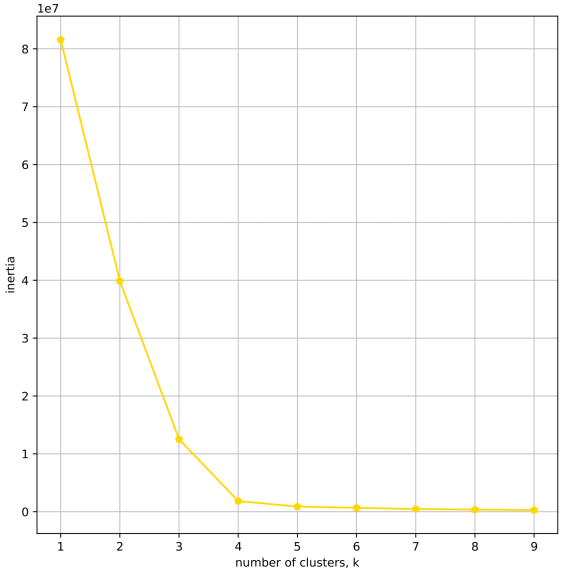

Figure A1: Scree plot showing a deterministic number of clusters = 4 (“elbow point”) determined by K-Means using 2 PCs (13% of the variance) from k-mer frequencies of four superpopulations (excluding AFR).

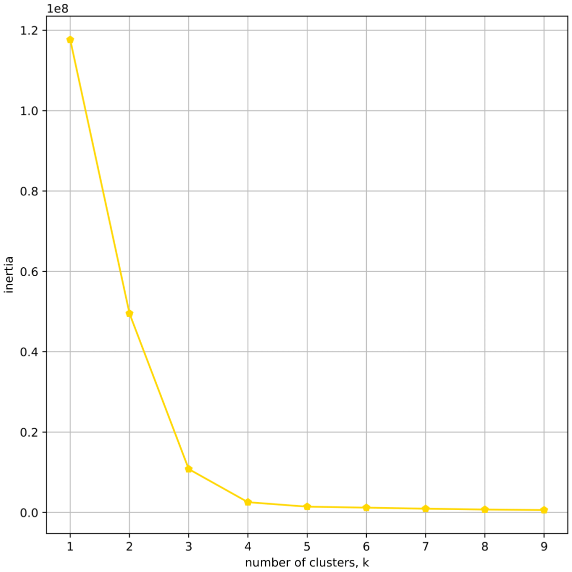

Figure A2: Scree plot showing a deterministic number of clusters = 4 (“elbow point”) determined by K-Means using 2 PCs (8.4% of the variance) from k-mer frequencies of four superpopulations with additional samples of admixed origin from EAS and EUR superpopulation.

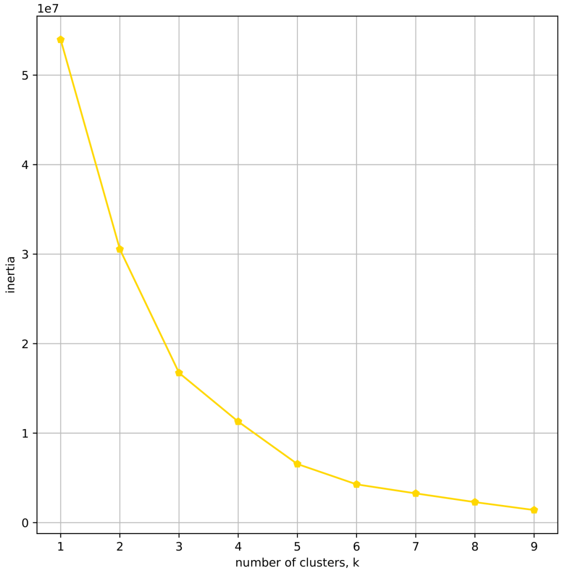

Figure A3: Scree plot showing no obvious ("elbow point") determined by K-Means using 2 PCs (13% of the variance) from k-mer frequencies of three populations, CDX (non-admixed), CHB (admixture of CDX and JPT), and JPT (non-admixed). All three populations are of a single superpopulation (EAS).

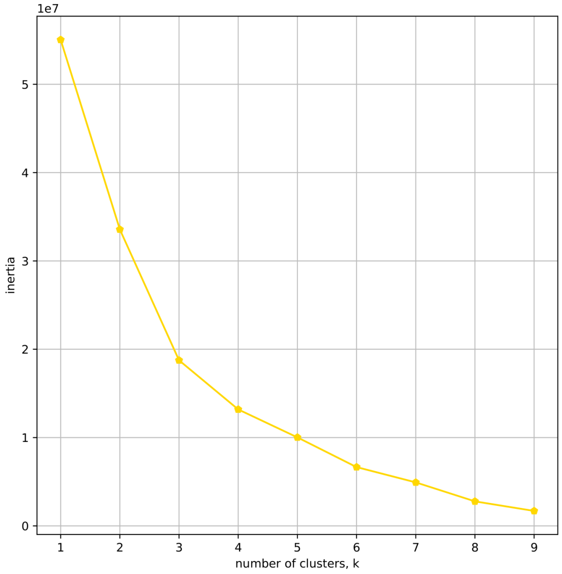

Figure A4: Scree plot showing no obvious ("elbow point") determined by K-Means using 2 PCs (13% of the variance) from k-mer frequencies of three populations, FIN (non-admixed), CEU (admixture of FIN and TSI), and TSI (non-admixed). All three populations are of a single superpopulation (EUR).

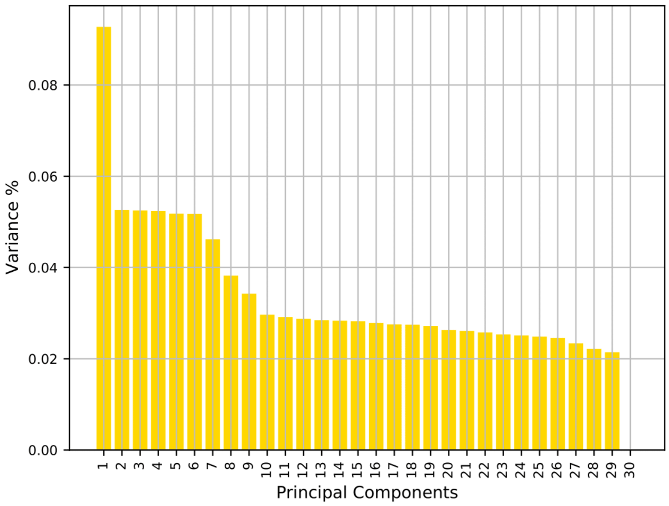

Figure A5: Bar plot showing explained variance per PC in the PCA analysis of 5 superpopulations of non-admixed origin.

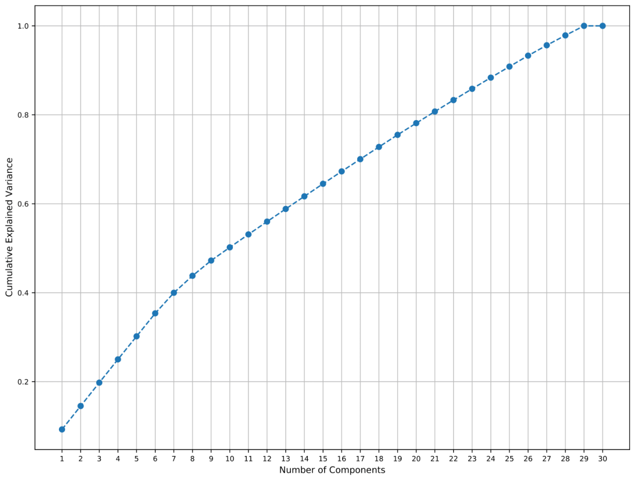

Figure A6: Scree plot showing cumulative explained variance per PC in the PCA in the analysis of 5 superpopulations of non-admixed origin.

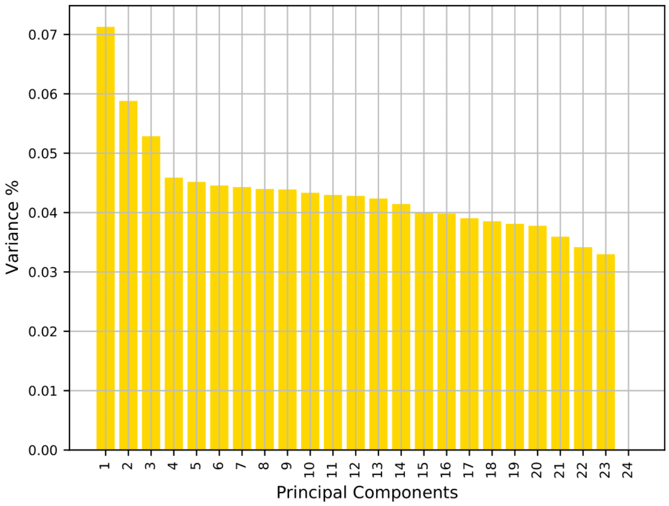

Figure A7: Bar plot showing explained variance per PC in the PCA analysis of 4 superpopulations of non-admixed origin (excluding AFR).

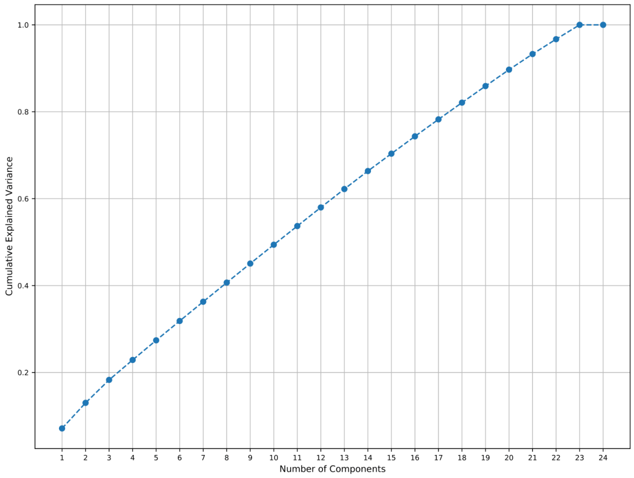

Figure A8: Scree plot showing cumulative explained variance per PC in the PCA analysis of 4 superpopulations of non-admixed origin (excluding AFR).

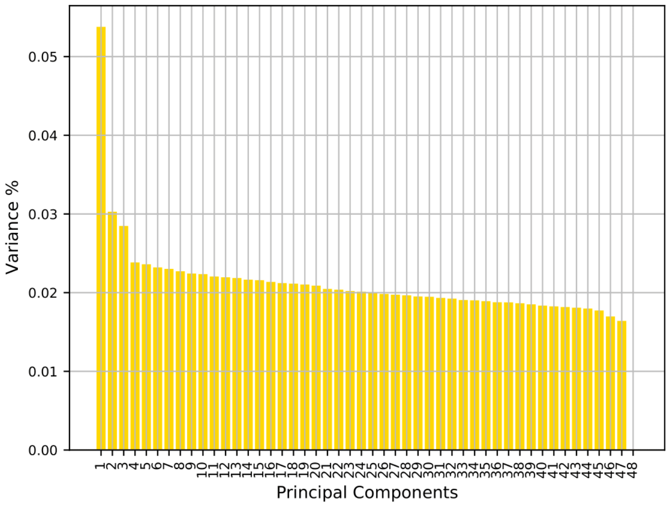

Figure A9: Bar plot showing explained variance per PC in the PCA analysis of 4 superpopulations including samples of admixed and non-admixed origin.

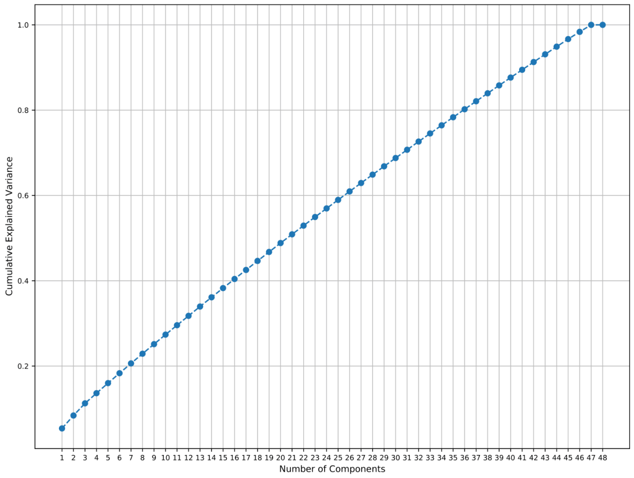

Figure A10: Scree plot showing cumulative explained variance per PC in the analysis PCA of 4 superpopulations including samples of admixed and non-admixed origin.

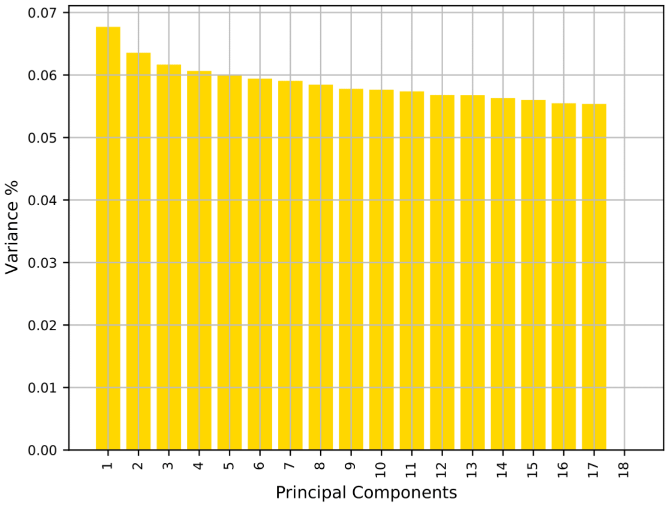

Figure A11: Bar plot showing explained variance per PC in the PCA analysis of 1 EAS superpopulation including samples of admixed and non-admixed origin.

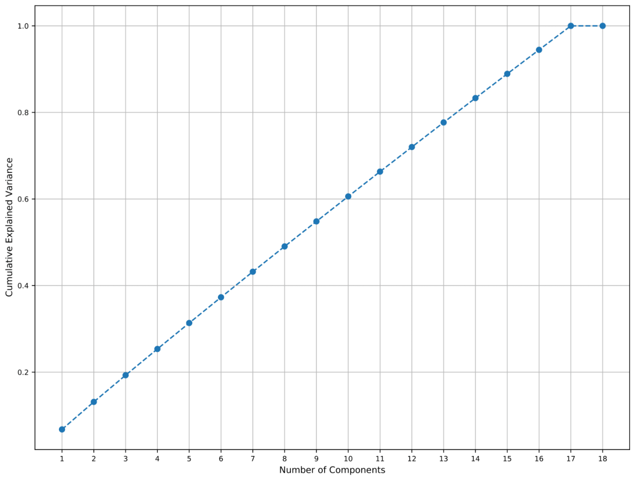

Figure A12: Scree plot showing cumulative explained variance per PC in the PCA analysis of 1 EAS superpopulation including samples of admixed and non-admixed origin.

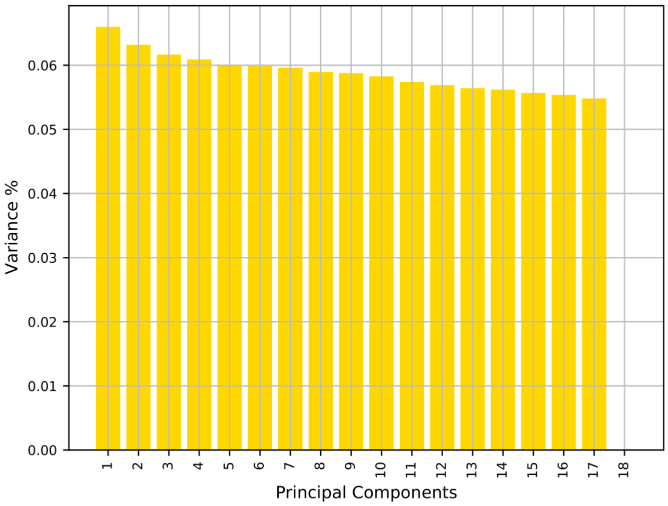

Figure A13: Bar plot showing explained variance per PC in the PCA analysis of 1 EUR superpopulation including samples of admixed and non-admixed origin.

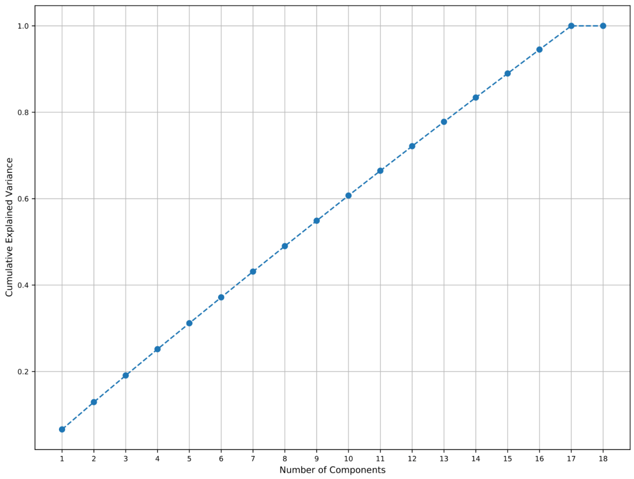

Figure A14: Scree plot showing cumulative explained variance per PC in the PCA analysis of 1 EUR superpopulation including samples of admixed and non-admixed origin.

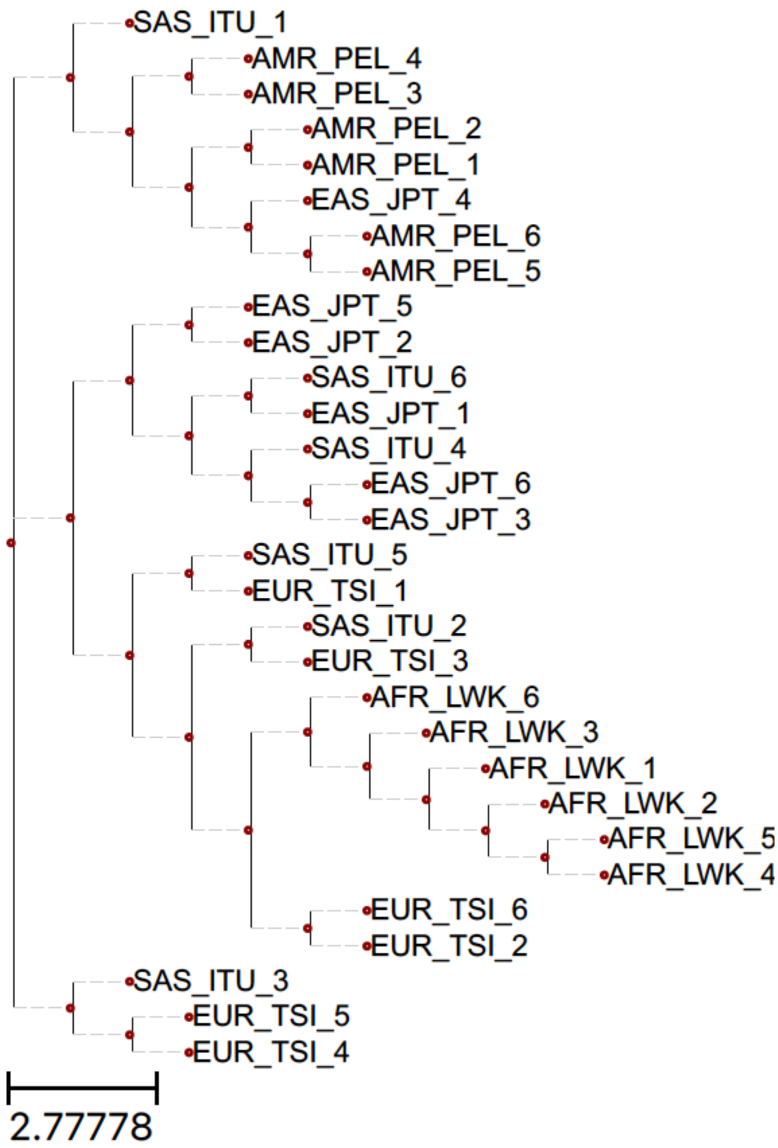

Figure A15: Phylogenetic tree built from pairwise mash distances using k-mer length = 21 and sketch size = 1000 (no monophyly).

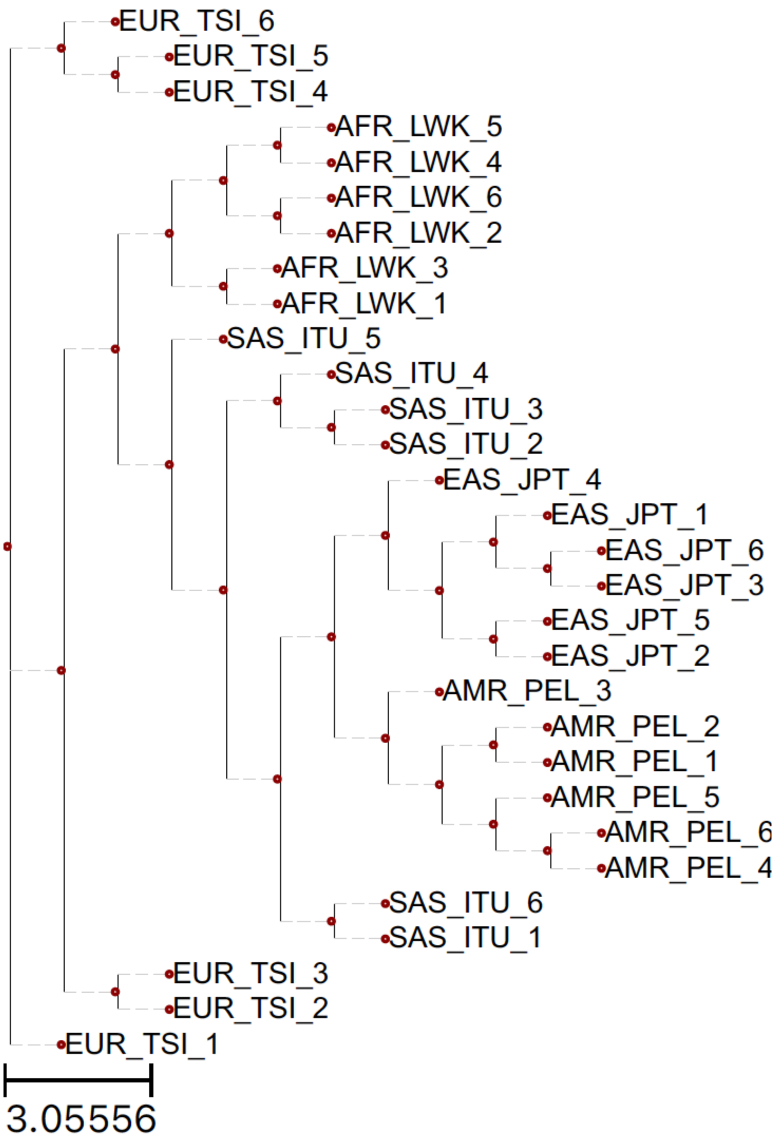

Figure A16: Phylogenetic tree built from pairwise mash distances using k-mer length = 21 and sketch size = 5000 (no monophyly).

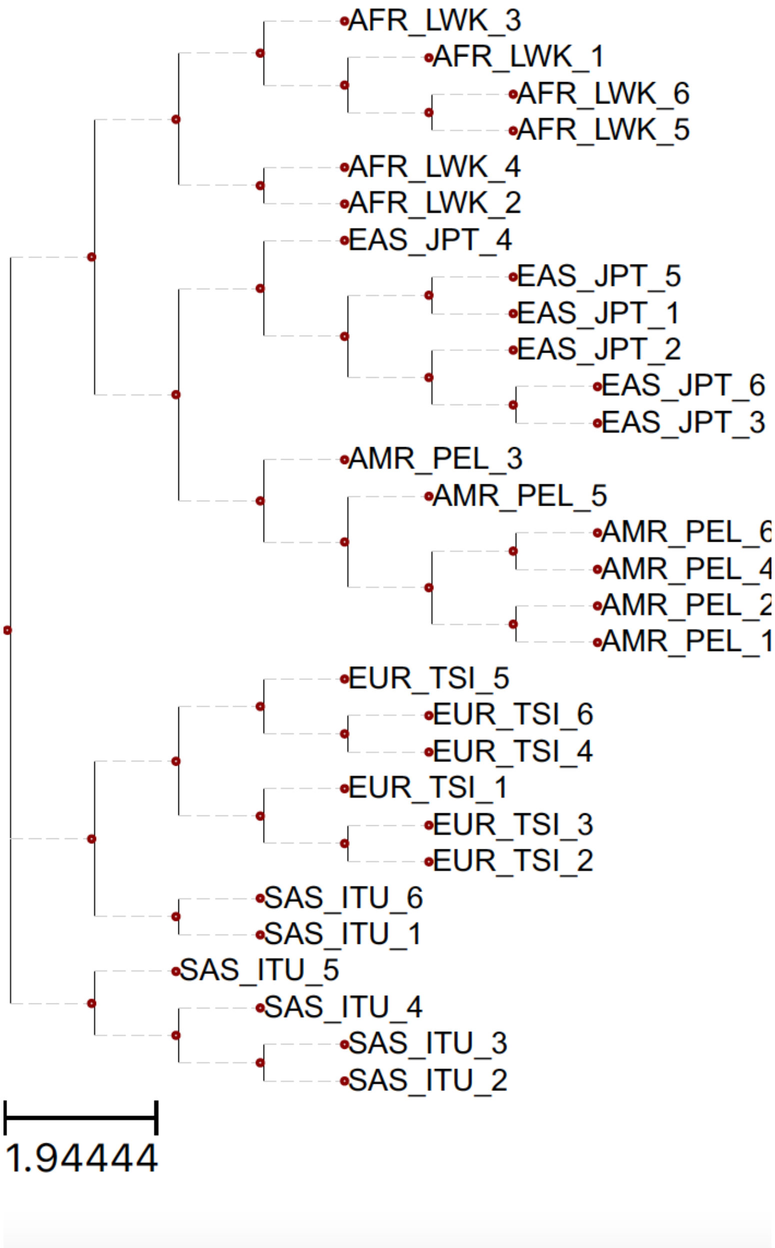

Figure A17: Phylogenetic tree built from pairwise mash distances using k-mer length = 21 and sketch size = 10000 (no monophyly).

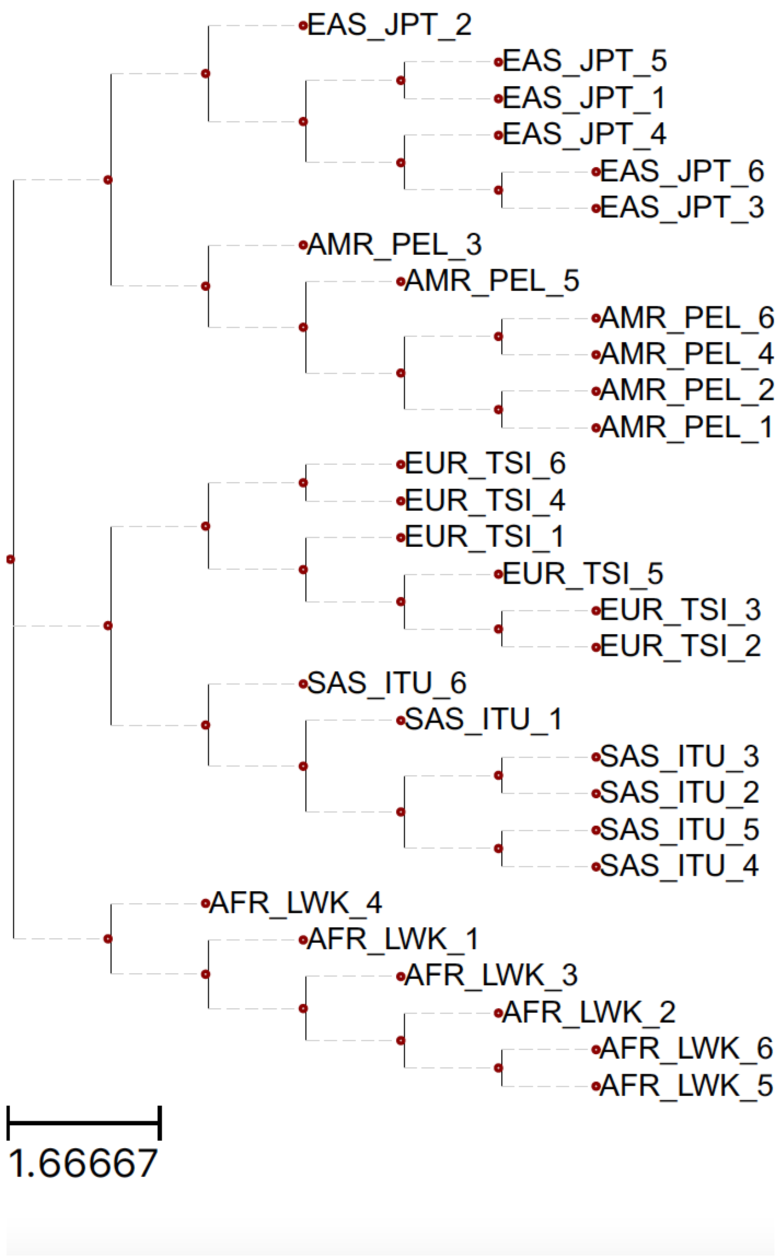

Figure A18: Phylogenetic tree built from pairwise mash distances using k-mer length = 21 and sketch size = 30000 (monophyly observed).