

# BEAST2 tree (space) tools

## Convergence + Means

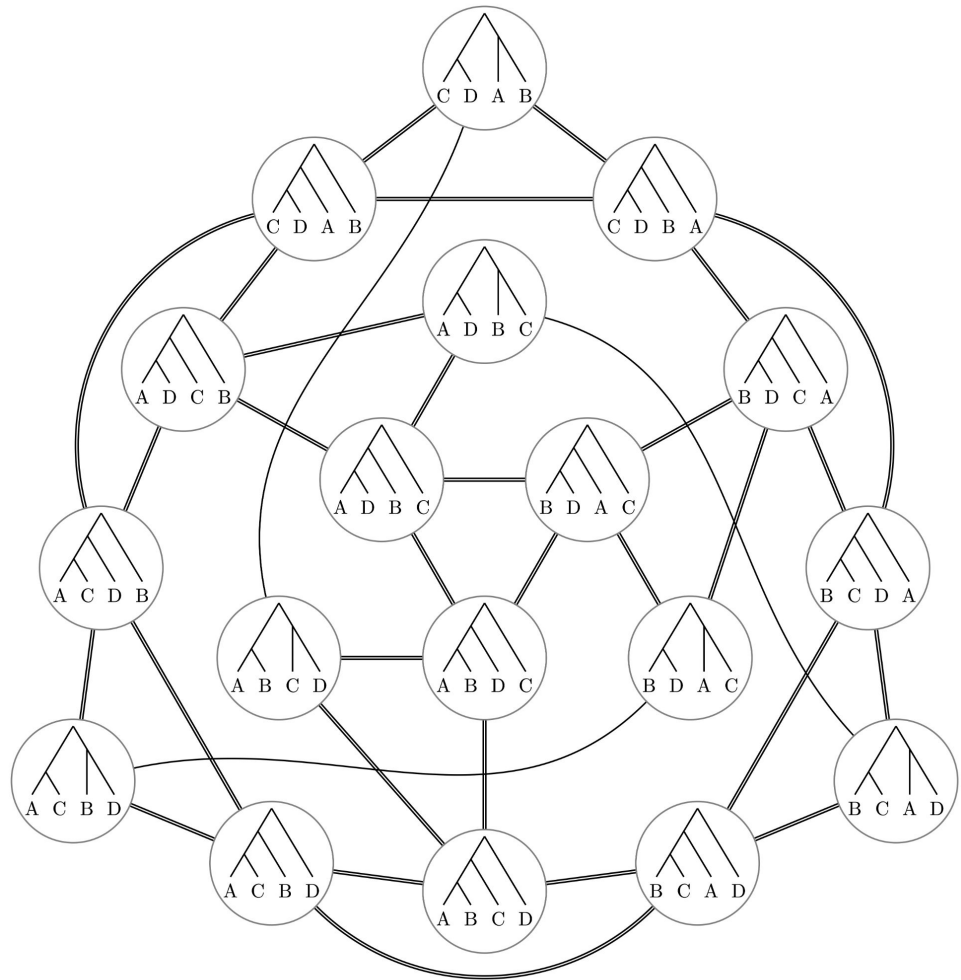
ASM and CCD

Lars Berling

Joint work with: Remco Bouckaert, Lena Collienne, Alexei Drummond, Alex Gavryushkin, Jonathan Klawitter, Walter Xie

## RNNI space with 4 taxa

- Rank moves – single edges
- NNI moves – double edges



# Convergence of Markov chain Monte Carlo

- In theory MCMC converges to the target distribution asymptotically
- Consider: convergence = indistinguishable samples from independent runs
- Phylogenetics rule of thumb for effective sample size (ESS):
  - ESS < 100 is bad
  - ESS > 100 is okay
  - ESS > 200 is converged
  - **!NOTE:** ESS is only for continuous parameter traces

→ No one is checking the trees!

**The ASM package automates best practises and adds tree space based diagnosis**

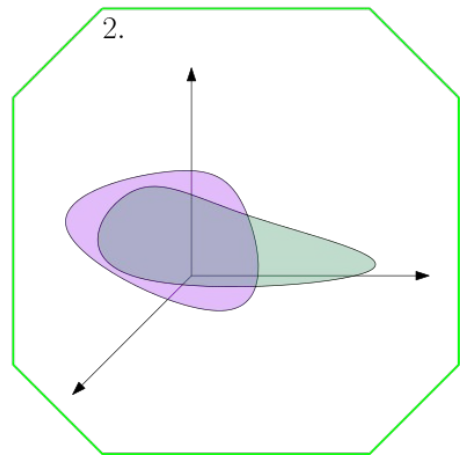
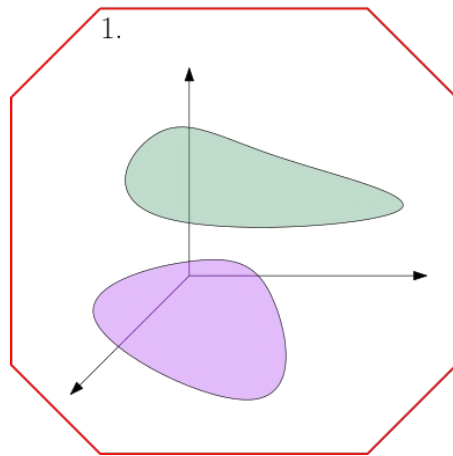
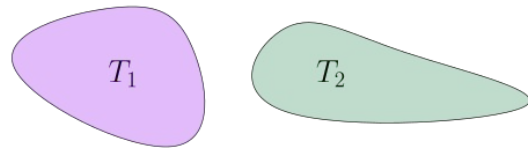
Use the RNNI space to define a convergence diagnostic:

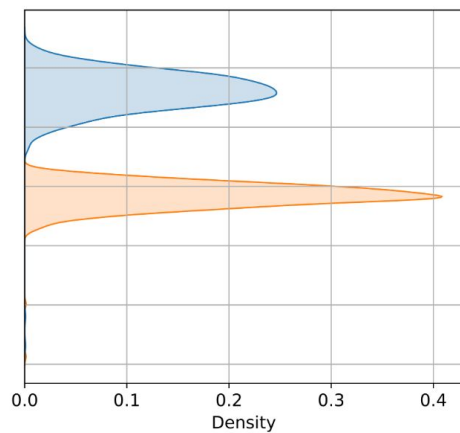
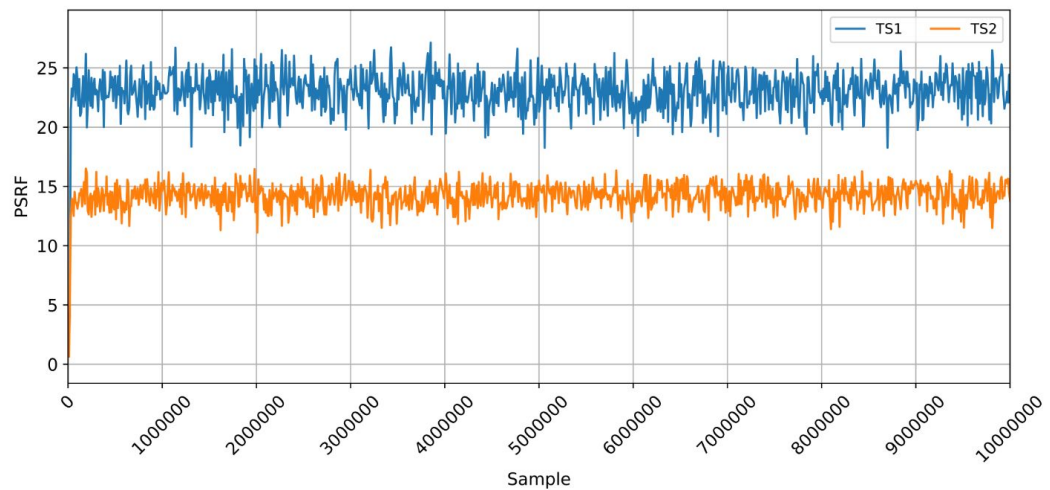
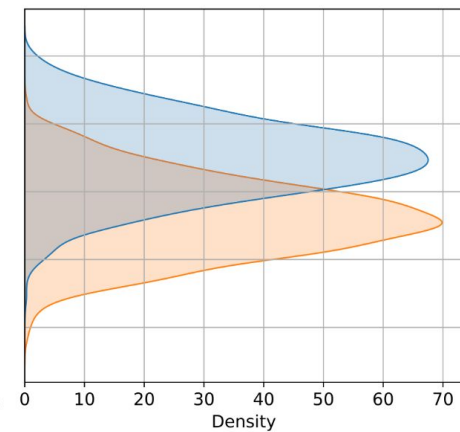
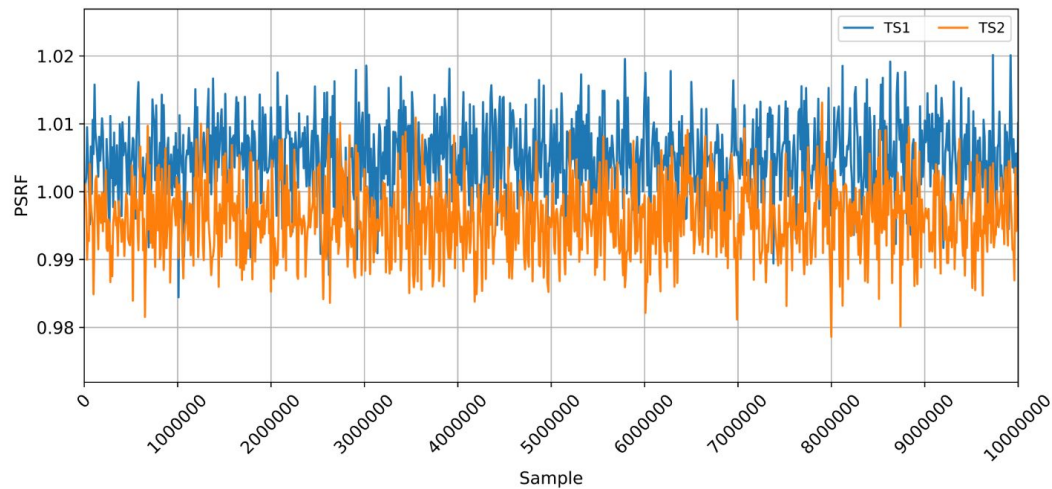
$$\text{Var}(t)_{\mathcal{T}} = \frac{\sum_{t_i \in \mathcal{T}} d(t_i, t)^2}{|\mathcal{T}|}.$$

$$\text{PSRF}(t; \mathcal{T}_1, \mathcal{T}_2) = \sqrt{\frac{\text{Var}(t)_{\mathcal{T}_2}}{\text{Var}(t)_{\mathcal{T}_1}}}, t \in \mathcal{T}_1$$

$$\text{GR}(t_i^1; \mathcal{T}_1, \mathcal{T}_2) = \frac{1}{i} \sum_{s=1}^i \text{PSRF}(t_s; \mathcal{T}_1, \mathcal{T}_2)$$

independently sampled sets of trees





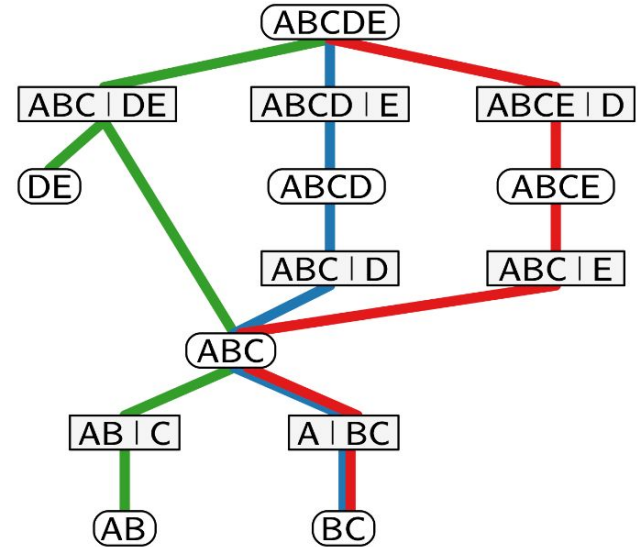
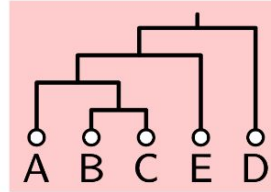
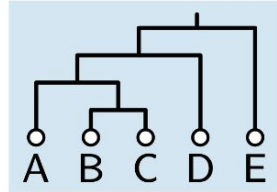
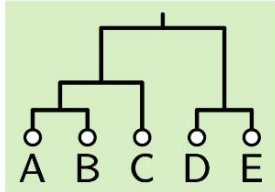
# Summary/Consensus/Mean tree

- Trees are now only topologies
- Tractable tree distribution:
  - A simple example would be a MCMC sample
  - Here we use: Conditional clade distributions (CCDs)
- We want an accurate representation of the posterior
- Use a MCMC sample to generate a CCD

# CCD graph & CCD1

- Based on observed **clade splits**

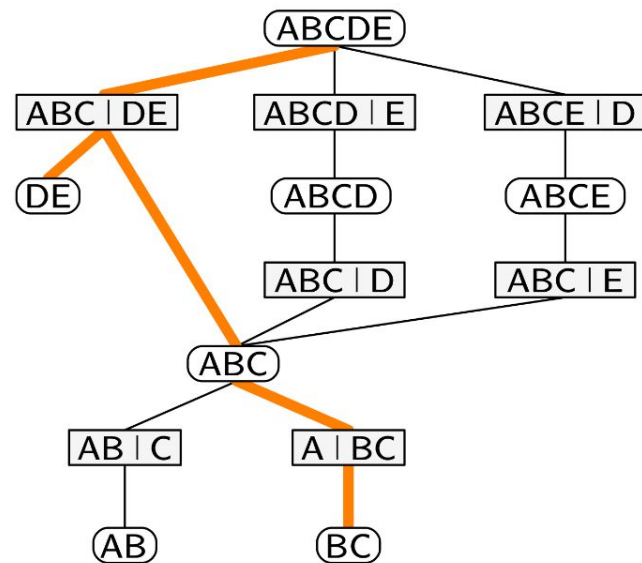
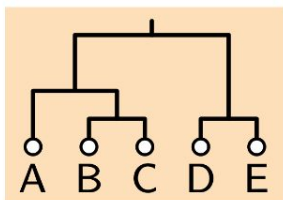
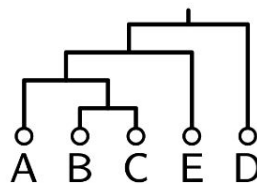
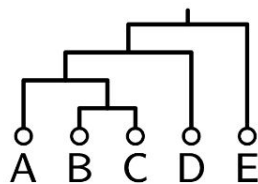
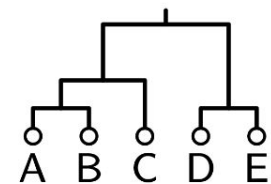
sampled trees



# CCD graph & CCD1

- Based on observed **clade splits**

sampled trees

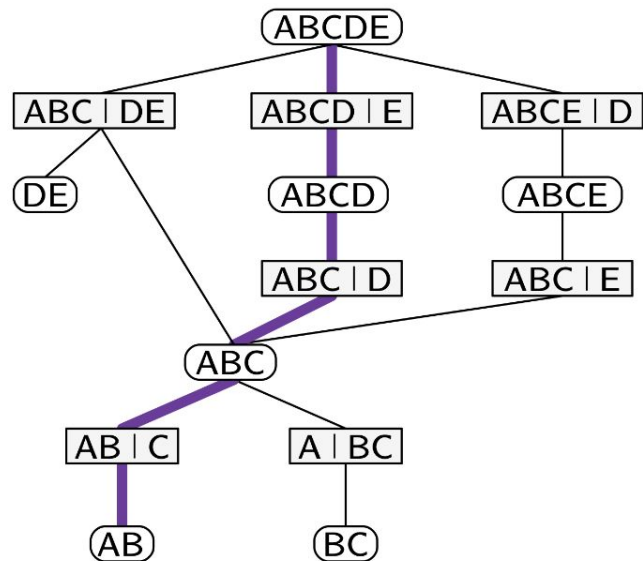
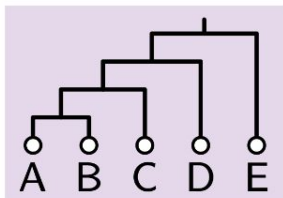
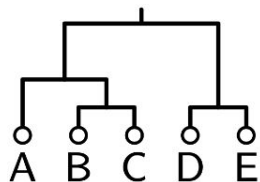
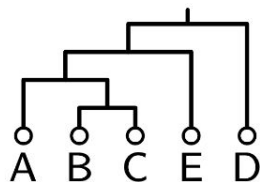
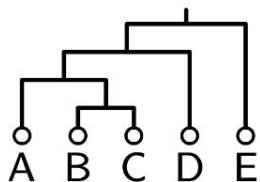
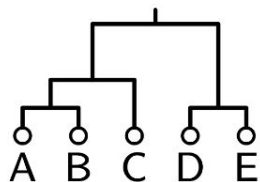




# CCD graph & CCD1

- Based on observed **clade splits**

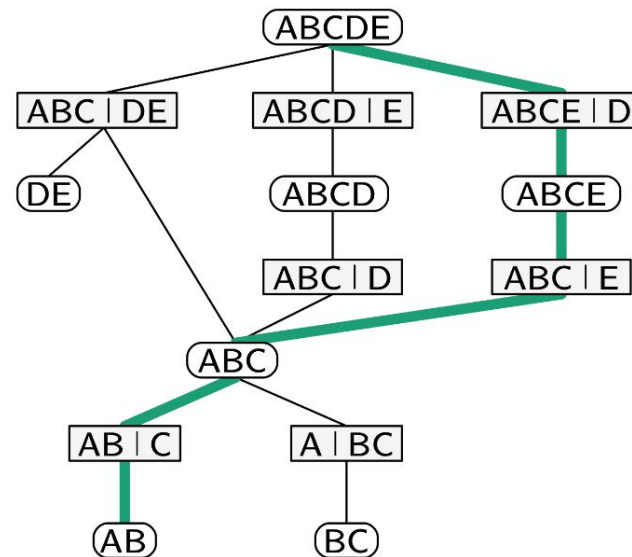
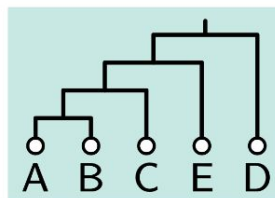
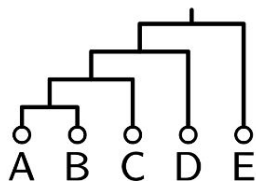
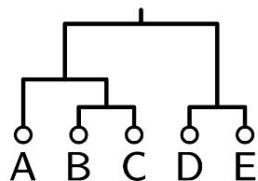
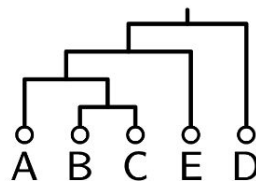
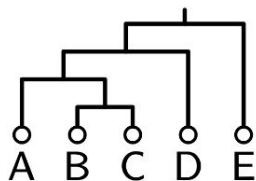
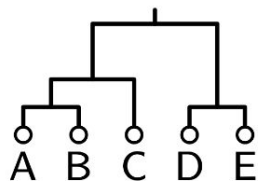
sampled trees



# CCD graph & CCD1

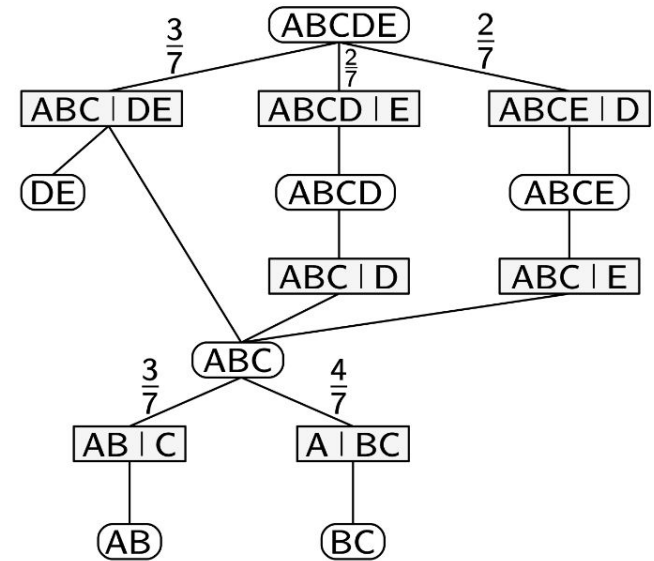
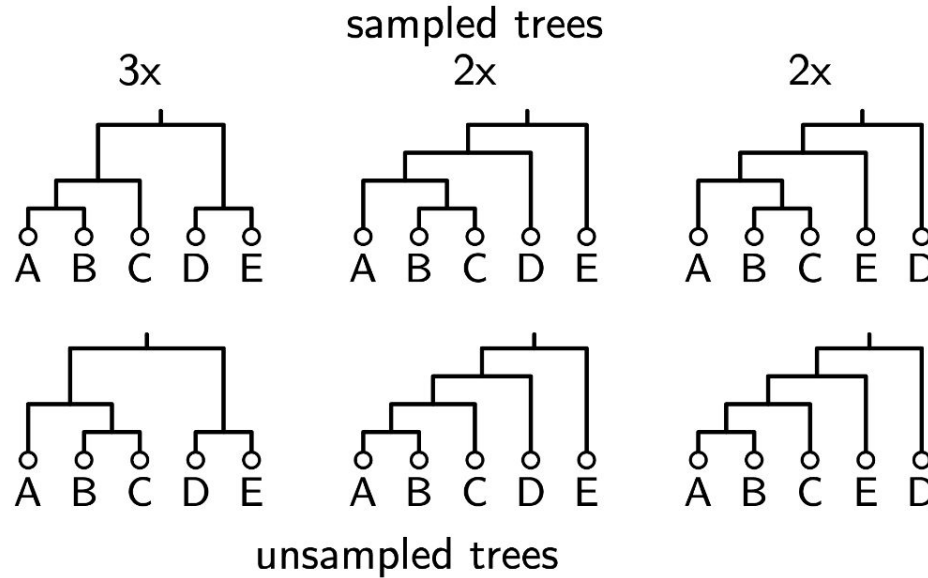
- Based on observed **clade splits**

sampled trees



# CCD graph & CCD1

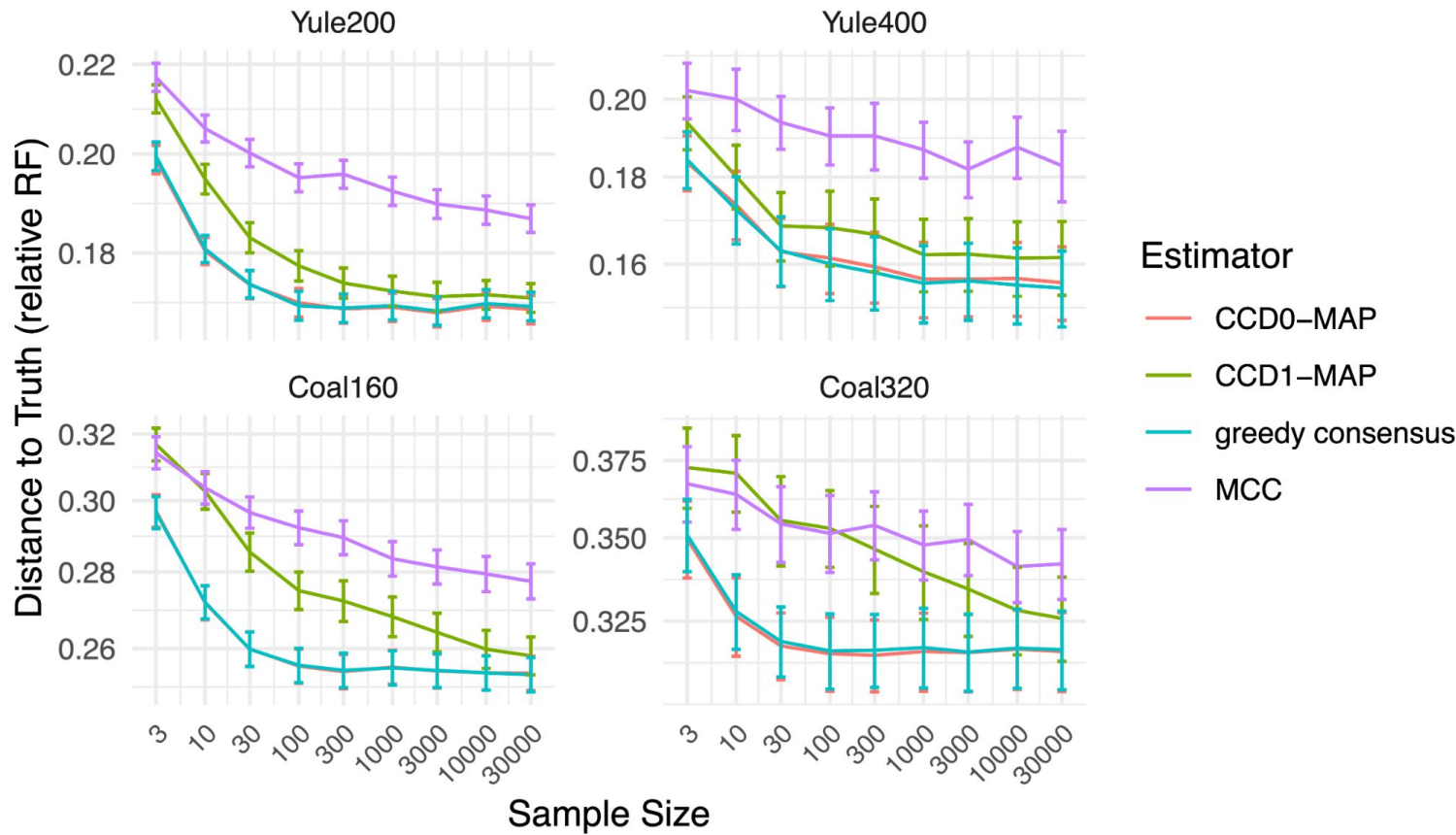
- Based on observed **clade splits**



- Tree probabilities ...

# Accuracy of CCD MAP trees

Mean distance to truth vs. Sample Size for Different Estimators



Phylogenetic tree of CCD0 strains. The tree is rooted at the top left with AUSA2s81. It shows a large cluster of strains on the right, including WV2780s79, WV1988s87, WV1234s84, and many others. The tree is color-coded by region: USA (blue), Europe (orange), Africa (green), Asia (red), and Oceania (purple). The tree is labeled with strain names and accession numbers.

**MCC**

USALongs56

AUSA2s81

WV2780s79

WV10083s87

BE131s90

BE138s90

MON1s90

MON1s92

CH34s94

BIR1734s89

WV6973s82

BE8073s92

BE1010s93

BE1011s92

BE2504s85

WV12342s84

MON1s92

BE13280s98

BE13281s98

BE13282s98

BE13283s98

BE13284s98

BE13285s98

BE13286s98

BE13287s98

BE13288s98

BE13289s98

BE13290s98

BE13291s98

BE13292s98

BE13293s98

BE13294s98

BE13295s98

BE13296s98

BE13297s98

BE13298s98

BE13299s98

BE13300s98

BE13301s98

BE13302s98

BE13303s98

BE13304s98

BE13305s98

BE13306s98

BE13307s98

BE13308s98

BE13309s98

BE13310s98

BE13311s98

BE13312s98

BE13313s98

BE13314s98

BE13315s98

BE13316s98

BE13317s98

BE13318s98

BE13319s98

BE13320s98

BE13321s98

BE13322s98

BE13323s98

BE13324s98

BE13325s98

BE13326s98

BE13327s98

BE13328s98

BE13329s98

BE13330s98

BE13331s98

BE13332s98

BE13333s98

BE13334s98

BE13335s98

BE13336s98

BE13337s98

BE13338s98

BE13339s98

BE13340s98

BE13341s98

BE13342s98

BE13343s98

BE13344s98

BE13345s98

BE13346s98

BE13347s98

BE13348s98

BE13349s98

BE13350s98

BE13351s98

BE13352s98

BE13353s98

BE13354s98

BE13355s98

BE13356s98

BE13357s98

BE13358s98

BE13359s98

BE13360s98

BE13361s98

BE13362s98

BE13363s98

BE13364s98

BE13365s98

BE13366s98

BE13367s98

BE13368s98

BE13369s98

BE13370s98

BE13371s98

BE13372s98

BE13373s98

BE13374s98

BE13375s98

BE13376s98

BE13377s98

BE13378s98

BE13379s98

BE13380s98

BE13381s98

BE13382s98

BE13383s98

BE13384s98

BE13385s98

BE13386s98

BE13387s98

BE13388s98

BE13389s98

BE13390s98

BE13391s98

BE13392s98

BE13393s98

BE13394s98

BE13395s98

BE13396s98

BE13397s98

BE13398s98

BE13399s98

BE13400s98

BE13401s98

BE13402s98

BE13403s98

BE13404s98

BE13405s98

BE13406s98

BE13407s98

BE13408s98

BE13409s98

BE13410s98

BE13411s98

BE13412s98

BE13413s98

BE13414s98

BE13415s98

BE13416s98

BE13417s98

BE13418s98

BE13419s98

BE13420s98

BE13421s98

BE13422s98

BE13423s98

BE13424s98

BE13425s98

BE13426s98

BE13427s98

BE13428s98

BE13429s98

BE13430s98

BE13431s98

BE13432s98

BE13433s98

BE13434s98

BE13435s98

BE13436s98

BE13437s98

BE13438s98

BE13439s98

BE13440s98

BE13441s98

BE13442s98

BE13443s98

BE13444s98

BE13445s98

BE13446s98

BE13447s98

BE13448s98

BE13449s98

BE13450s98

BE13451s98

BE13452s98

BE13453s98

BE13454s98

BE13455s98

BE13456s98

BE13457s98

BE13458s98

BE13459s98

BE13460s98

BE13461s98

BE13462s98

BE13463s98

BE13464s98

BE13465s98

BE13466s98

BE13467s98

BE13468s98

BE13469s98

BE13470s98

BE13471s98

BE13472s98

BE13473s98

BE13474s98

BE13475s98

BE13476s98

BE13477s98

BE13478s98

BE13479s98

BE13480s98

BE13481s98

BE13482s98

BE13483s98

BE13484s98

BE13485s98

BE13486s98

BE13487s98

BE13488s98

BE13489s98

BE13490s98

BE13491s98

BE13492s98

BE13493s98

BE13494s98

BE13495s98

BE13496s98

BE13497s98

BE13498s98

BE13499s98

BE13500s98

BE13501s98

BE13502s98

BE13503s98

BE13504s98

BE13505s98

BE13506s98

BE13507s98

BE13508s98

BE13509s98

BE13510s98

BE13511s98

BE13512s98

BE13513s98

BE13514s98

BE13515s98

BE13516s98

BE13517s98

BE13518s98

BE13519s98

BE13520s98

BE13521s98

BE13522s98

BE13523s98

BE13524s98

BE13525s98

BE13526s98

BE13527s98

BE13528s98

BE13529s98

BE135

- Check trees for convergence!
- Don't use the MCC tree!

Thank you!

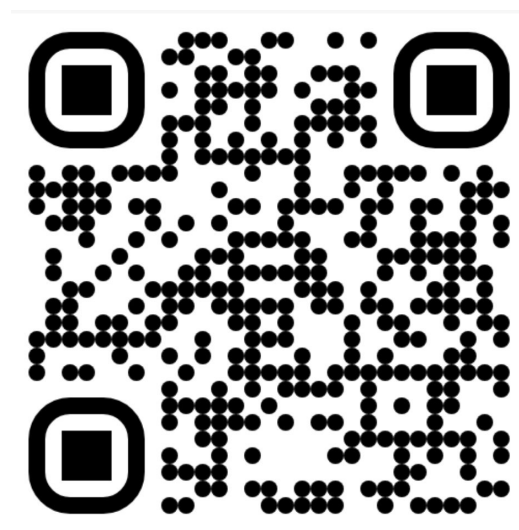
### ASM package



"An automated convergence diagnostic for phylogenetic MCMC analyses."

*IEEE/ACM Transactions on Computational Biology and Bioinformatics* (2024)

### CCD package



"Accurate Bayesian phylogenetic point estimation using a tree distribution parameterized by clade probabilities." *PLOS Computational Biology* (2025)