Sc2ts vs Usher cophylogenies

To plot cophylogenies, we aim to find identical "representative samples" for Pango lineages that exist in both the Usher tree and the sc2ts tree, and have the same Pango assignment in both.

However, the earliest sample of each Pango type could be an erroneous classication. To avoid this, we identify "originating nodes" for each pango. An originating node of (say) B.1.1.7 is the earliest node that has > 50% of the B.1.1.7 samples as descendants, and which itself is labelled B.1.1.7. As this is an ARG, there are many trees: we count the maximum number of samples in any tree. To find a representative sample, we pick the oldest descendant sample node of the origination node which has entirely B.1.1.7 samples as descendants in a tree.

To reduce the number of tips to compare, we also remove samples which are known Pango-X recombinants (and descendants of them).

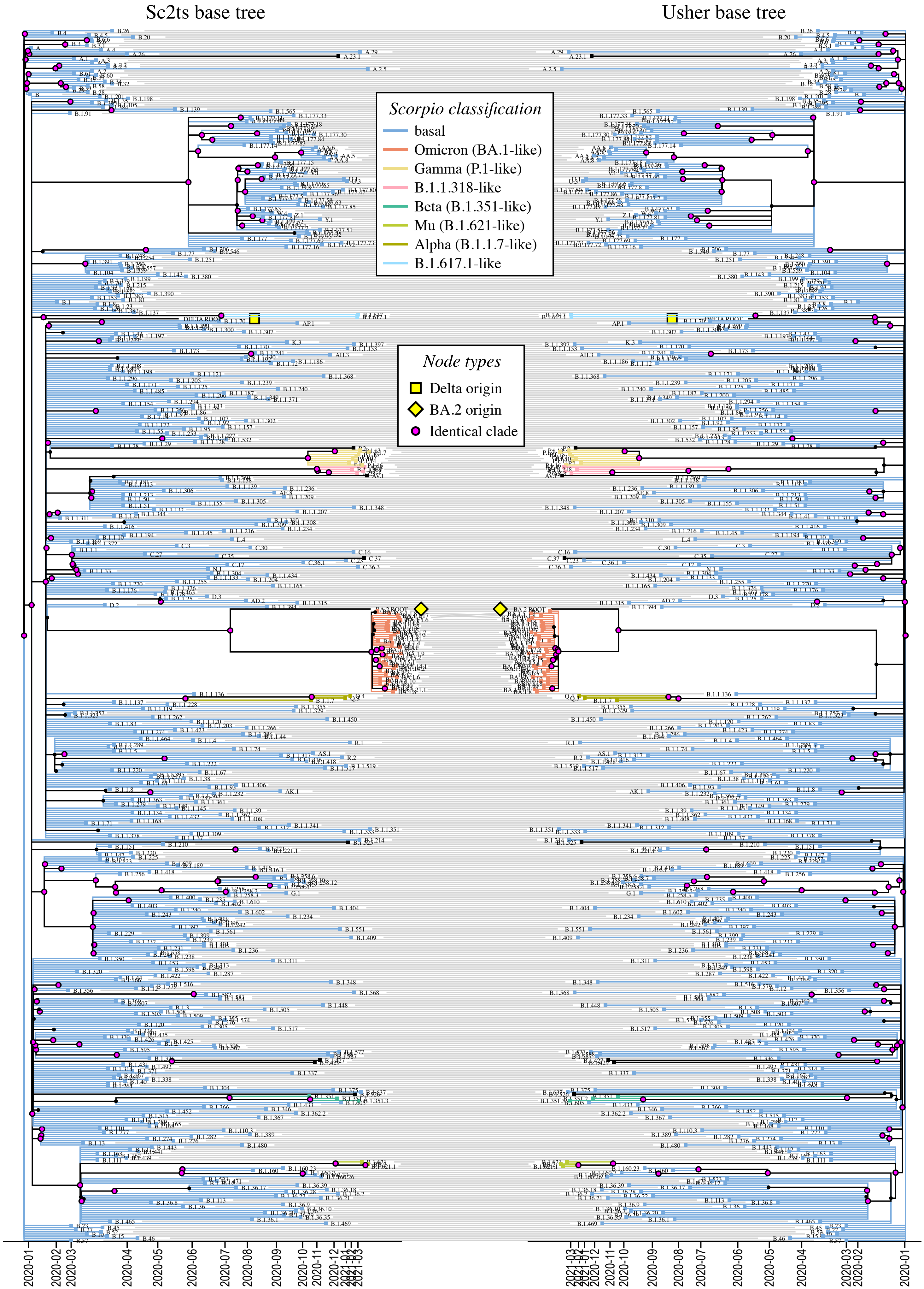
```
Using a 434.8 megabyte ARG up to 2023-02-21, with 2482157 sampled SARS-CoV2 sequences
(317 trees, 2484587 mutations over 29904.0bp with 855 recomb. events)
ARG has 2057 pango designations

Found 1986 Pango sample nodes out of 2057 pango groups
The following 71 pango designations were omitted due to having no focal pango origination node
with > 50.0% focal pango descendants:
 BF.38, BA.5.2.6, BA.5.2.12, EE.3, B.33, BA.1.1.9, BA.4.8, BQ.1.11.1, XBB.1.11, B.1.1.142, BA.
5.2.18, AY.112, BA.1.13.1, B.1.1.148, BA.2.75.7, BA.5.2.20, AY.4.3, BQ.1.1.46, XBB.1.5.96, BA.
2.24, BA.2.45, BF.7, BA.2.25, Q.1, XBB.1.5.37, BA.2.27, BA.5.2.29, BQ.1.1.53, BQ.1.22, B.1.560,
BA.2.9.3, BF.7.16, XBB.1.5.47, B.1.280, BA.2.9.5, BF.7.19, BA.2.3.13, BA.2.56, BA.5.2.36, B.1.
1.174, BA.2.3.16, BA.5.2.38, B.1.1.118, B.1.572, BA.2.6, B.1.1.177, XBB.1.5.100, BA.5.2.41, BQ.
1.1.21, XBB.1.5.59, AM.1, B.1.302, BQ.1.1.66, XAC, BA.2.63, XAD, B.1.1.122, XBB.1.5.12, B.1.1.1
89, B.1.586, B.1.1.34, CA.2, B.1.36.38, BA.1.1.16, B.1.1.268, B.1.590, BA.5.2.55, BF.7.9, XBB.
1.5.2, XBB.1.5.73, XBF.2
```

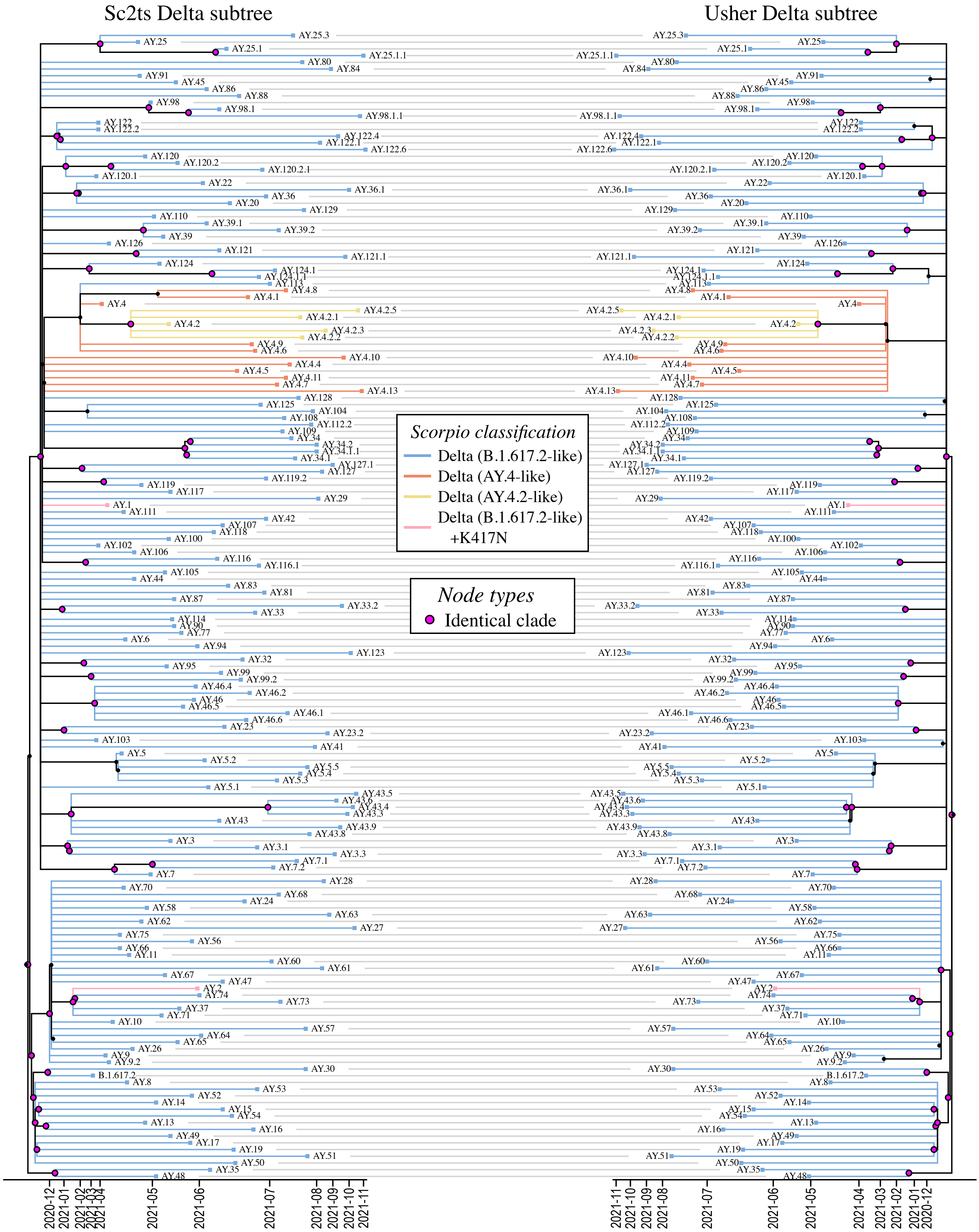
We also treat descendants of the 1030562 fake "recombinant" (see https://github.com/jeromekelleher/sc2ts-paper/discussions/672) as non-recombining, by trimming away the RHS from 27382.0 onwards.

By comparing sample ids with the UShER tree, we find 1749 shared Pango-representative samples

Finally, to create manageably-sized phylogeny plots, we restrict to Pango lineages accounting for >= 10 samples, and break the cophylogenies down into a base tree and several subtrees.

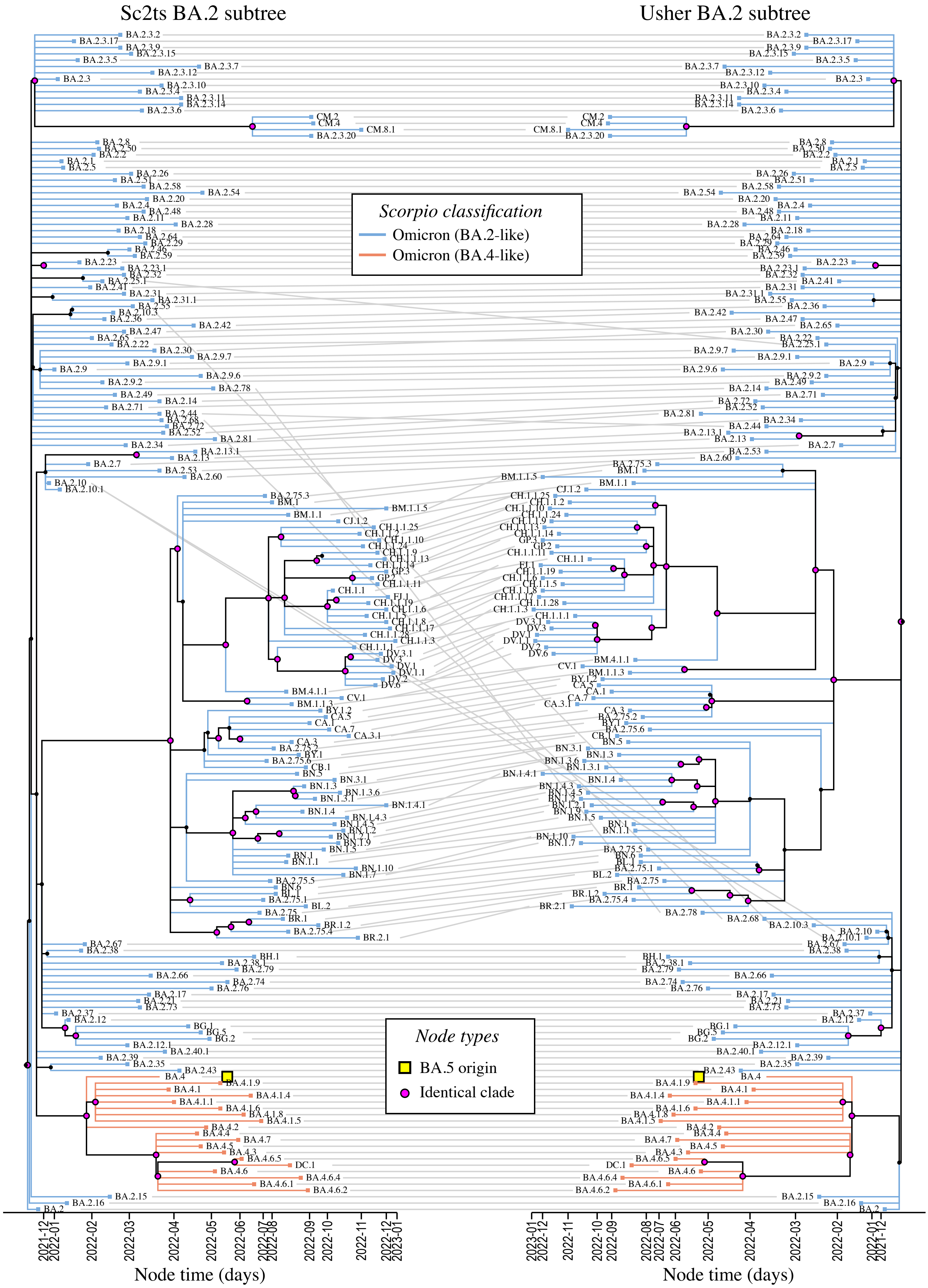
In the following cophylogeny tanglegram plots, we have used *dendroscope* (Huson and Scornavacca, DOI:10.1093/sysbio/sys062) to untangle the trees. We have found this to be the most effective software for untangling trees with polytomies.


Sc2ts base tree
Usher base tree
Scorpio classification
basal
Omicron (BA.1-like)
Gamma (P.1-like)
B.1.1.318-like
Beta (B.1.351-like)
Mu (B.1.621-like)
Alpha (B.1.1.7-like)
B.1.617.1-like
Node types
Delta origin
BA.2 origin
Identical clade
DELTA ROOT
BA.2 ROOT
2020-01
2020-02
2020-03
2020-04
2020-05
2020-06
2020-07
2020-08
2020-09
2020-10
2020-11
2020-12
2021-01
2021-02
2021-03



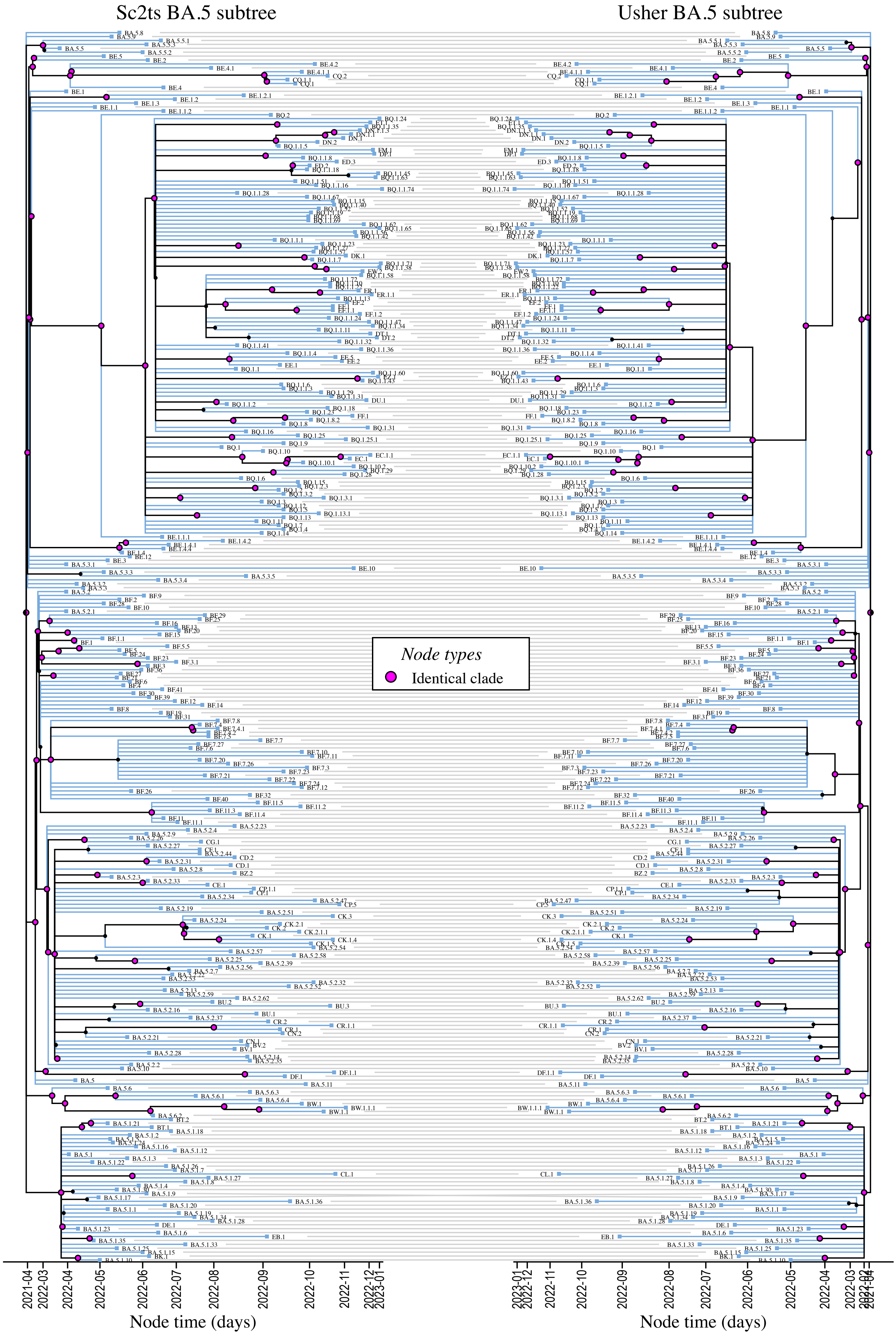
Sc2ts Delta subtree

Usher Delta subtree


Sc2ts BA.2 subtree
Usher BA.2 subtree
Scorpio classification
Omicron (BA.2-like)
Omicron (BA.4-like)
Node types
BA.5 origin
Identical clade
2021-12
2022-01
2022-02
2022-03
2022-04
2022-05
2022-06
2022-07
2022-08
2022-09
2022-10
2022-11
2022-12
2023-01
Node time (days)
Node time (days)

Sc2ts BA.5 subtree

Usher BA.5 subtree


Node types
Identical clade
2021-04
2022-03
2022-04
2022-05
2022-06
2022-07
2022-08
2022-09
2022-10
2022-11
2022-12
2023-01
Node time (days)
2023-01
2022-12
2022-11
2022-10
2022-09
2022-08
2022-07
2022-06
2022-05
2022-04
2022-03
2022-02
2021-04
Node time (days)