Directed networks reveal genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes

Ovidiu Popa, Einat Hazkani-Covo, Giddy Landan, William Martin, Tal Dagan

SUPPLEMENTARY MATERIAL

Figure S1. Edge weight in the directed network of recent lateral gene transfers....................... 2
Figure S2. Community structure within the dLGT network. 3
Figure S3-A. Donor-recipient similarity by *Sgs*.. 4
Figure S3-B. Donor-recipient similarity by *Spr*. 5
Figure S3-C. Donor-recipient similarity by *Sgc*.. 6
Figure S4. Intergeneric network. 7
Figure S5. Distribution of transferred genes having unknown function across taxa. 8
Figure S6. Pathogens and nonpathogens in the dLGT network. 9
Figure S7. Distribution of IN and OUT degree in pathogens and nonpathogens 10
Figure S8. Overlap between the orthogroup by rBBH and MCL algorithms. 11
Figure S9. Model for change in GC content similarity as a function of evolutionary distance. 12
Figure S10. Phylogenetic inference of donor and recipient species. 13

Figure S1. Edge weight in the directed network of recent lateral gene transfer
Edges in the network are colored by the number of genes transferred from the donor to the recipient that they connect (see color legend at the bottom).

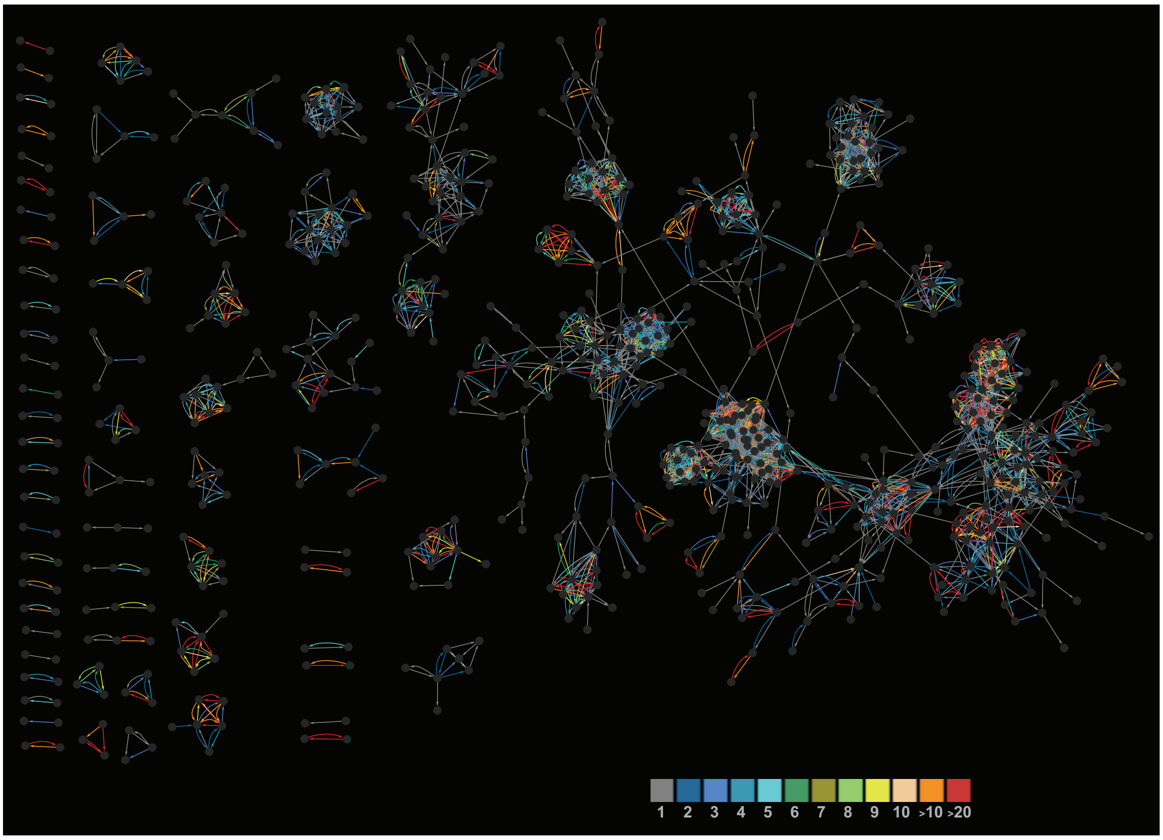


Figure S2. Community structure within the dLGT network.
Vertices that are grouped into the same module are colored the same.

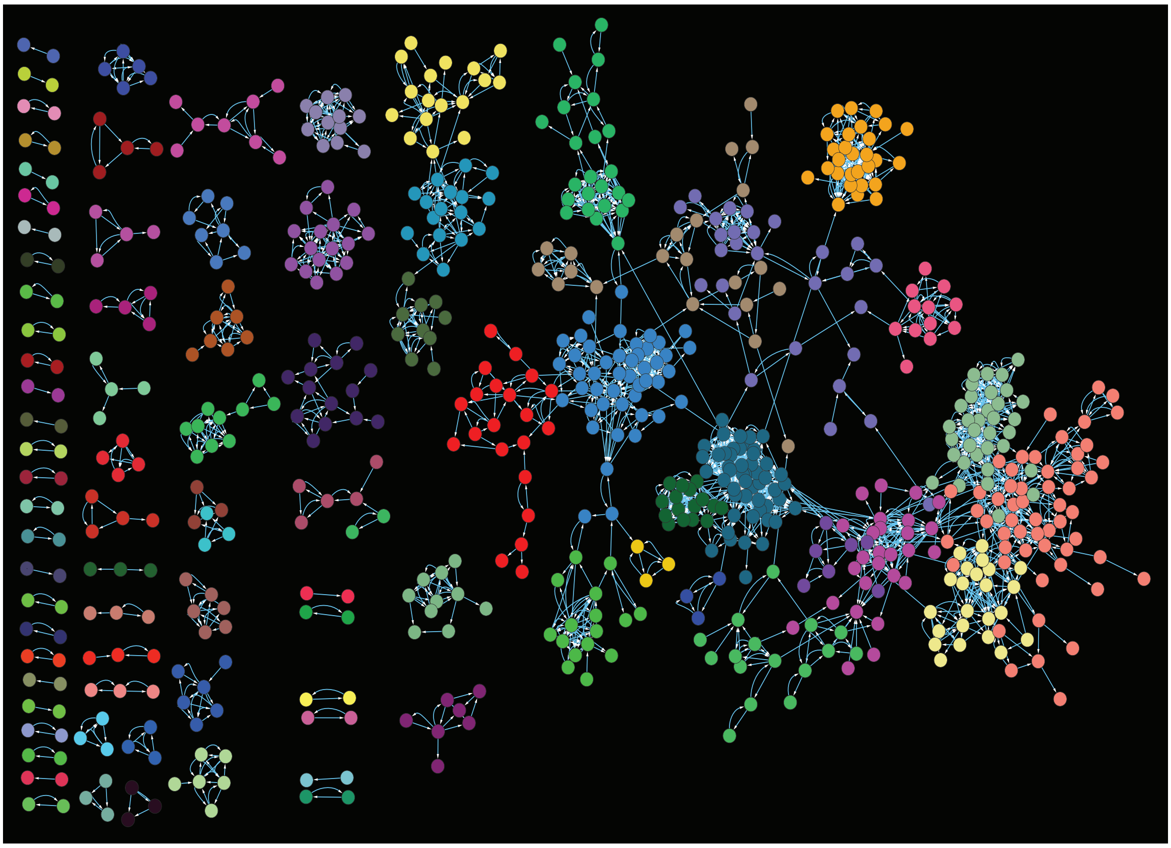


Figure S3-A. Donor-recipient similarity by *Sgs*.
Color-coded matrix of donor (y-axis) and recipient (x-axis) genome similarity (S_{gs}) for all genome pairs, sorted by taxonomical classification.

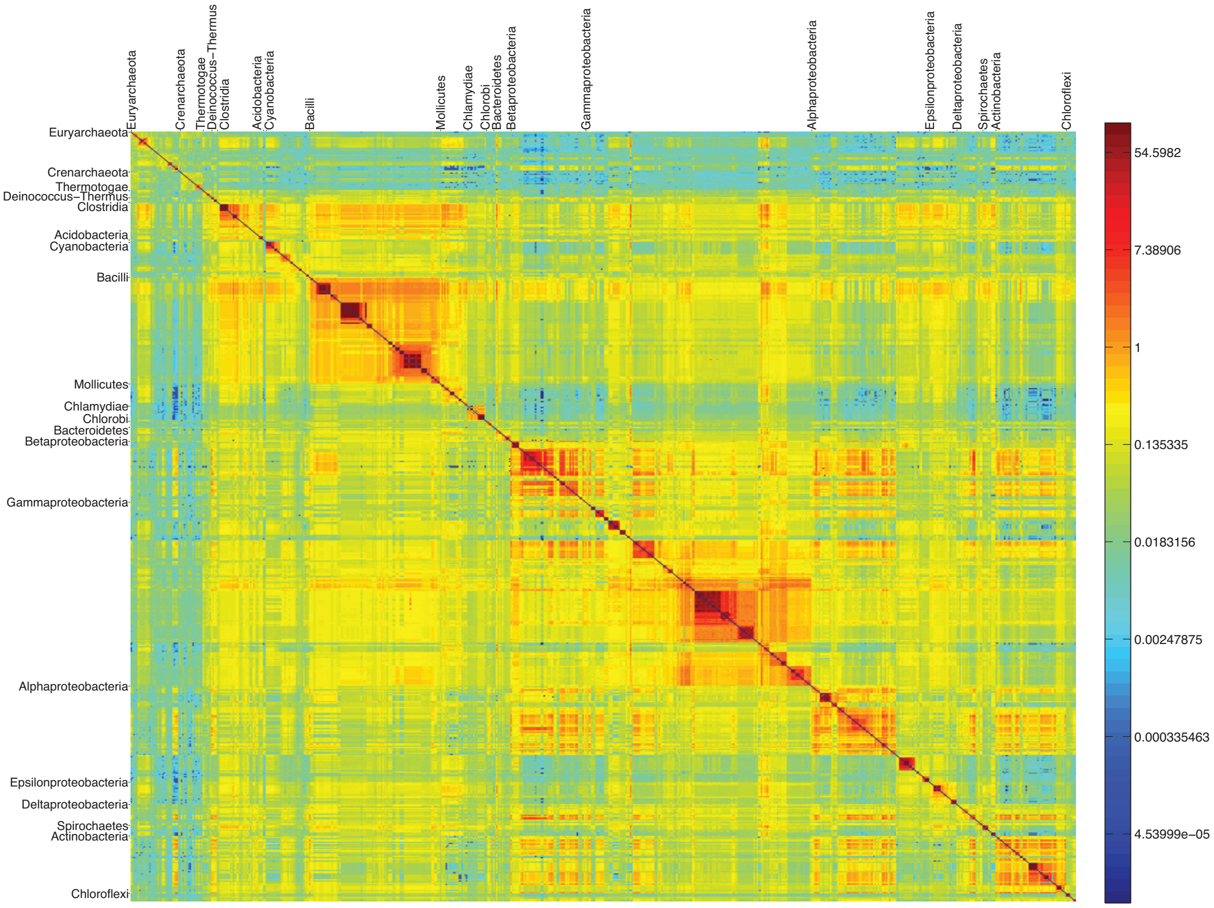


Figure S3-B. Donor-recipient similarity by *Spr*.
Color-coded matrix of donor (y-axis) and recipient (x-axis) proteome similarity(S_{pr}) for all genome pairs, sorted by taxonomical classification.

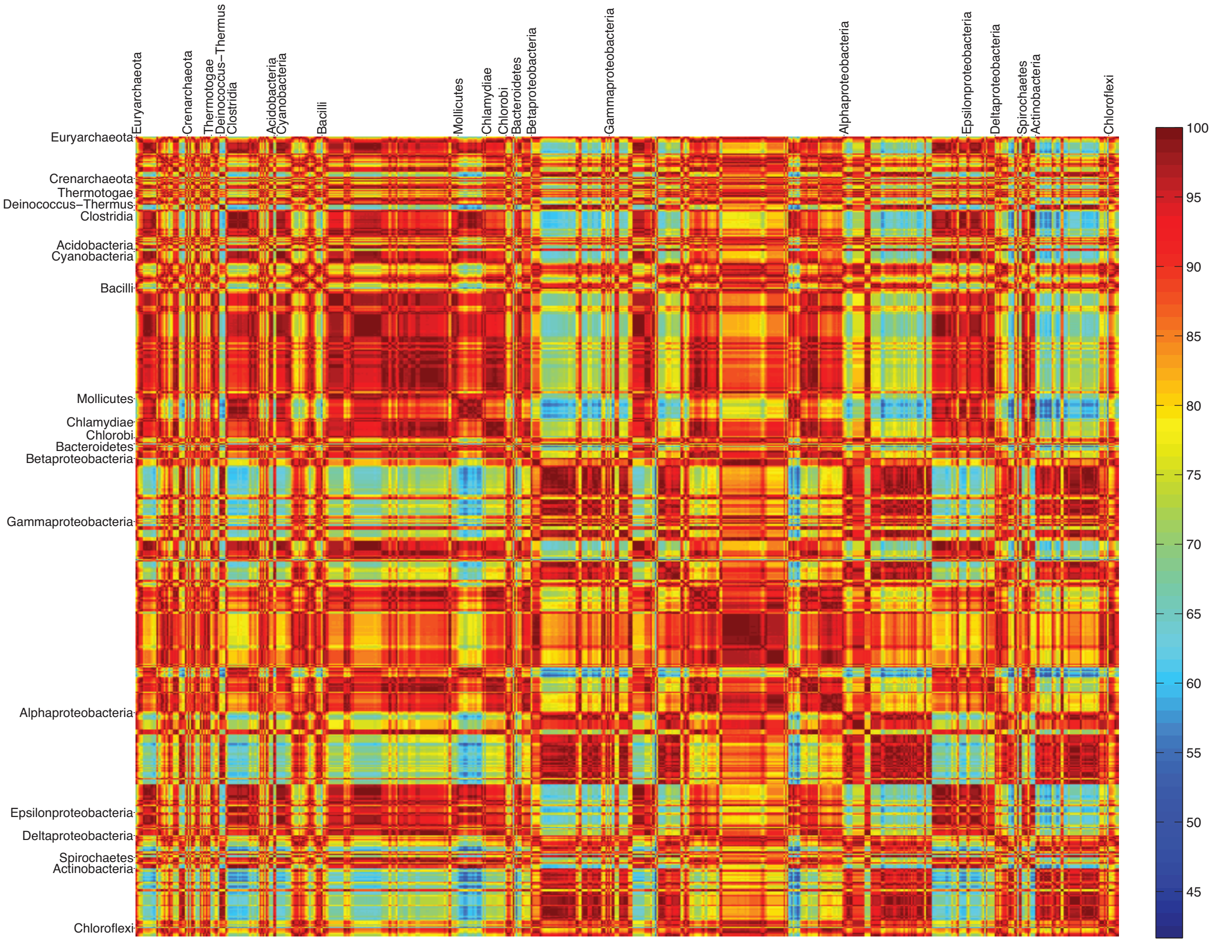


Figure S3-C. Donor-recipient similarity by *Sgc*.
Color-coded matrix of donor (y-axis) and recipient (x-axis) GC content similarity (S_{gc}) for all genome pairs, sorted by taxonomical classification.

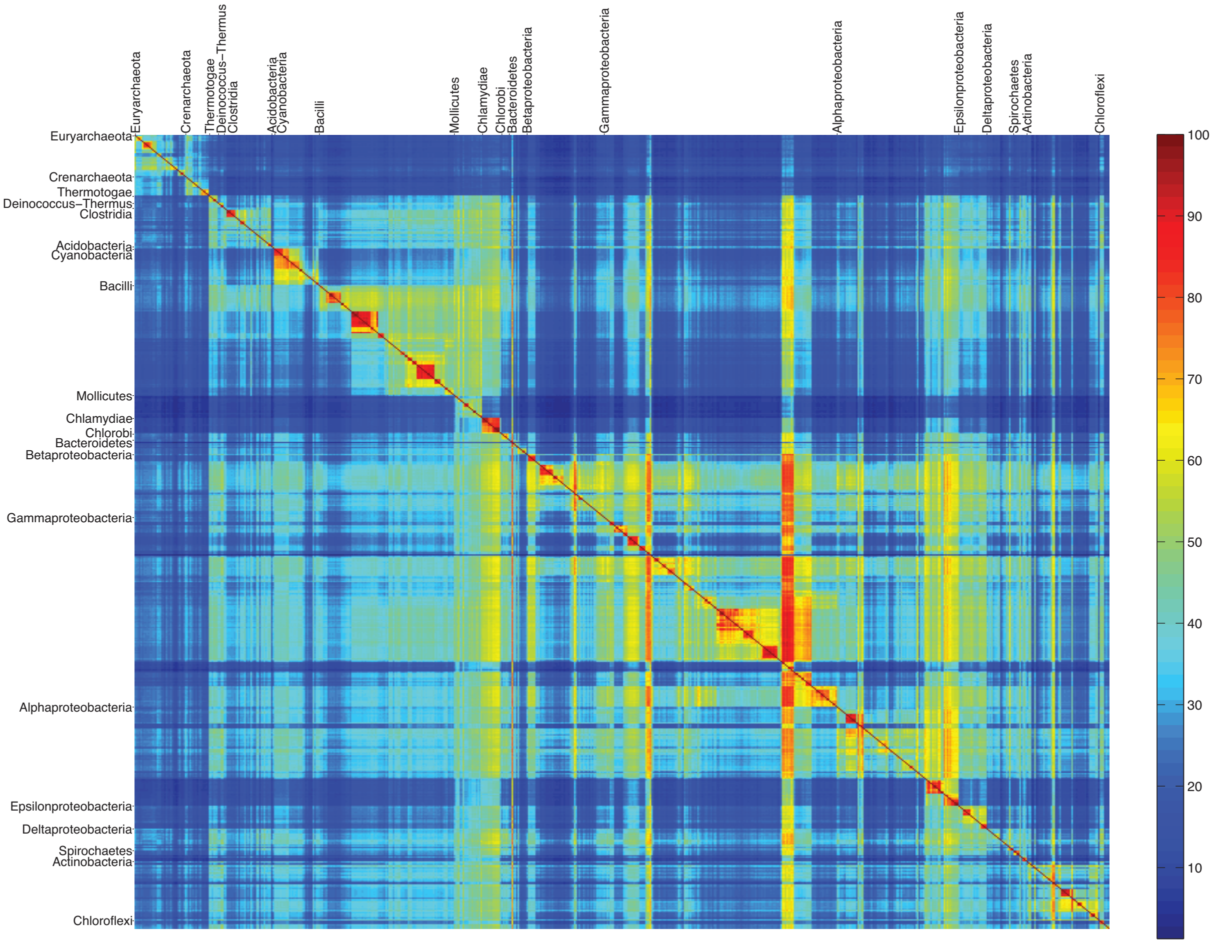


Figure S4. Intergeneric network.
(A) the intergeneric network. Each vertice represents all species from a certain genus. Edges represent one or more recent LGT between species from the genus-vertex they connect. (B) distribution of IN degree and OUT degree in the network. (C) Distribution of edge weight in the network.

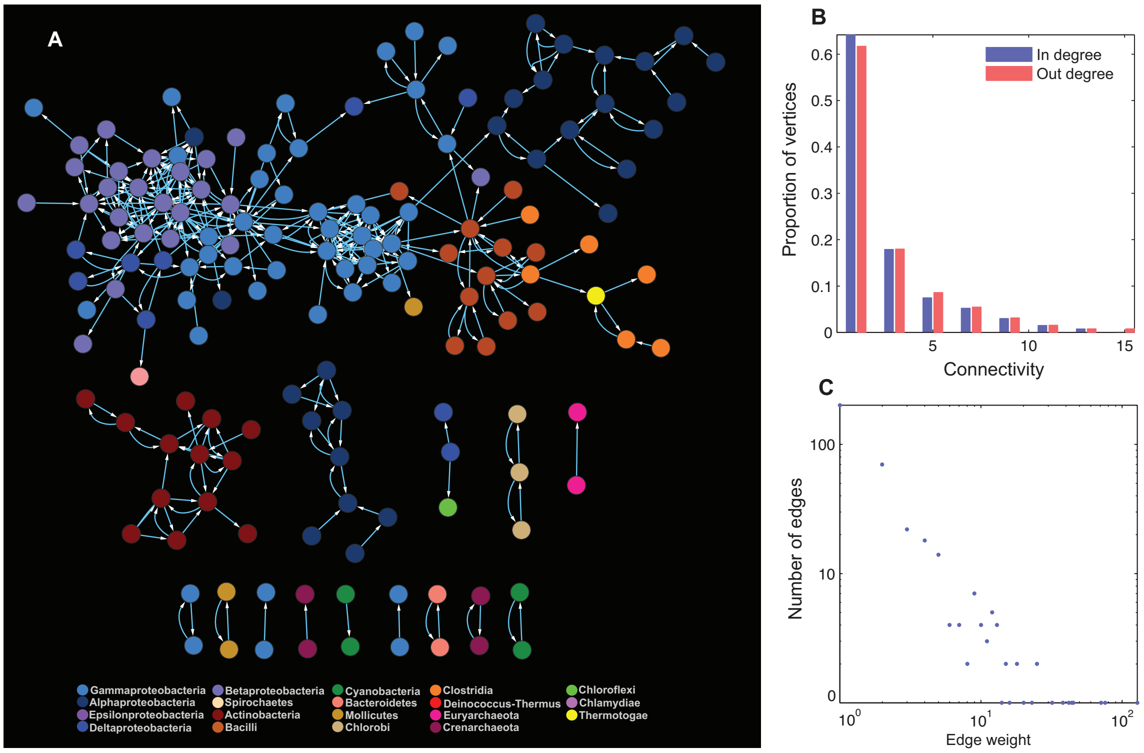
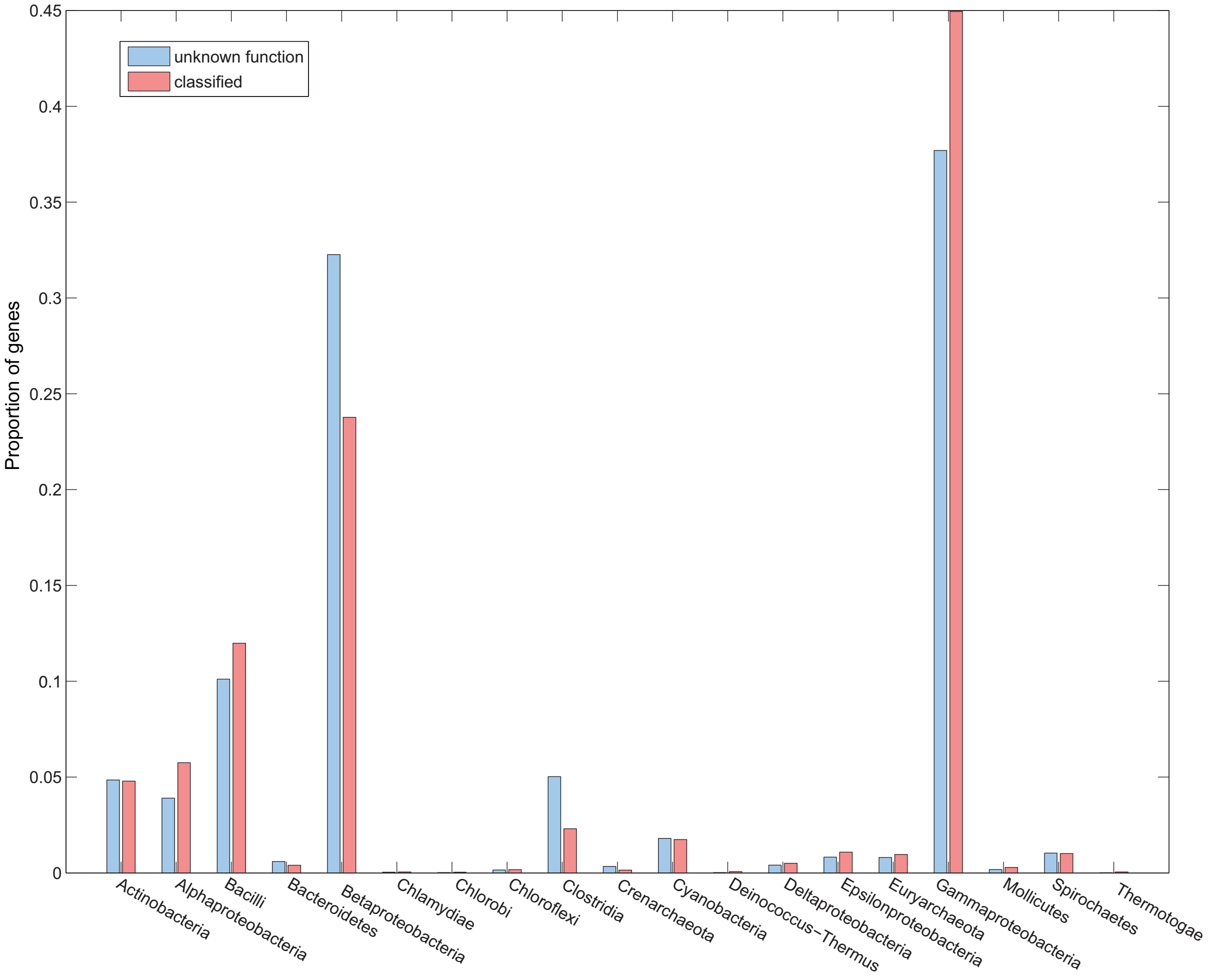


Figure S5. Distribution of transferred genes having unknown function across taxa.



SUPPLEMENTARY MATERIAL

Figure S6. Pathogens and nonpathogens in the dLGT network.

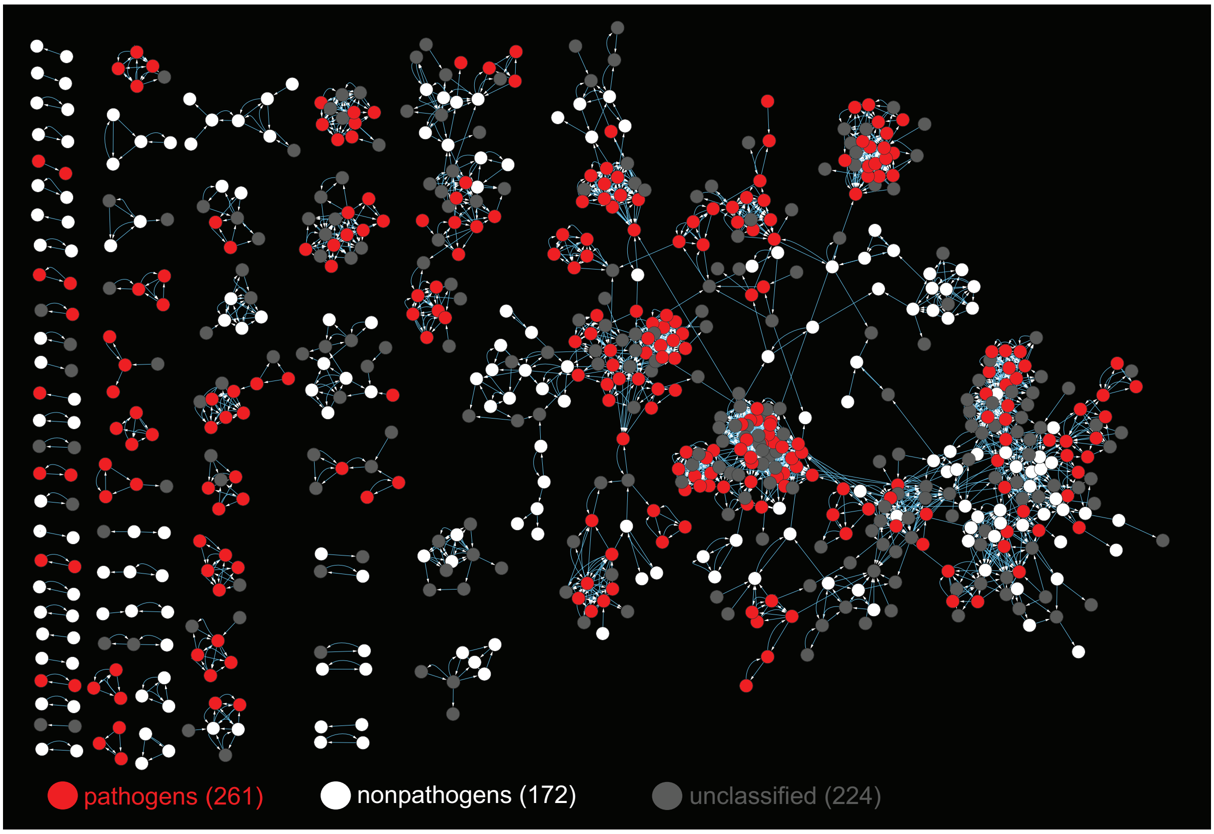


Figure S7. Distribution of IN and OUT degree in pathogens and nonpathogens

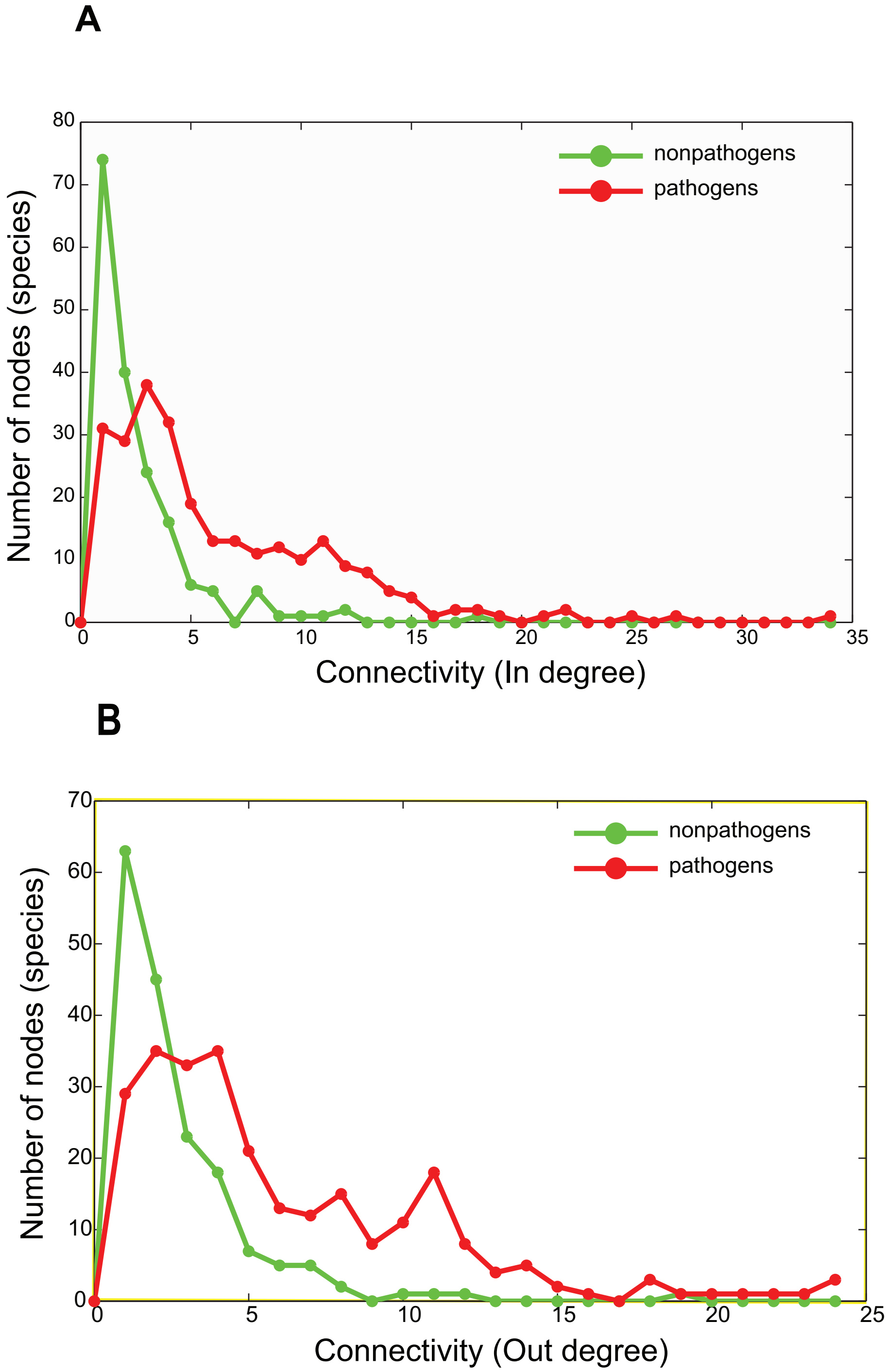


Figure S8. Overlap between the orthogroup by rBBH and MCL algorithms.

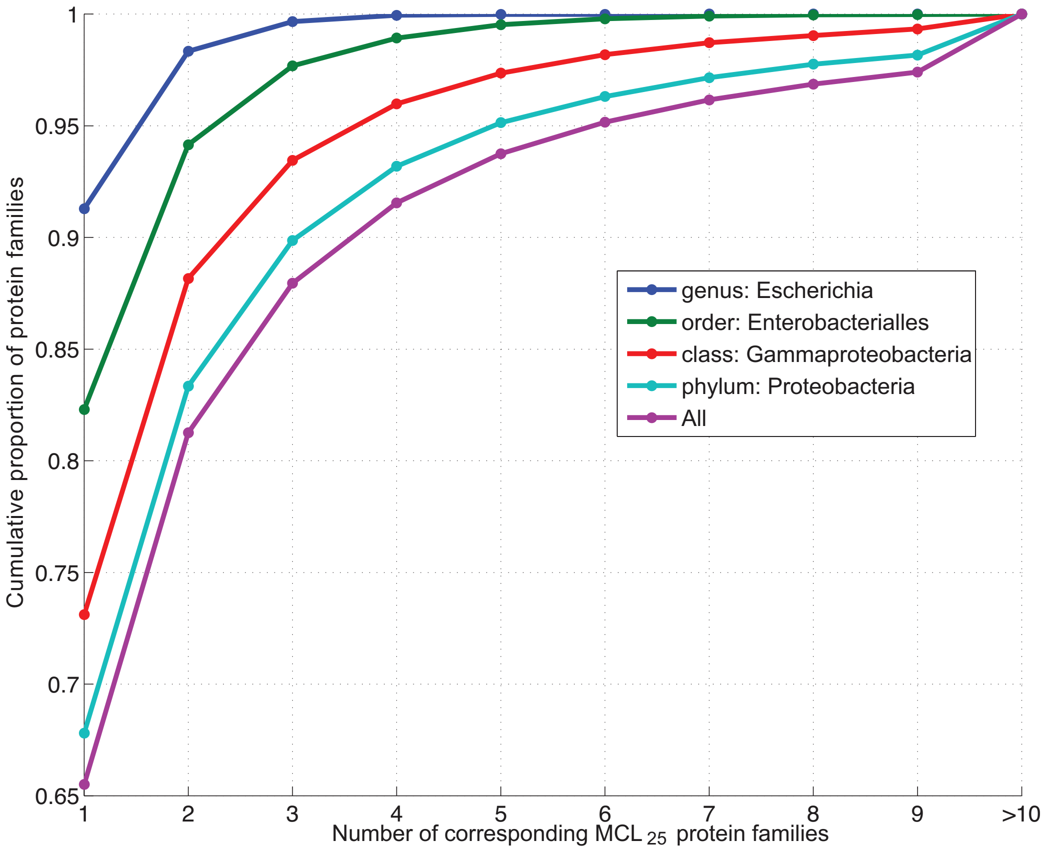


Figure S9. Model for change in GC content similarity as a function of evolutionary distance.
The top figure shows the distribution of sample size in the model. The figure at the bottom shows 95% percentile of GC content similarity for the samples at the top by the number of different nucleotides.

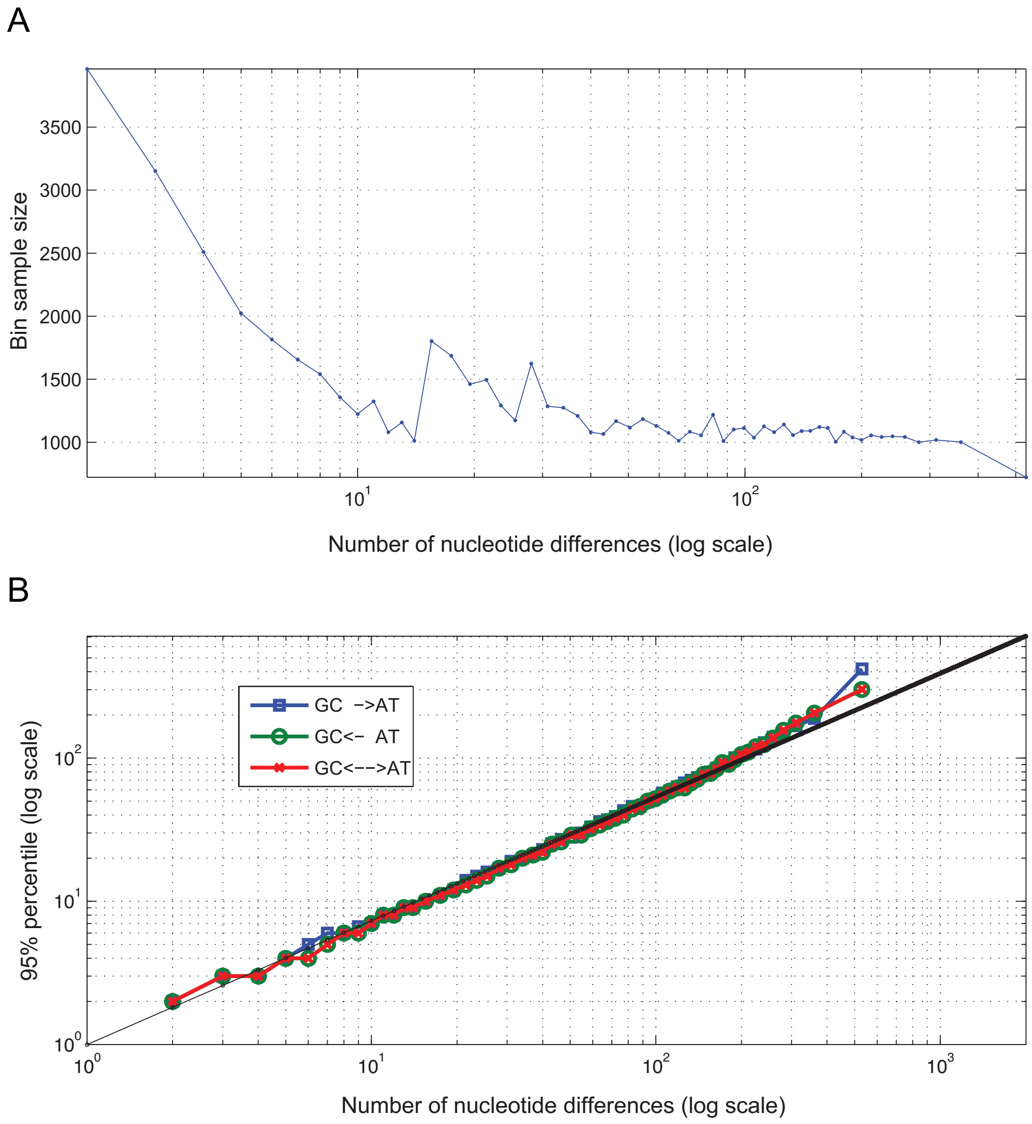


Figure S10. Phylogenetic inference of donor and recipient species.