ALLAN WILSON CENTRE

Massey University
COLLEGE OF SCIENCES

Lineage Specific Evolution

Using the best model!

Reconstructing the worst tree…

Liat Shavit Grievink

Doctoral Candidate
Allan Wilson Center for Molecular Ecology and Evolution
And
Institute of Fundamental Sciences
Massey University
Palmerston North

Te Kunenga
ki Pūrehuroa

Unmodelled biochemical realism

Evolution: a homogeneous, time reversible, stationary process?

As sequences diverge they acquire their own properties.

The proportion of variable sites can evolve in a lineage-specific manner.

LineageSpecificSeqgen

An extention of seqgen & seqgen-amino-cov.

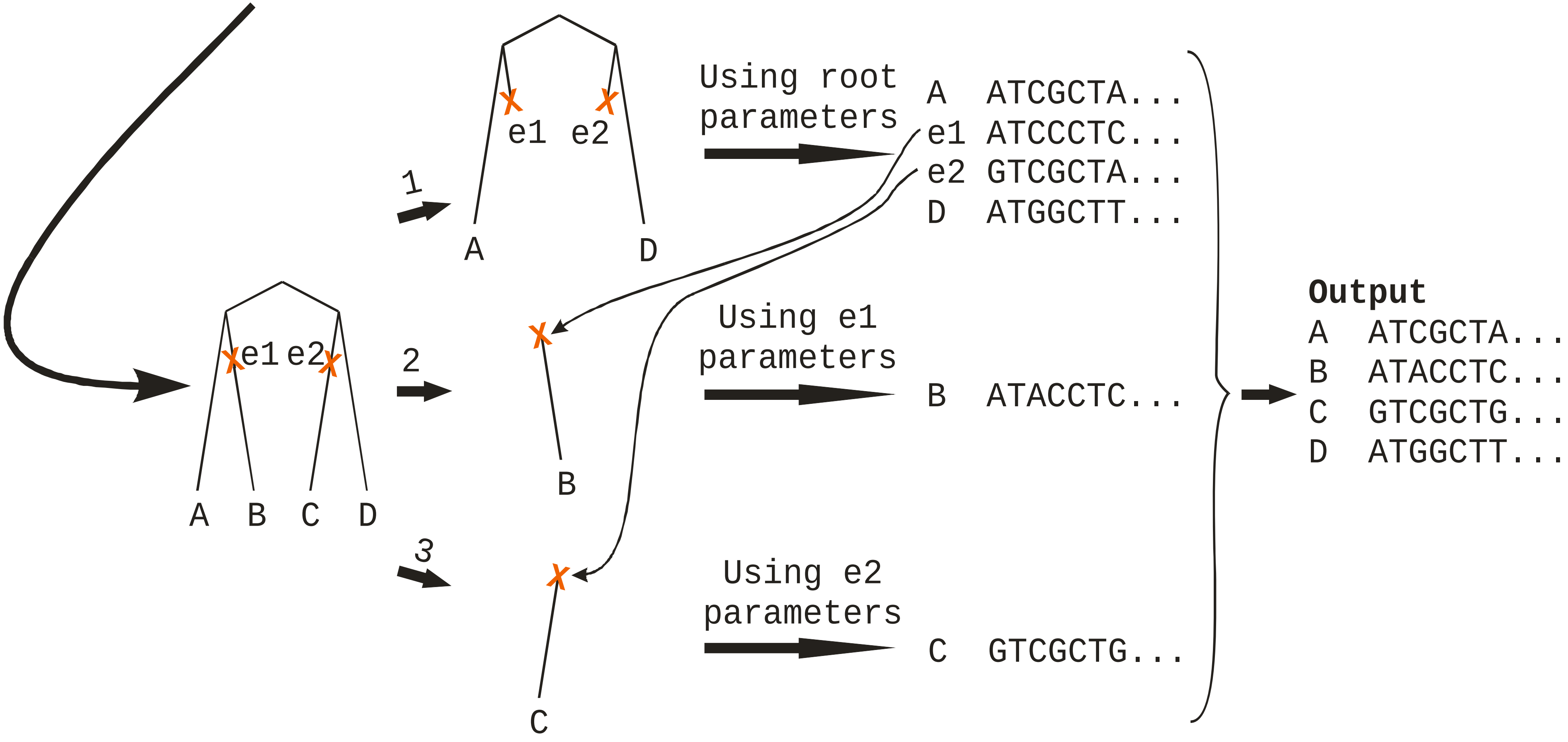
Allows generation of sequence data containing changes in Pvar and changes in the on/off switch rate of the covarion model.

Available from http://awcmee.massey.ac.nz/downloads.htm.

LineageSpecificSeqgen cont.

Input tree:

`((A:0.2,B:0.1$e1:0.1):0.01,(c:0.1$e2:0.1,D:0.2):0.01);`



Using root parameters

```
A   ATCGCTA...
e1  ATCCCTC...
e2  GTCGCTA...
D   ATGGCTT...
```

Using e1 parameters

```
B   ATACCTC...
```

Using e2 parameters

```
C   GTCGCTG...
```

Output

```
A   ATCGCTA...
B   ATACCTC...
C   GTCGCTG...
D   ATGGCTT...
```

Rescaling

How do we define the branch lengths?

In seqgen branch lengths are rescaled to compensate for invariable sites.

I=0.8, branch length=0.4 ➡ rescaled=2

We have changed the default, but scaling is still an option.

A simulation study

100 data sets

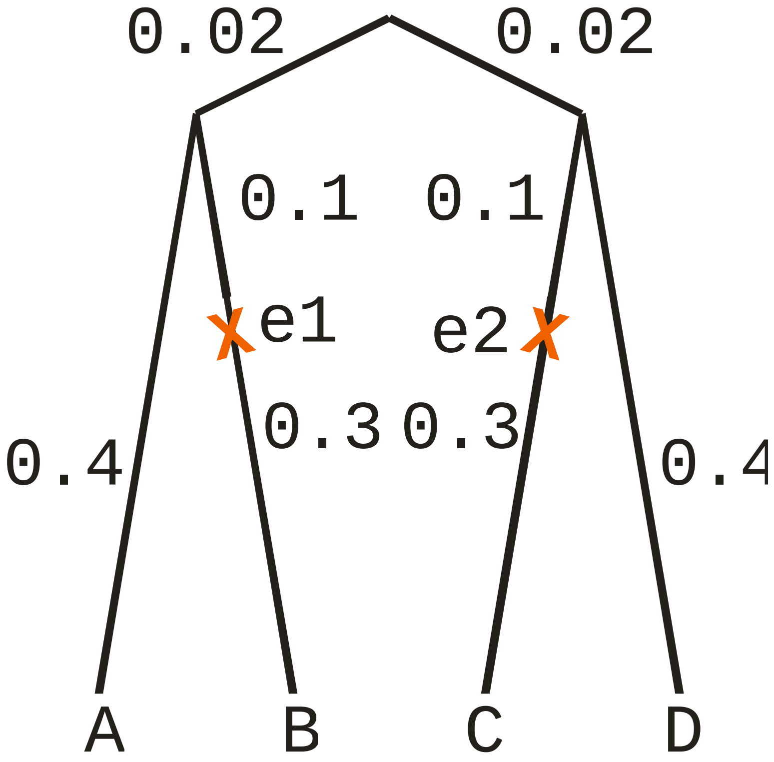
10000 bp

JC model

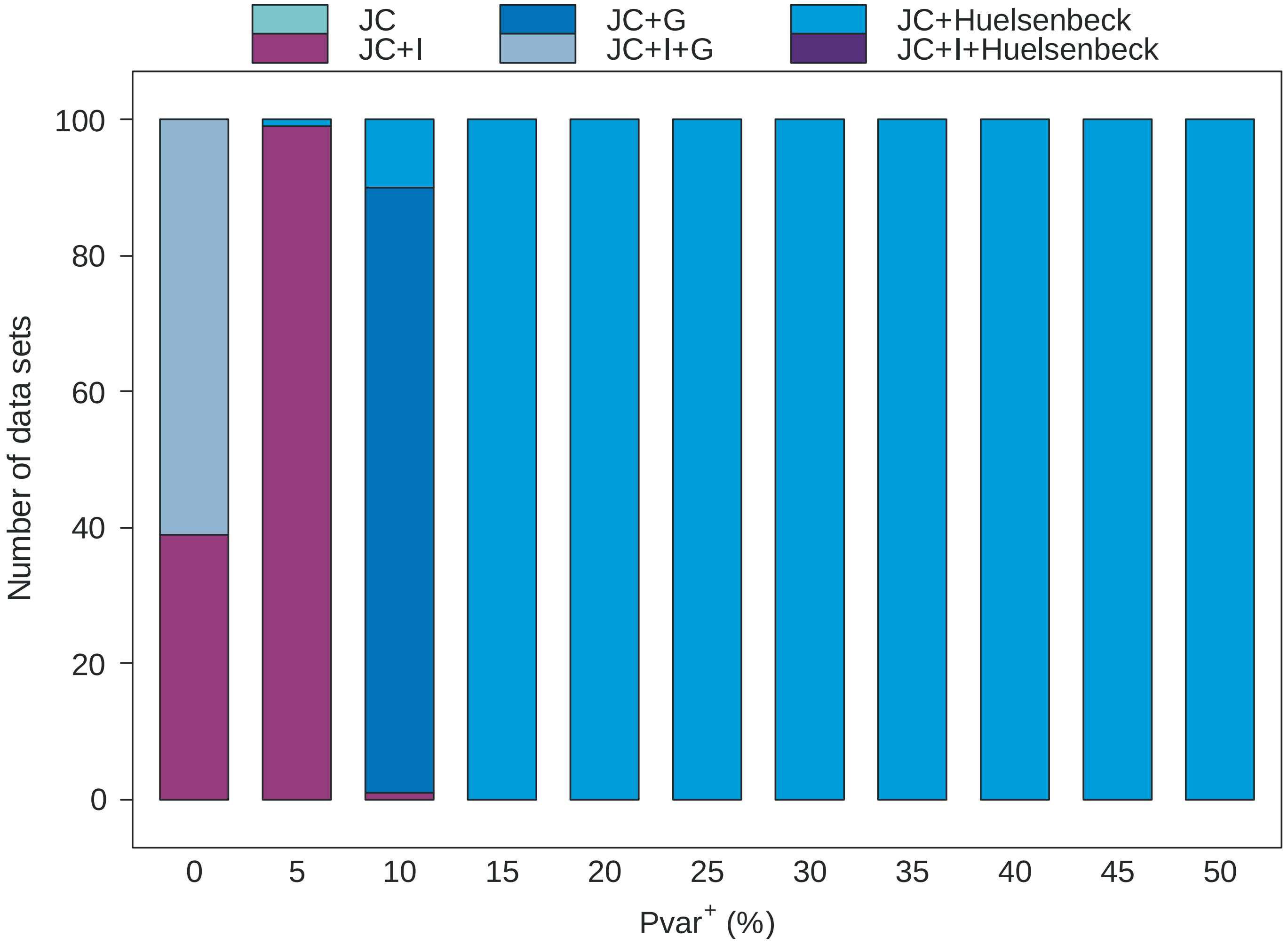
I = 0.8

Pvar$^+$ = (0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5)

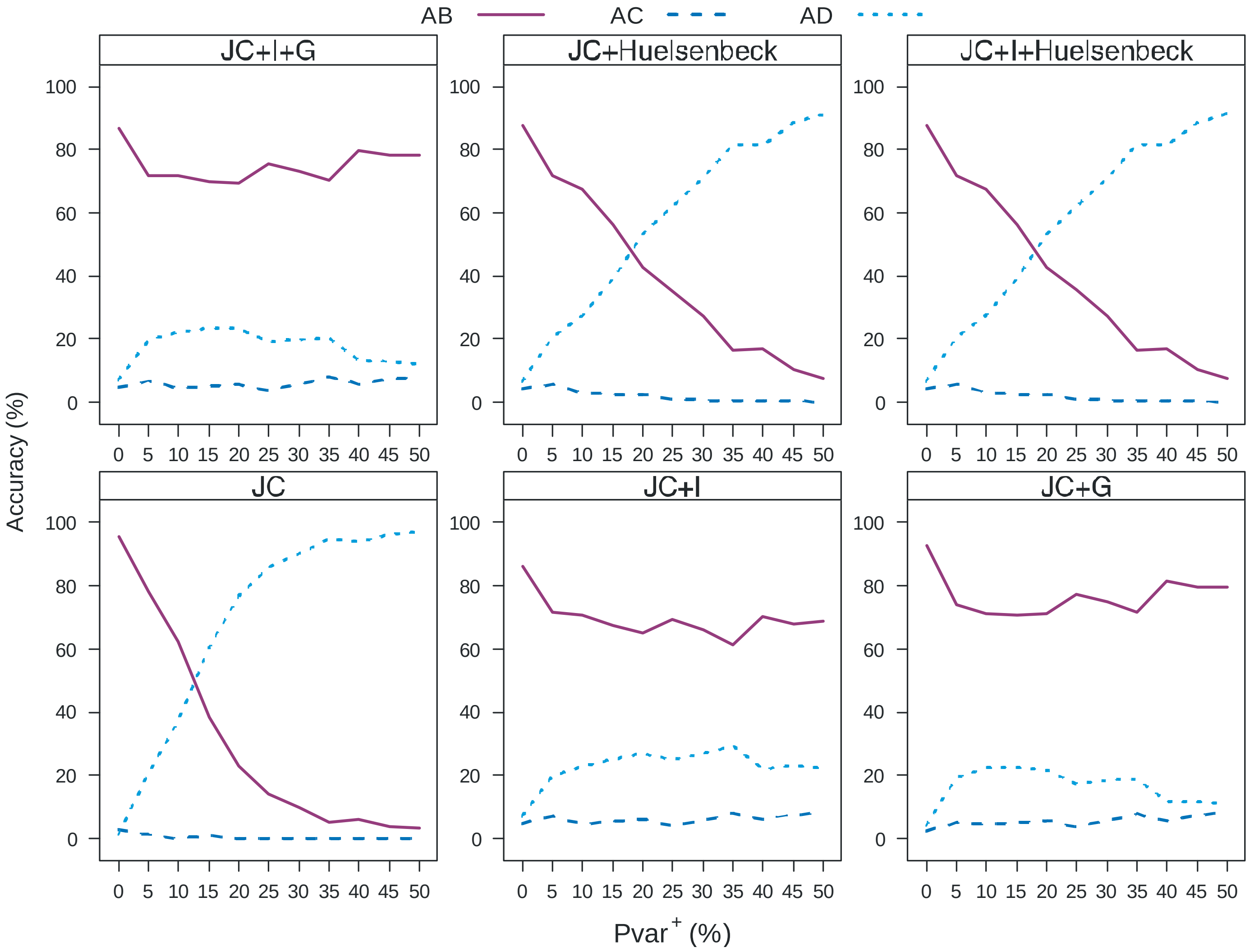
Covarion on = 0.6
switch rate = 0.1



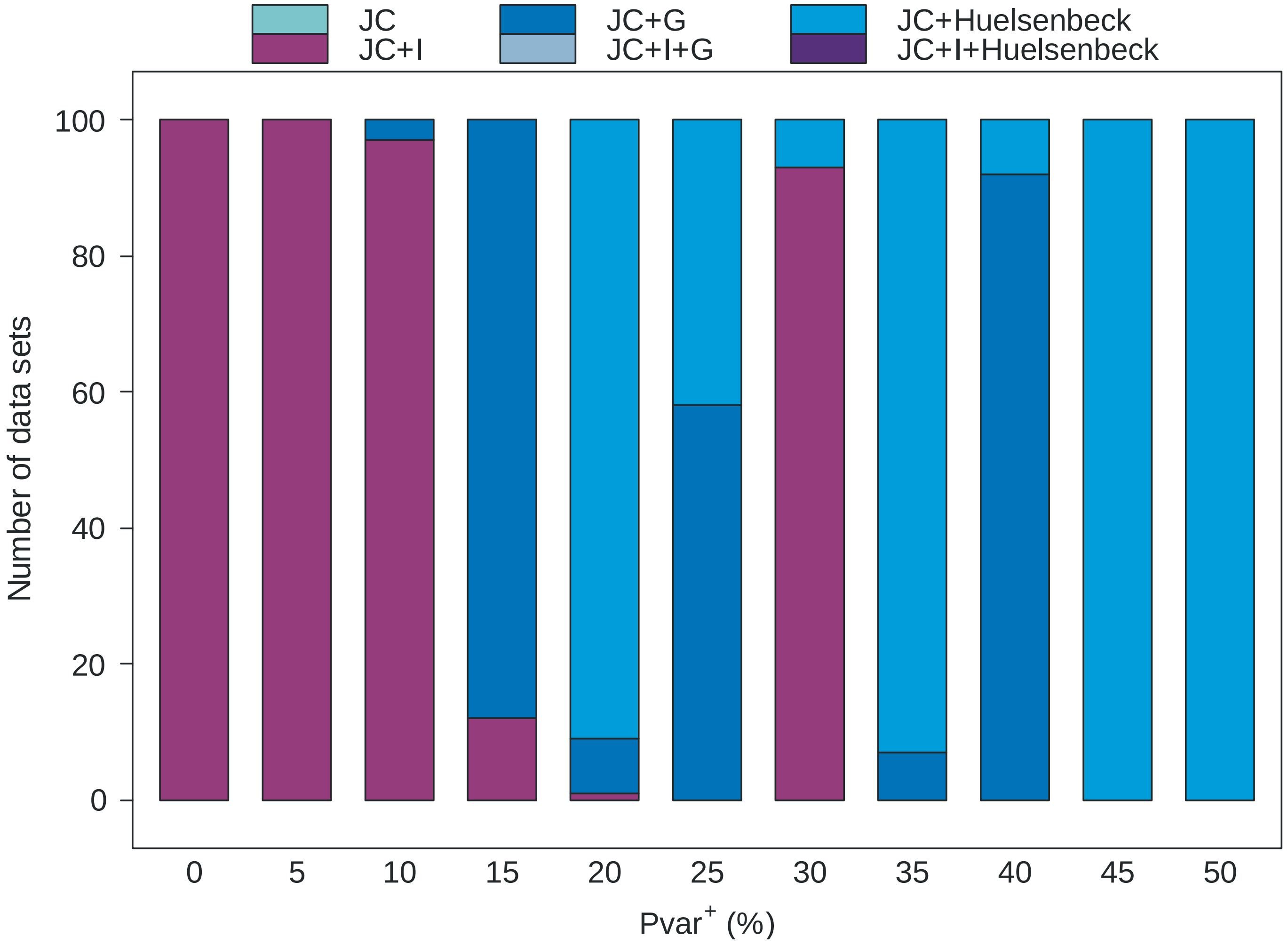
Model fit
no rescaling no covarion



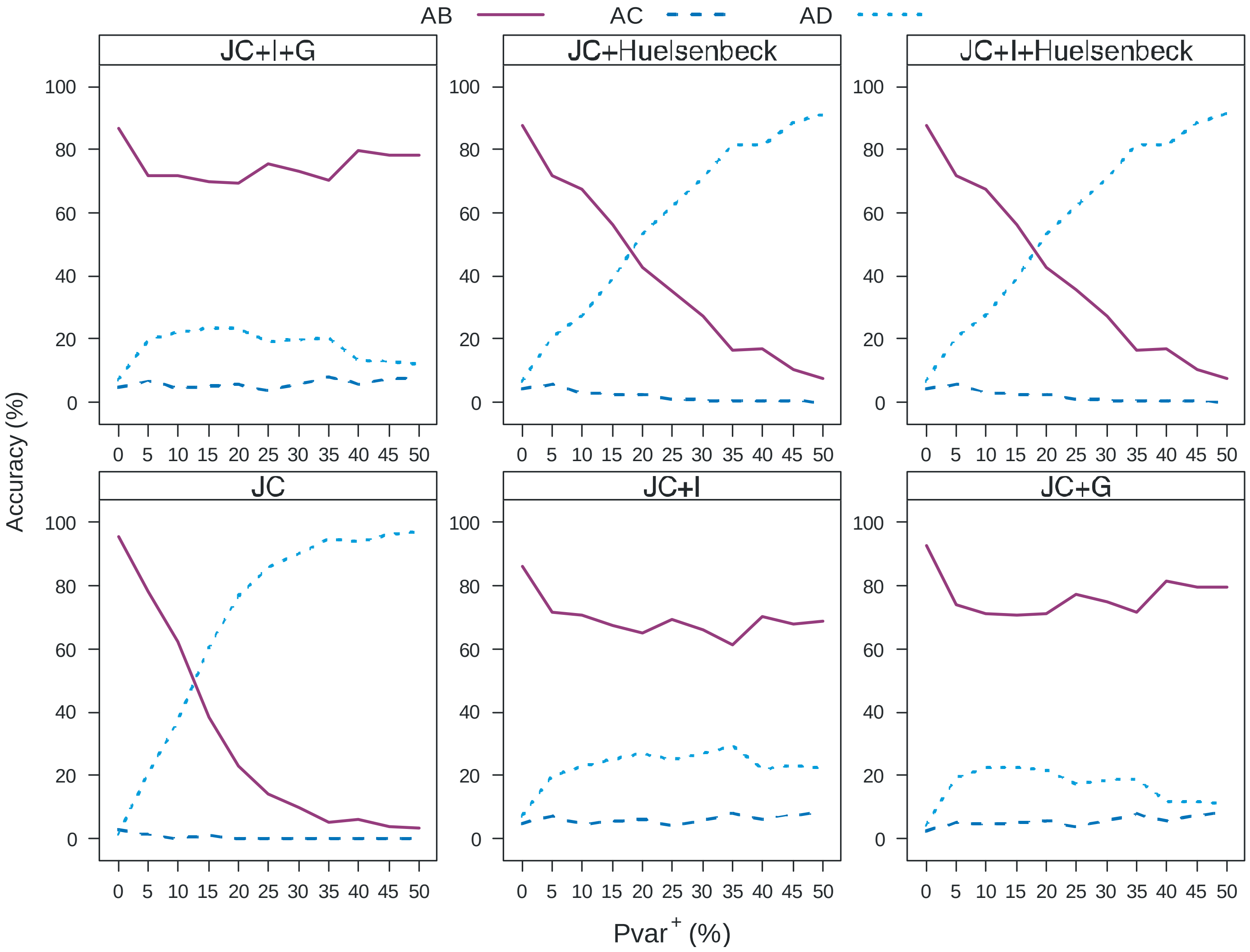
Reconstruction accuracy no rescaling no covarion



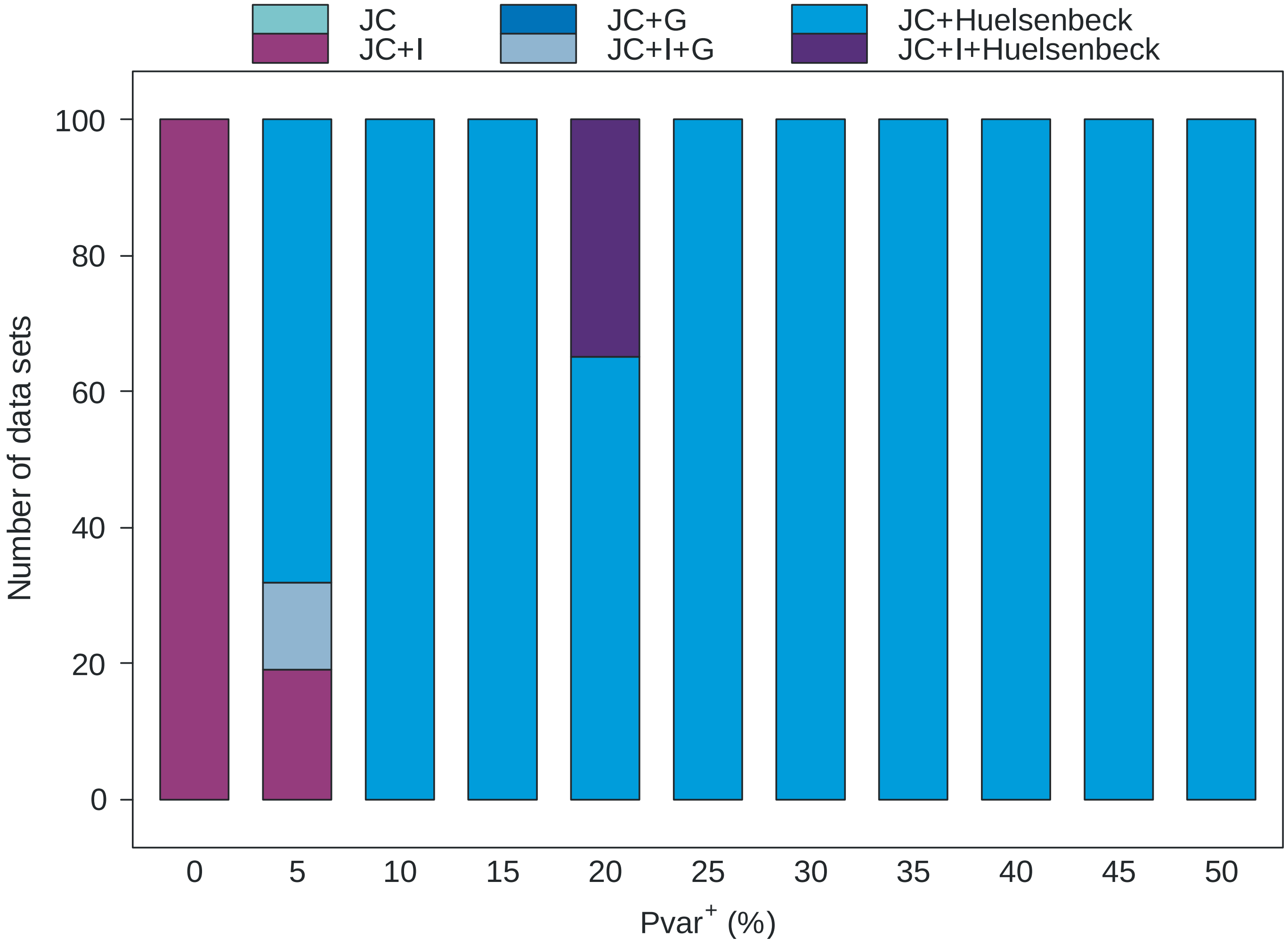
Model fit
no rescaling with covarion



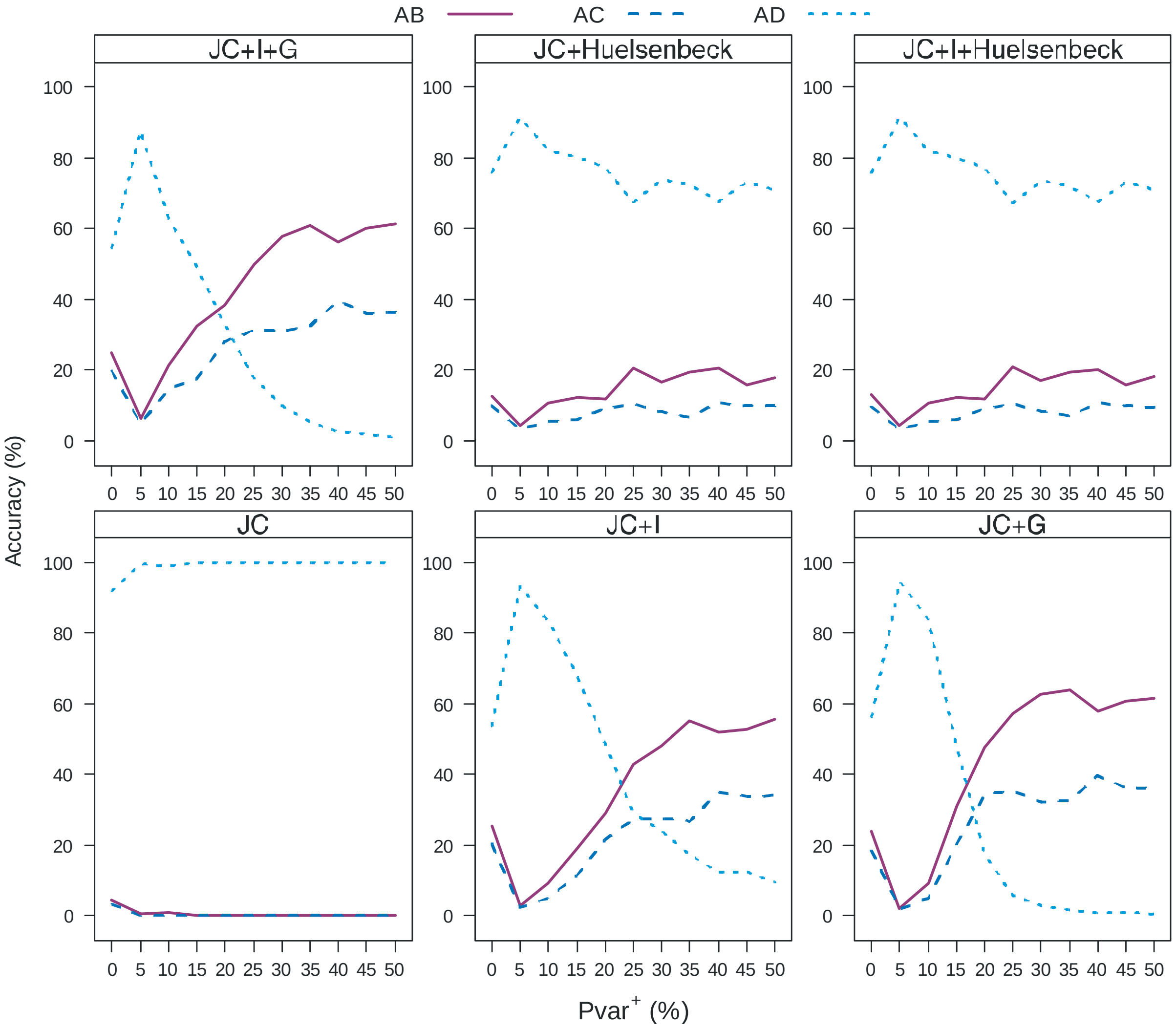
Reconstruction accuracy
no rescaling with covarion



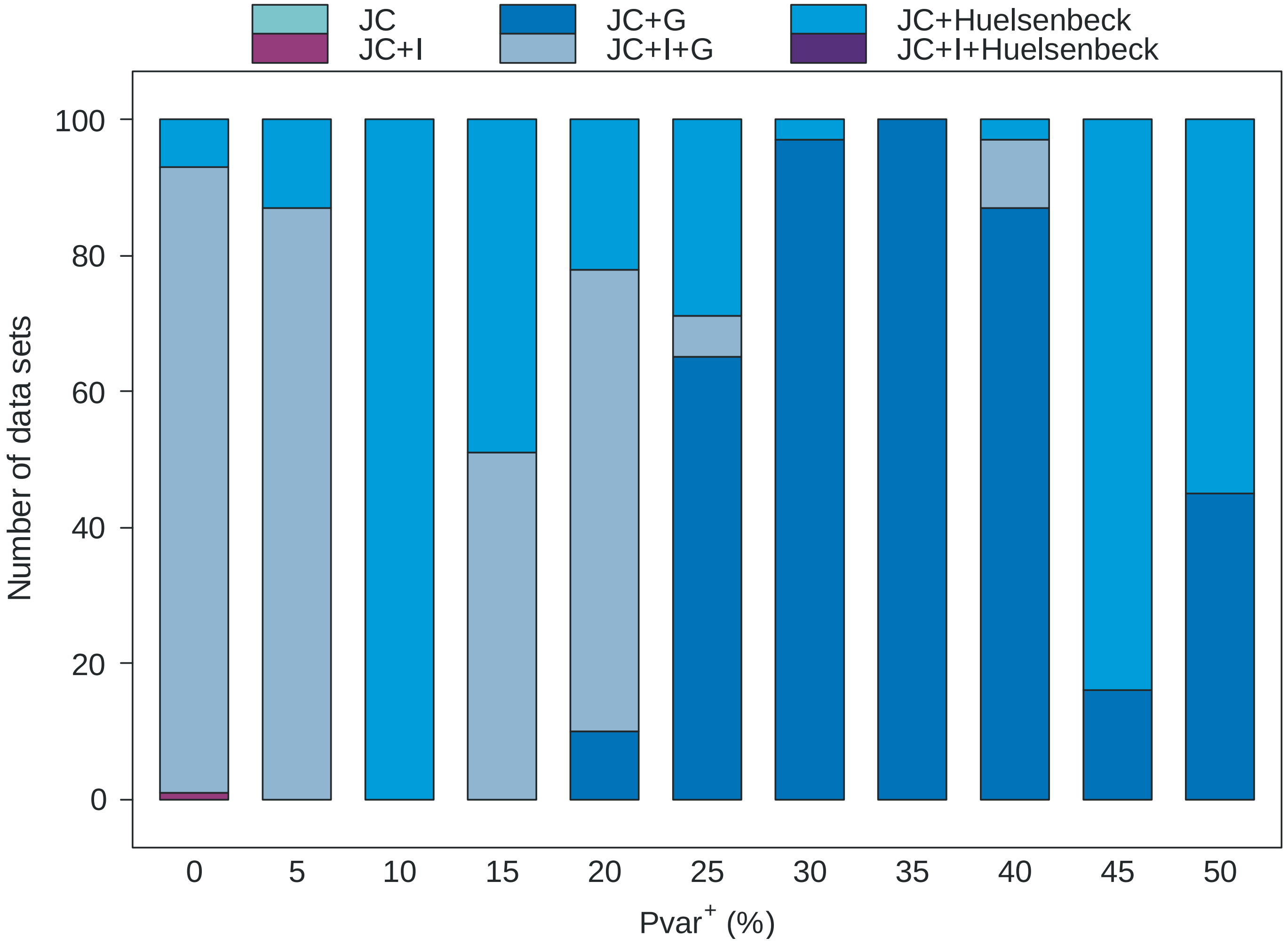
Model fit
with rescaling no covarion



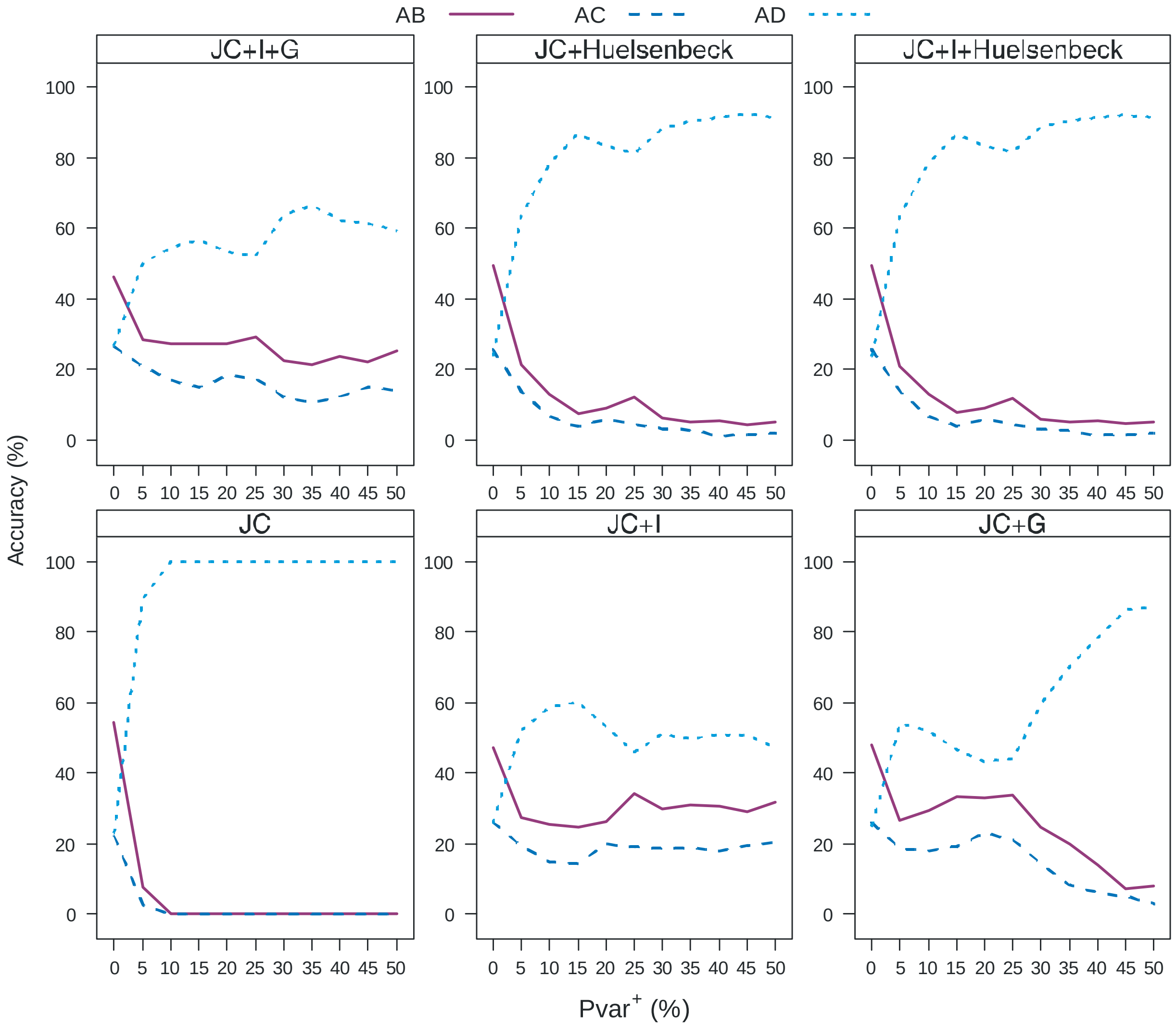
Reconstruction accuracy with rescaling no covarion



Model fit
with rescaling with covarion



Reconstruction accuracy with rescaling with covarion



Conclusion

Tree reconstruction, even when using the best fit model, is hindered by change in Pvar and may result in the wrong tree.

There is a need for more adequate and biochemically realistic models for sequence evolution.

Thanks

Barbara Holland
David Penny
Mike Hendy
Pete Lockhart