Figures and figure supplements

Improved characterisation of MRSA transmission using within-host bacterial sequence diversity

Matthew D Hall *et al*

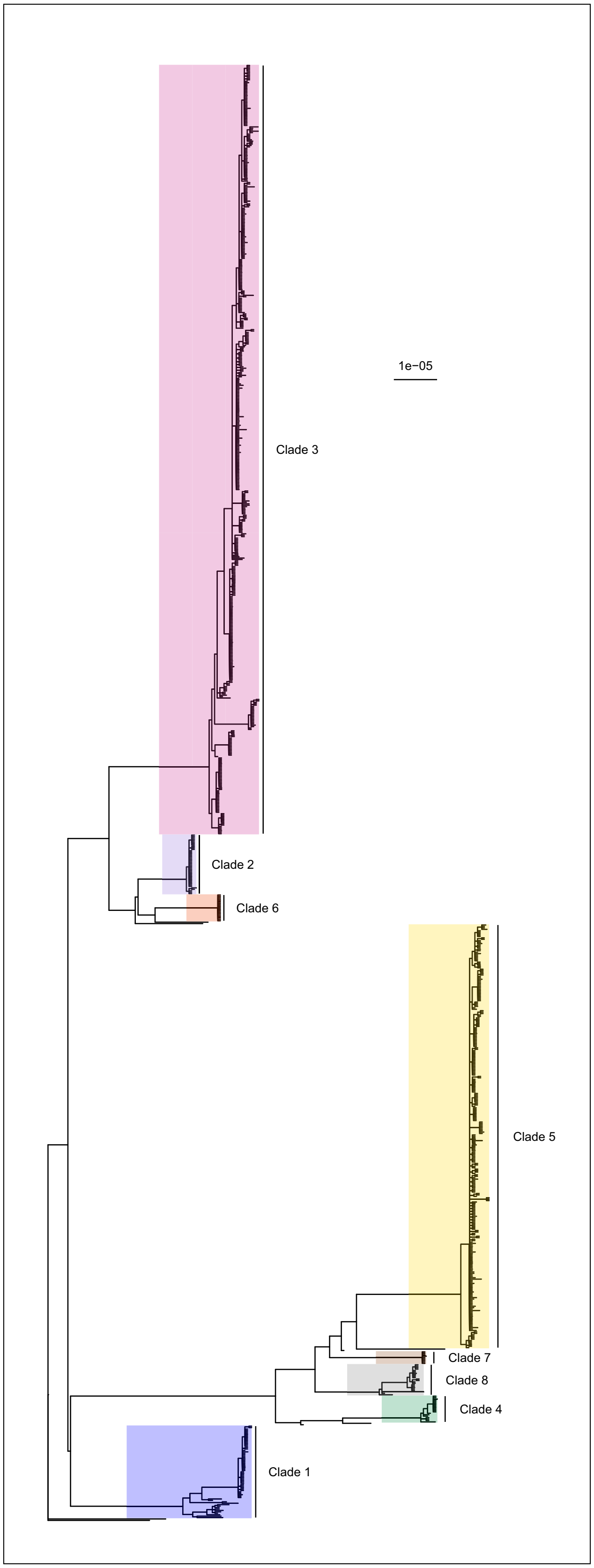

Figure 1. 50% majority-rule consensus tree of the posterior distribution of ExaBayes phylogenies. Branch lengths are in substitutions per site. Eight clades are highlighted. Clades 1 to 5 correspond to those identified by ***Tong et al. (2015)***, while the additional three are newly designated.

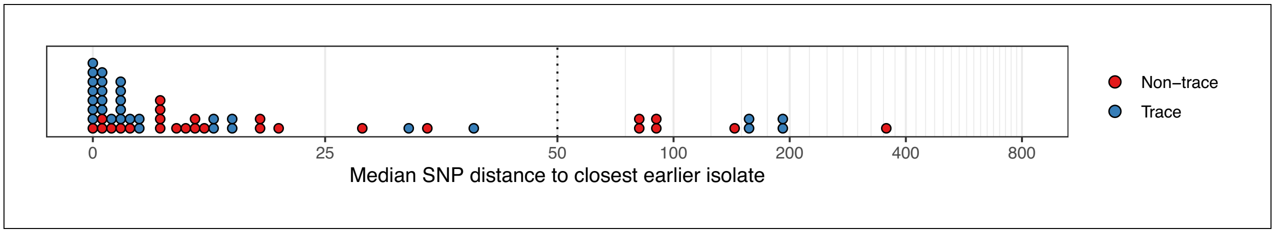

Figure 2. The median number of SNPs separating sequences from each colonisation from the most similar sequence from an isolate acquired before the date of the first positive sample from that colonisation. Blue dots are colonisations isolated from a single swab only (trace colonisations), while red were acquired from colonisations where multiple swabs were isolated. For three colonisations (two trace) the first collection date was the commencement of the study and hence there was no such earlier isolate. The x-axis transfers to a log scale on the right of the dotted line.

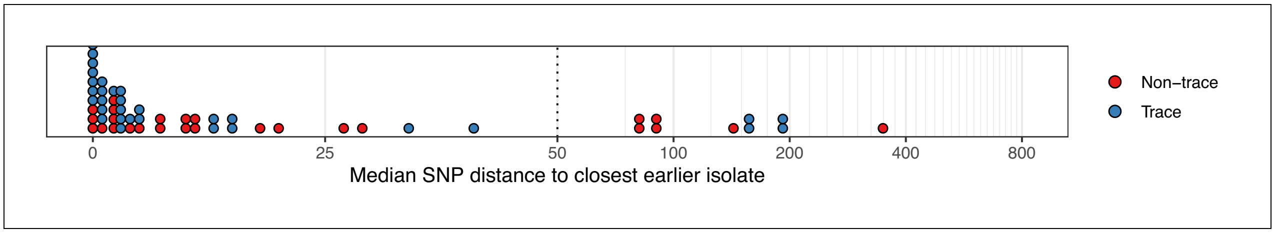

Figure 2—figure supplement 1. The median number of SNPs separating sequences from each colonisation, restricted to those obtained at the time of the first positive swab, from the most similar sequence from an isolate acquired before the date of the first positive sample from that colonisation. Blue dots are colonisations isolated from a single swab only (trace colonisations), while red were acquired from colonisations where multiple swabs were isolated. For three colonisations (two trace) the first collection date was the commencement of the study and hence there was no such earlier isolate. The x-axis transfers to a log scale on the right of the dotted line.

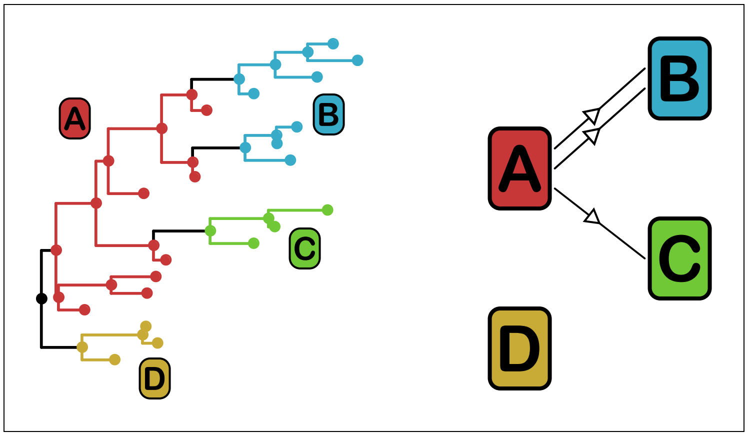

Figure 3. *phyloscanner* identifies transmission pairs by ancestral state reconstruction of hosts (left) and subsequent classification of the topological relationships between the subgraphs reconstructed to each host (right). In this example the hosts are designated (A to D). Here host A is inferred to be the infector of hosts (B and C). The transmission from (A to C) was of only a single pathogen lineage, while that from A to B was of two, with the result that host B has two subgraphs. The subgraph from host D forms a sibling clade to the rest of the phylogeny and, as a result, no inference is made about transmission.

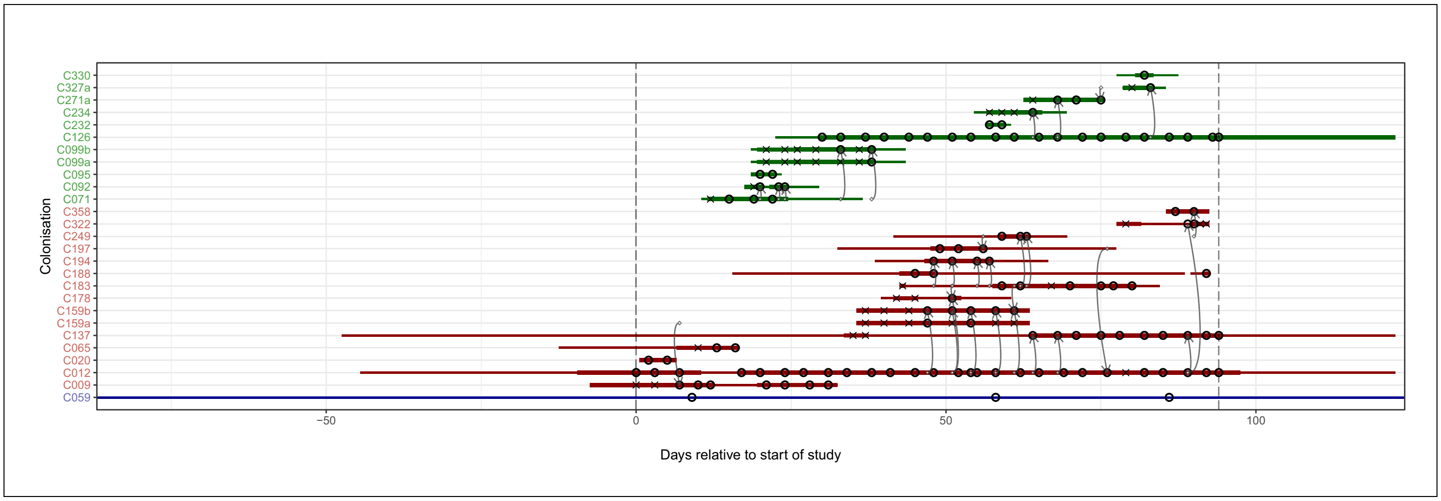

Figure 4. The reconstruction of the transmission process using the consensus tree overlaid on a timeline of hospital and ICU stays and sampling events. Each row represents a colonisation, with thin lines representing the colonised subject's presence in the hospital and thick lines their presence as a patient in an ICU. Colours of the lines and the *y*-axis labels indicate surgical ICU patients (green), paediatric ICU patients (red) and HCWs (blue). Crosses represent times of screens that were negative for MRSA, while circles those that returned positive swabs and sequenced isolates. The grey arrows represent reconstructed transmission events. These appear when at least one subgraph from the recipient is descended from an adjacent subgraph from the infector. Such a transmission may also involve unsampled intermediaries or the environment. The timings of these arrows represent the upper bound for the time at which they could have occurred rather than an exact estimate. The dotted vertical lines demarcate the period of sampling.

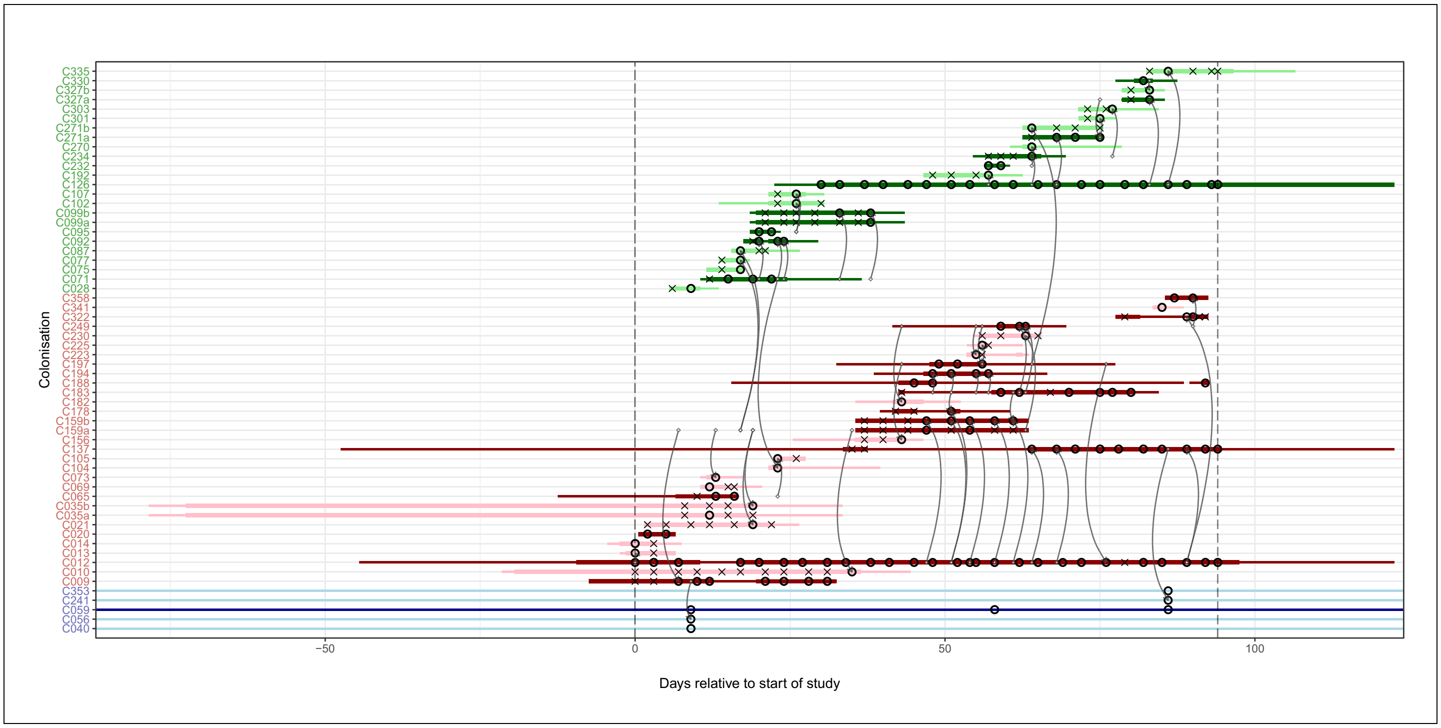

Figure 4—figure supplement 1. The reconstruction of the transmission process using the consensus tree overlaid on a timeline of hospital and ICU stays and sampling events, with trace colonisations included. Each row represents a colonisation, with thin lines representing the colonised subject's presence in the hospital and thick lines their presence as a patient in an ICU. Colours of the lines and the *y*-axis labels indicate surgical ICU patients (green), paediatric ICU patients (red) and HCWs (blue). Light colours are colonisations providing only trace isolates, whereas dark lines were identified on multiple dates or at different body sites. Crosses represent times of screens that were negative for MRSA, while circles those that returned positive swabs and sequenced isolates. The grey arrows represent reconstructed transmission events. These appear when at least one subgraph from the recipient is descended from an adjacent subgraph from the infector. Such a transmission may also involve unsampled intermediaries or the environment. The timings represent the last possible time at which these could have occurred, which is the earliest time of sampling of all the sequences involved in the recipient subgraphs. The dotted vertical lines demarcate the period of sampling.

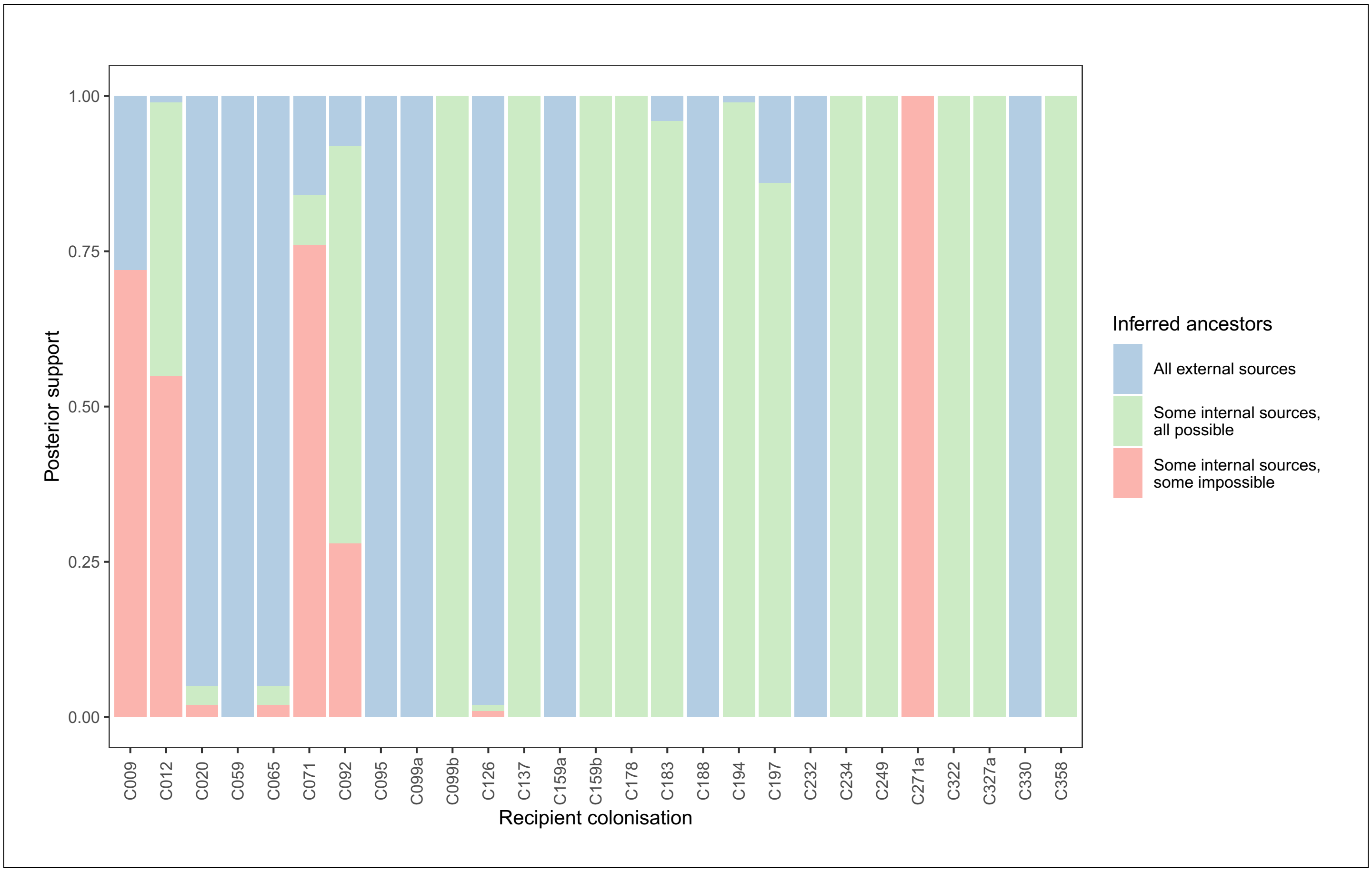

Figure 5. Concordance of inferred infectors for each colonisation with recorded timings of hospital stays and sampling dates. Each bar represents a colonisation, and the colours represent the proportions of the posterior set of trees where the transmission chain prior to that host involves no sampled subjects (blue), involves one or more sampled subjects all of which are possible given known timings of entry and departure to the hospital and sampling of isolates (green) and at least one sampled subject where the timings are in conflict, with the infector entering the hospital after isolates from the recipient were acquired (red).

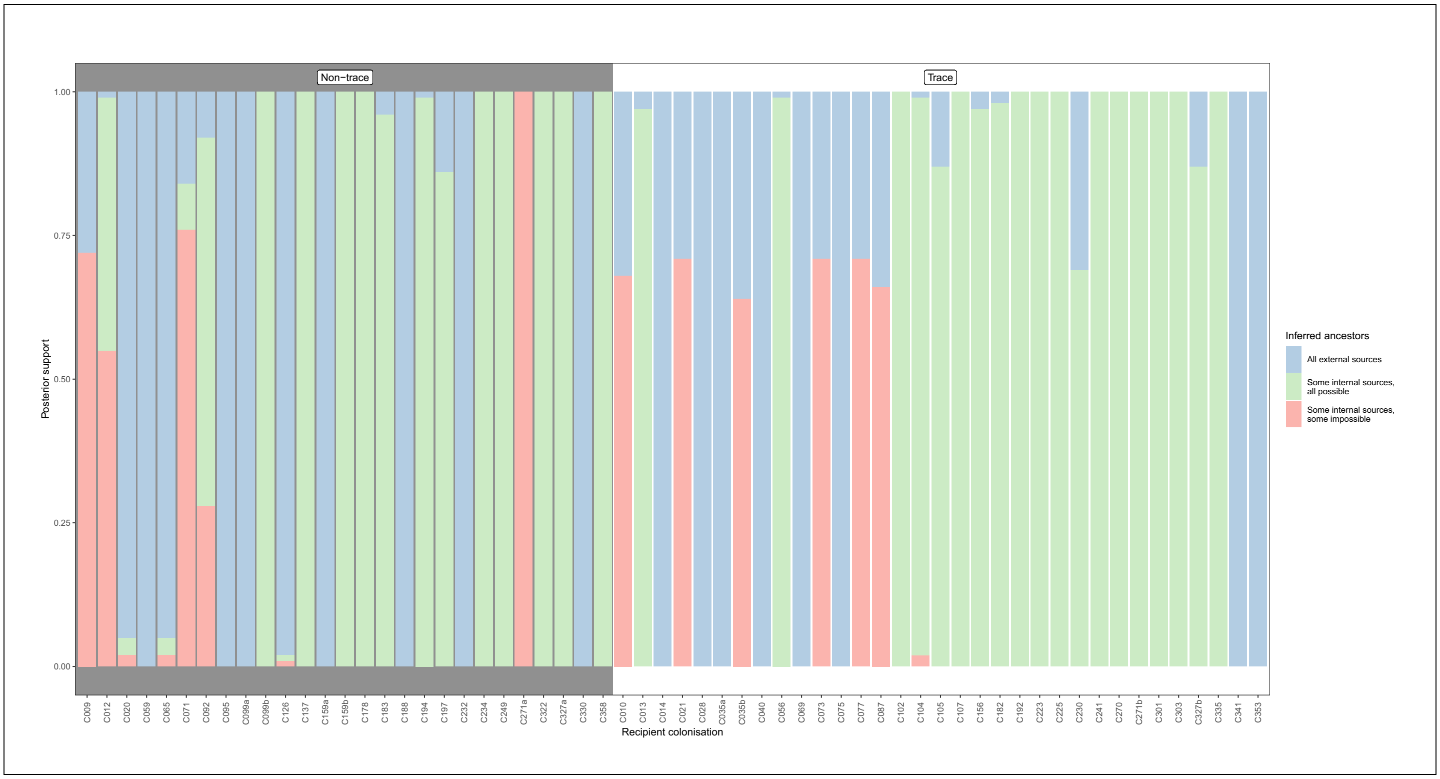

Figure 5—figure supplement 1. Concordance of inferred infectors for each colonisation with recorded timings of hospital stays and sampling dates, with trace colonisations included. Each bar represents a colonisation, and the colours represent the proportions of the posterior set of trees where the transmission chain prior to that host involves no sampled subjects (blue), involves one or more sampled subjects all of which are possible given known timings of entry and departure to the hospital and sampling of isolates (green) and at least one sampled subject where the timings are in conflict, with the infector entering the hospital after isolates from the recipient were acquired (red). The graph is divided into non-trace (left) and trace (right) colonisations.

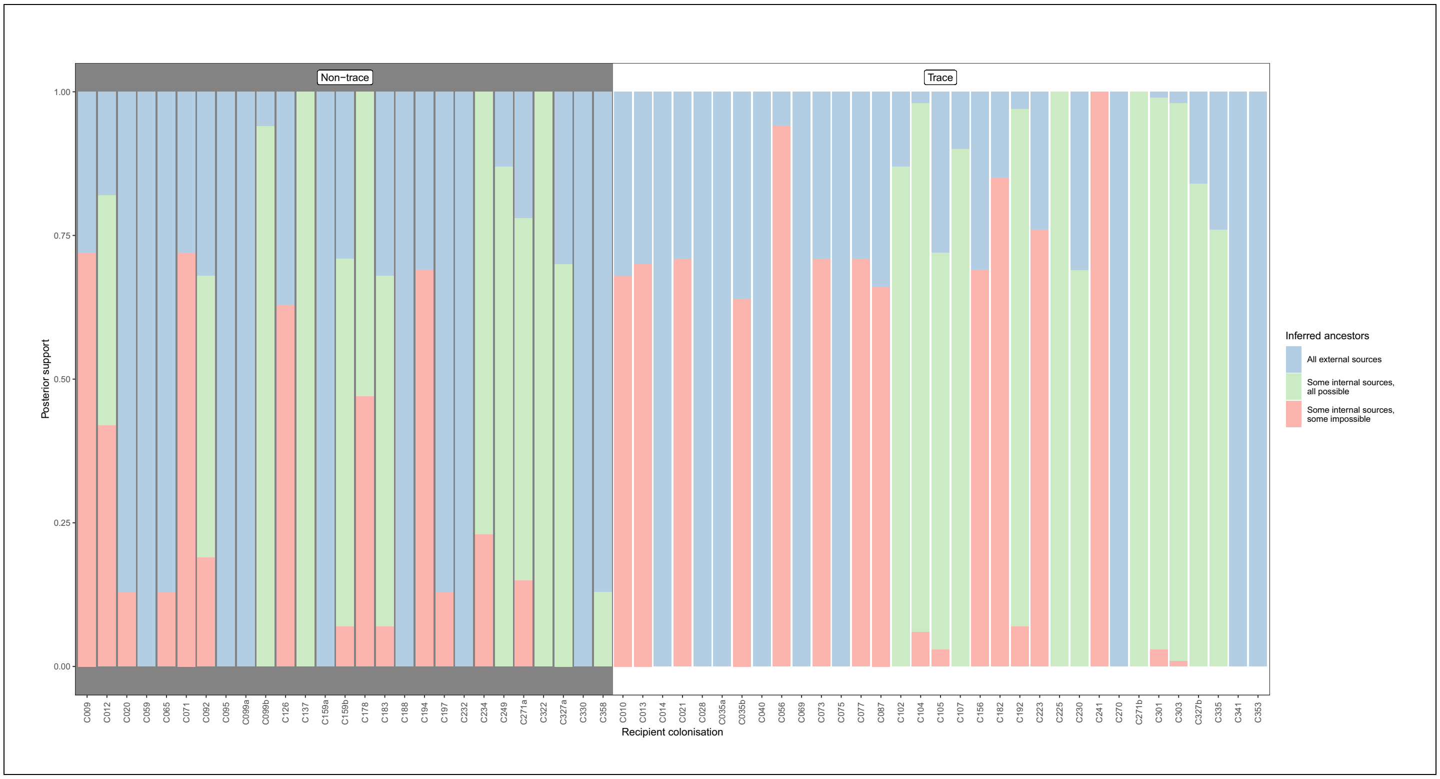

Figure 5—figure supplement 2. Concordance of inferred infectors for each colonisation with recorded timings of hospital stays and sampling dates, from a secondary *phyloscanner* analysis in which the tip set from each subject was randomly downsampled until a maximum of five remained. Each bar represents a colonisation, and the colours represent the proportions of the posterior set of trees where the transmission chain prior to that host involves no sampled subjects (blue), involves one or more sampled subjects all of which are possible given known timings of entry and departure to the hospital and sampling of isolates (green) and at least one sampled subject where the timings are in conflict, with the infector entering the hospital after isolates from the recipient were acquired (red). The graph is divided into non-trace (left) and trace (right) colonisations.

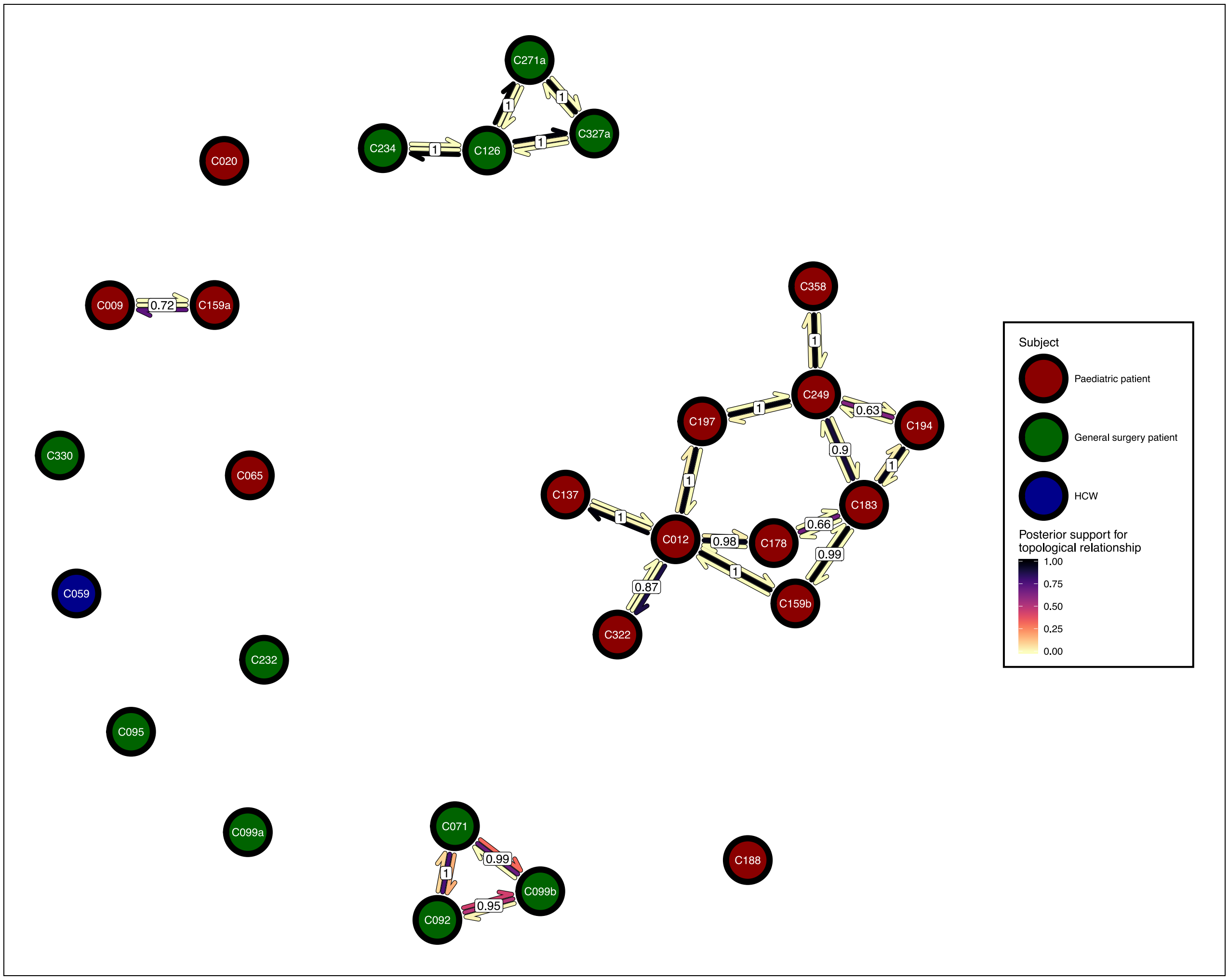

Figure 6. The *phyloscanner* host relationship diagram. Each node represents all the sequences for one colonisation. Node fill colours designate patients in the two hospital ICUs and the HCWs. Edges appear where colonisations share a relationship with posterior support of at least 0.5 and consist of three elements: arrows representing transmission in either direction and a central line segment representing the 'complex' topological relationship, which is indicative of transmission but the direction is ambiguous. Each of these is coloured according to the proportion of posterior trees showing the corresponding relationship. Edges are also labelled with the overall posterior support for any topology suggesting transmission.

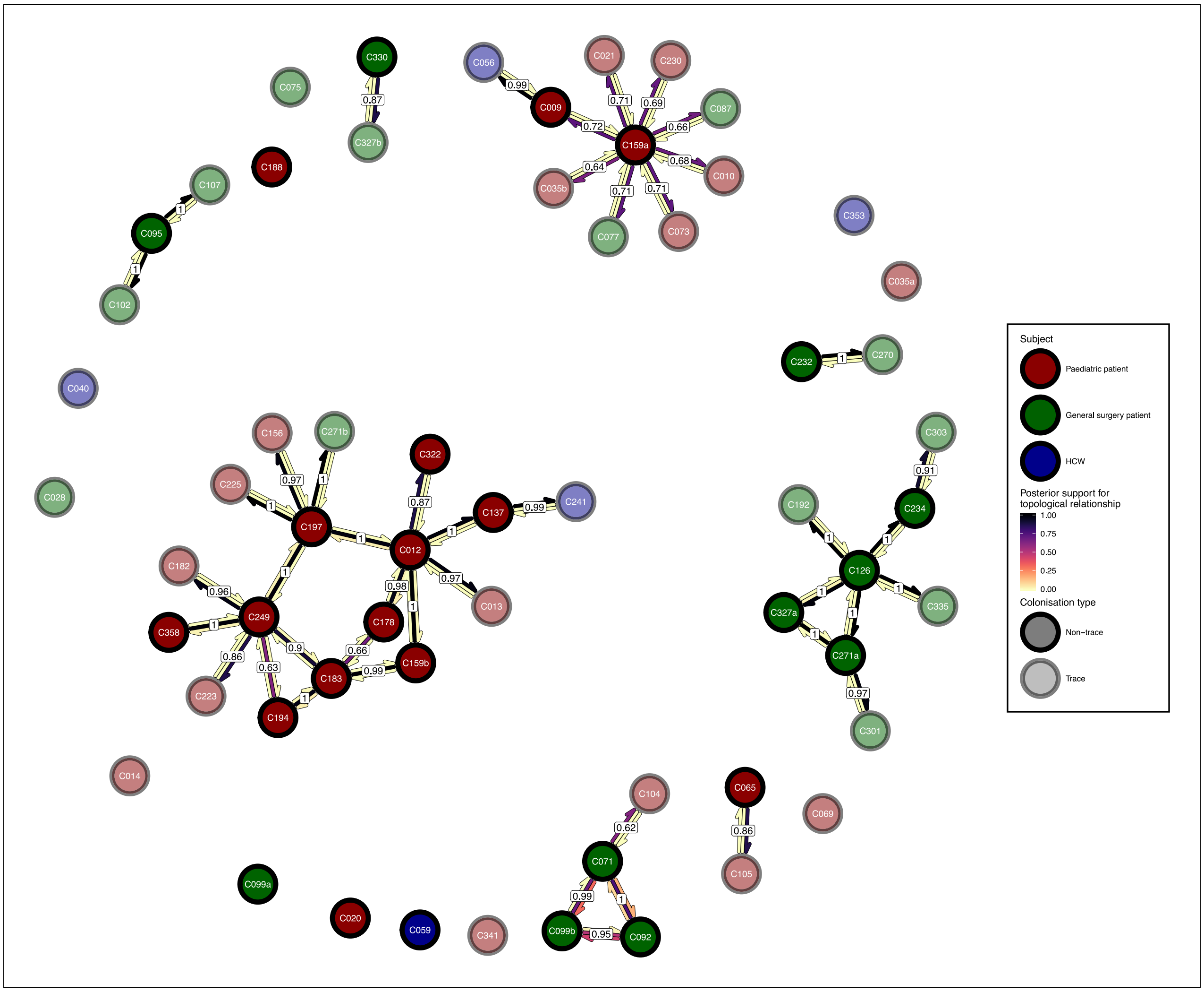

Figure 6—figure supplement 1. The *phyloscanner* host relationship diagram, with trace colonisations included. Each node represents all the sequences for one colonisation. Node fill colours designate patients in the two hospital ICUs and the HCWs and nodes with faded colours representing trace colonisations. Edges appear where colonisations share a relationship with posterior support of at least 0.5 and consist of three elements: arrows representing transmission in either direction and a central line segment representing the ‘complex’ topological relationship, which is indicative of transmission but the direction is ambiguous. Each of these is coloured according to the proportion of posterior trees showing the corresponding relationship. Edges are also labelled with the overall posterior support for any topology suggesting transmission.

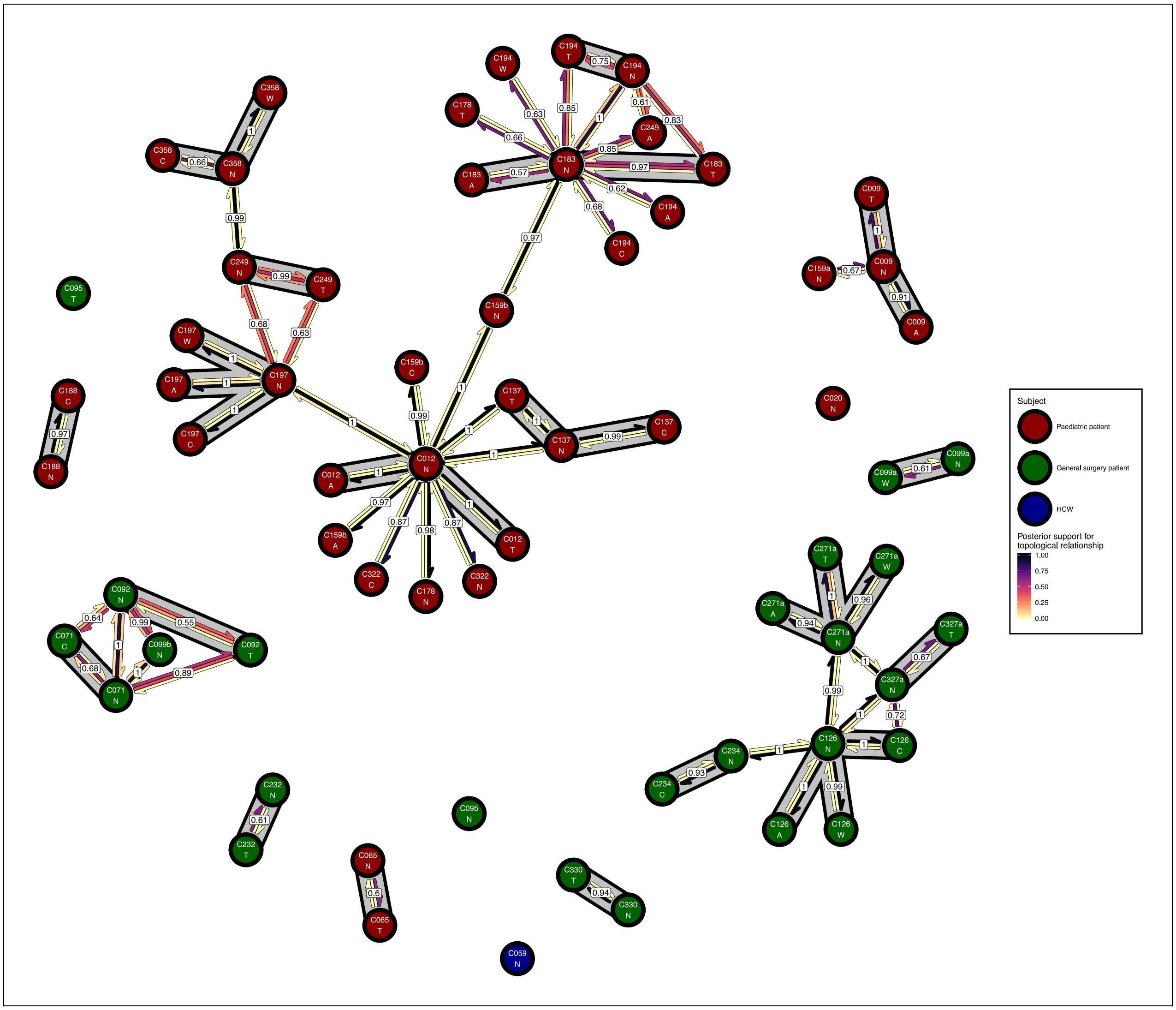

Figure 7. The *phyloscanner* host relationship diagram for a separate analysis where samples taken from distinct body sites on the same subject were treated as separate 'hosts'. Each node represents all the sequences for one colonisation of one body site. Node fill colours designate patients in the two hospital ICUs and the HCWs. Edges appear where colonisations share a relationship with posterior support of at least 0.5 and consist of three elements: arrows representing transmission in either direction and a central line segment representing the 'complex' topological relationship, which is indicative of transmission but the direction is ambiguous. Each of these is coloured according to the proportion of posterior trees showing the corresponding relationship. Edges are also labelled with the overall posterior support for any topology suggesting transmission, and edges connecting colonisations from sites from the same subject have a grey background. Nodes are annotated with colonisation IDs and a code for body site: A = axilla, C = endotracheal suction, N = nose, T = throat, W = wound.

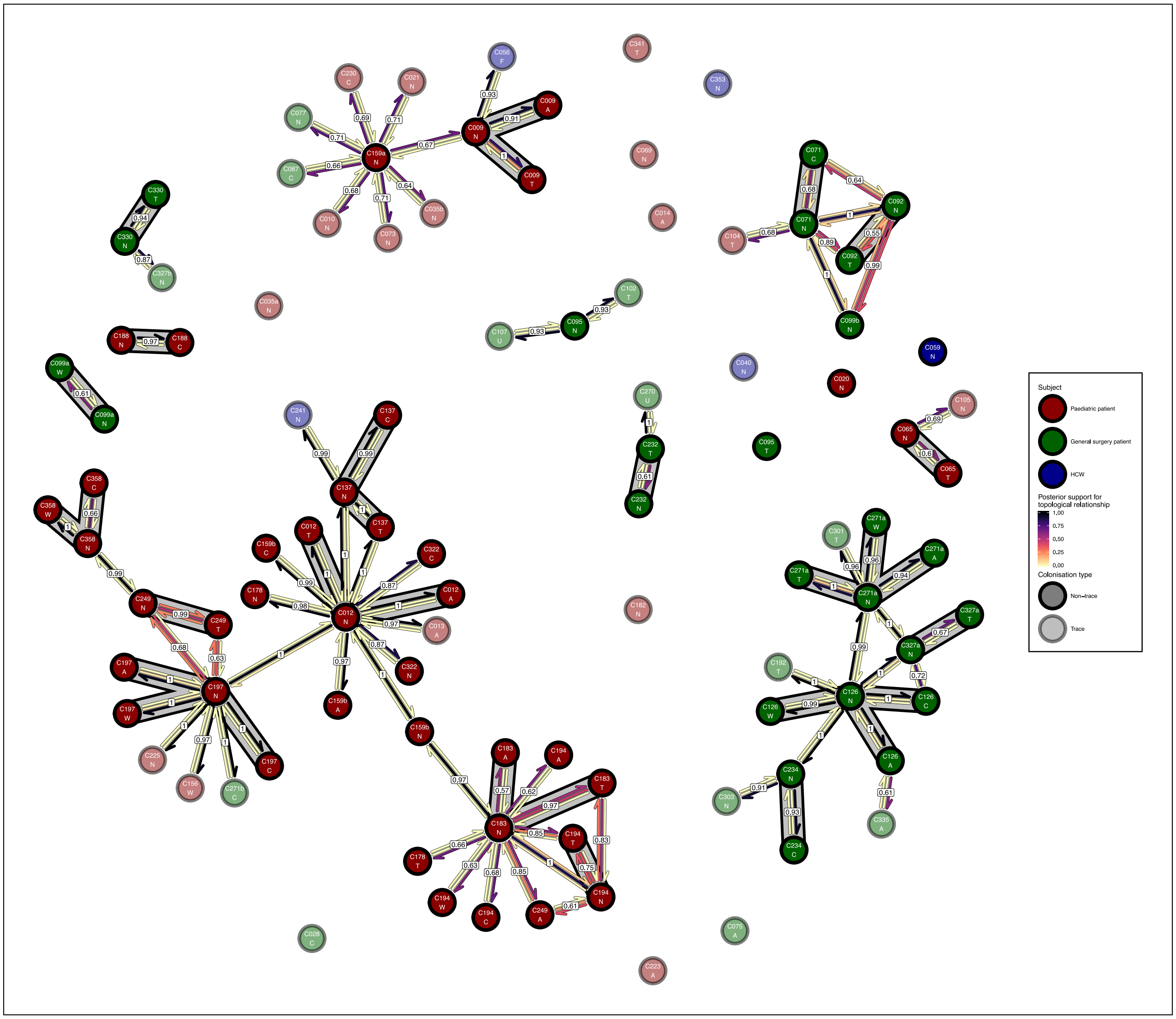

Figure 7—figure supplement 1. The *phyloscanner* host relationship diagram for the body site analysis, with trace colonisations included. Each node represents all the sequences for one colonisation of one body site. Node fill colours designate patients in the two hospital ICUs and the HCWs and nodes with faded colours representing trace colonisations. Edges appear where colonisations share a relationship with posterior support of at least 0.5 and consist of three elements: arrows representing transmission in either direction and a central line segment representing the 'complex' topological relationship, which is indicative of transmission but the direction is ambiguous. Each of these is coloured according to the proportion of posterior trees showing the corresponding relationship. Edges are also labelled with the overall posterior support for any topology suggesting transmission, and edges connecting colonisations from sites from the same subject have a grey background. Nodes are annotated with colonisation IDs and a code for body site: A = axilla, C = endotracheal suction, F = fingertips, N = nose, T = throat, U = urine, W = wound.

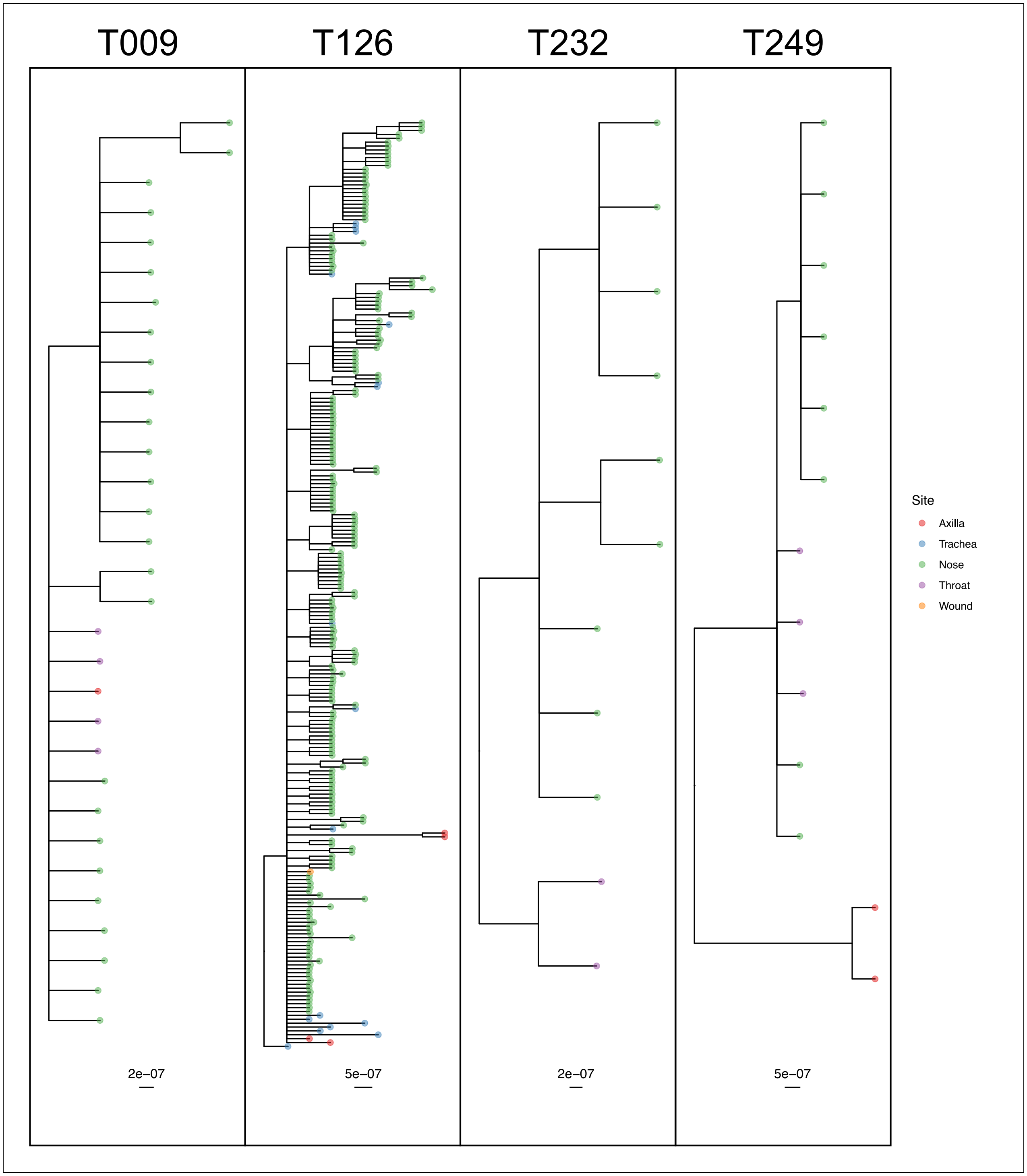

Figure 8. 50% majority-rule consensus phylogenies for the ExaBayes phylogenetic analyses of the sequences from patients T009, T126, T232 and T249. Tips are coloured by body site of origin. Branch lengths are in substitutions per site. Trees were rooted using the TW20 outgroup (not shown).

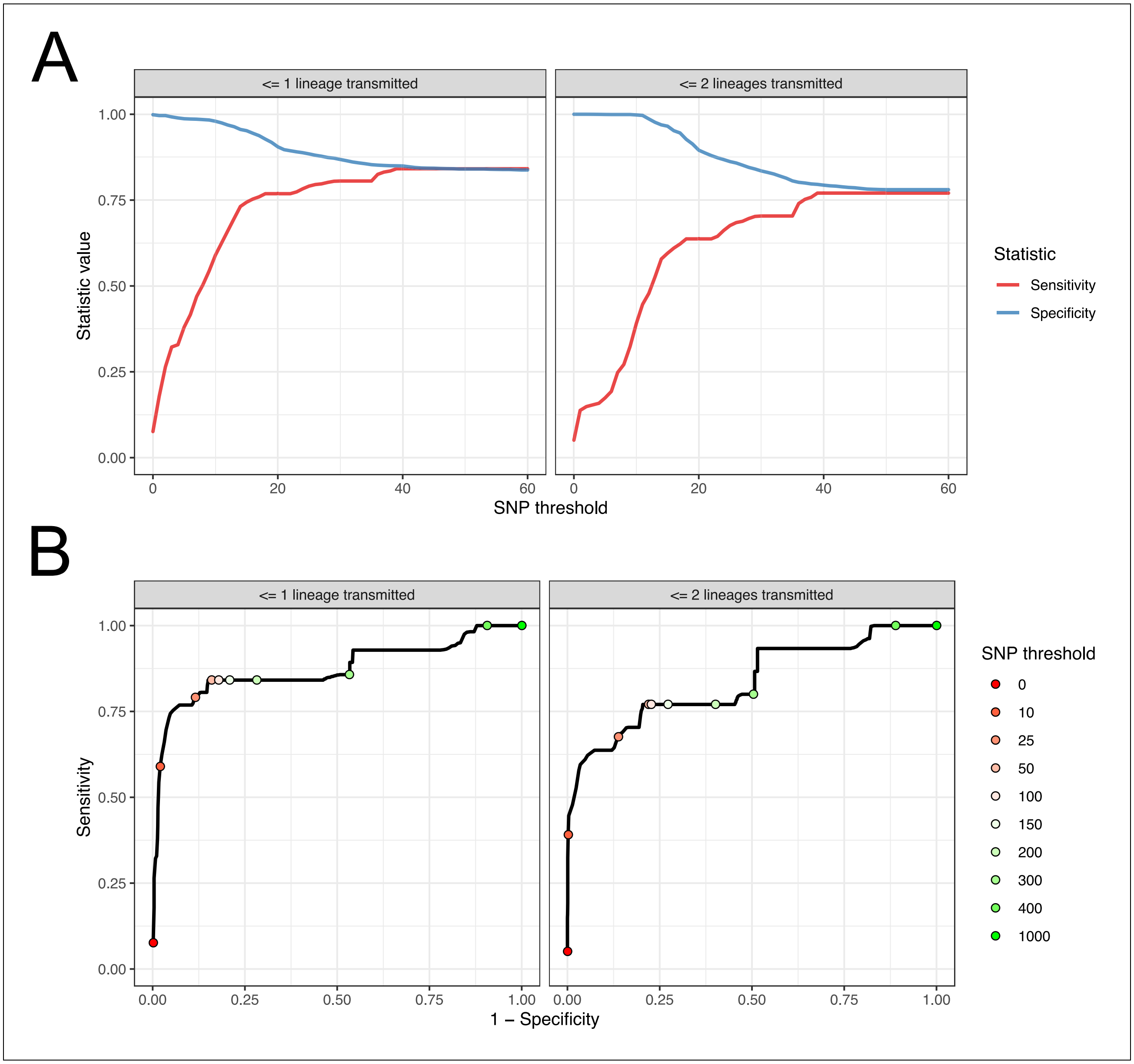

Figure 9. Performance of SNP distance as a method for identifying transmission pairs. (**A**) Plots of the sensitivity and specificity of using the number of SNPs to identify transmission pairs, for different distance thresholds. (**B**) ROC curves plotting true positive rate (sensitivity) against false positive rate (1-specificity) for different thresholds. The curve is annotated with selected threshold values. The gold standard for identifying transmission pairs in the version on the left is a topological relationship suggesting at least one transmitted lineage, while on the right at least two are required, a criterion which will occur much less often if there is a missing intermediary in transmission.

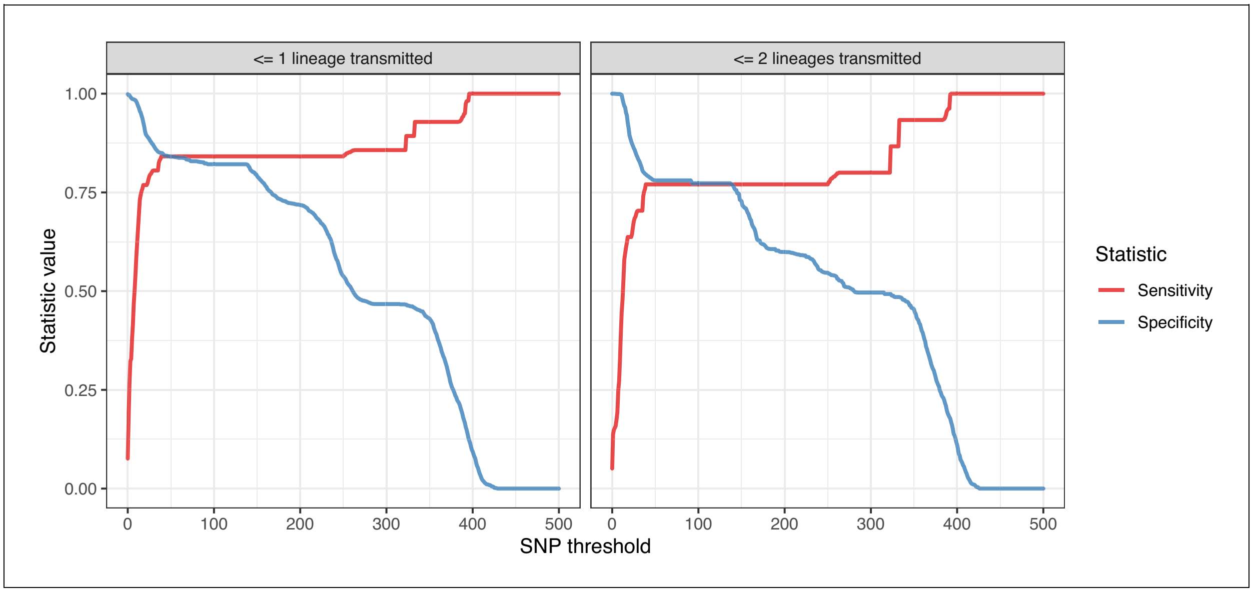

Figure 9—figure supplement 1. Plots of the sensitivity and specificity of using the number of SNPs to identify transmission pairs, for different distance thresholds. The gold standard for identifying transmission pairs in the version on the left is a topological relationship suggesting at least one transmitted lineage, while on the right at least two are required, a criterion which will occur much less often if there is a missing intermediary in transmission.