                                     hgt 0      hybhgt *     tipp 100000 rootp 100000 hybp *
bA                                                     hgt 0      hybhgt *     tipp 100000 rootp 100000 hybp *
cA                                                     hgt 0      hybhgt *     tipp 100000 rootp 100000 hybp *
dA                                                     hgt 0      hybhgt *     tipp 100000 rootp 100000 hybp *
eA                                                     hgt 0      hybhgt *     tipp 100000 rootp 100000 hybp *
fA                                                     hgt 0      hybhgt *     tipp 100000 rootp 100000 hybp *
zA                                                     hgt 0      hybhgt *     tipp 100000 rootp 100000 hybp *
zB                                                     hgt 0      hybhgt *     tipp 100000 rootp 100000 hybp *
yA                                                     hgt 0      hybhgt *     tipp 100000 rootp 100000 hybp *
yB                                                     hgt 0      hybhgt *     tipp 100000 rootp 100000 hybp *
xA                                                     hgt 0      hybhgt *     tipp 100000 rootp 100000 hybp *
xB                                                     hgt 0      hybhgt *     tipp 100000 rootp 100000 hybp *
wA                                                     hgt 0      hybhgt *     tipp 100000 rootp 100000 hybp *
wB                                                     hgt 0      hybhgt *     tipp 100000 rootp 100000 hybp *
vA                                                     hgt 0      hybhgt 0.004 tipp 100000 rootp 100000 hybp 100000
vB                                                     hgt 0      hybhgt 0.004 tipp 100000 rootp 100000 hybp 100000
uA                                                     hgt 0      hybhgt *     tipp 100000 rootp 100000 hybp *
uB                                                     hgt 0      hybhgt *     tipp 100000 rootp 100000 hybp *
tA                                                     hgt 0      hybhgt *     tipp 100000 rootp 100000 hybp *
tB                                                     hgt 0      hybhgt *     tipp 100000 rootp 100000 hybp *
(uA,tA)                                                hgt 0.003  hybhgt 0.008 tipp *      rootp 100000 hybp 100000
(uB,tB)                                                hgt 0.003  hybhgt 0.008 tipp *      rootp 100000 hybp 100000
(dA,vA)                                                hgt 0.006  hybhgt *     tipp *      rootp 100000 hybp *
(eA,vB)                                                hgt 0.006  hybhgt *     tipp *      rootp 100000 hybp *
(zA,yA)                                                hgt 0.010  hybhgt *     tipp *      rootp 100000 hybp *
(zB,yB)                                                hgt 0.010  hybhgt *     tipp *      rootp 100000 hybp *
(cA,(dA,vA))                                           hgt 0.011  hybhgt *     tipp *      rootp 100000 hybp *
((eA,vB),(uA,tA))                                      hgt 0.012  hybhgt *     tipp *      rootp 100000 hybp *
((zA,yA),xA)                                           hgt 0.013  hybhgt *     tipp *      rootp 100000 hybp *
((zB,yB),xB)                                           hgt 0.013  hybhgt *     tipp *      rootp 100000 hybp *
(((zA,yA),xA),wA)                                      hgt 0.016  hybhgt 0.019 tipp *      rootp 100000 hybp 100000
(((zB,yB),xB),wB)                                      hgt 0.016  hybhgt 0.019 tipp *      rootp 100000 hybp 100000
(aA,bA)                                                hgt 0.017  hybhgt *     tipp *      rootp 100000 hybp *
((cA,(dA,vA)),(((zA,yA),xA),wA))                       hgt 0.023  hybhgt *     tipp *      rootp 100000 hybp *
((((zB,yB),xB),wB),((eA,vB),(uA,tA)))                  hgt 0.025  hybhgt *     tipp *      rootp 100000 hybp *
((aA,bA),((cA,(dA,vA)),(((zA,yA),xA),wA)))             hgt 0.027  hybhgt *     tipp *      rootp 100000 hybp *
(((((zB,yB),xB),wB),((eA,vB),(uA,tA))),(uB,tB))        hgt 0.030  hybhgt *     tipp *      rootp 100000 hybp *
((((((zB,yB),xB),wB),((eA,vB),(uA,tA))),(uB,tB)),fA)   hgt 0.032  hybhgt *     tipp *      rootp 100000 hybp *
(((aA,bA),((cA,(dA,vA)),(((zA,yA),xA),wA))),((((((zB,yB),xB),wB),((eA,vB),(uA,tA))),(uB,tB)),fA))
                                                       hgt 0.035  hybhgt *     tipp *      rootp * hybp *
```