Figure S3  Principal component analysis of SNV data. In order to reduce bias by short range LD, principal component analysis was performed by EIGENSOFT 3.0, using one biallelic SNV with 5-95% allele frequency per 50kb window. (A) First two principal components (PC) for all strains. Both PCs were significant (P<0.001) according to Tracy-Widom statistics. (B) Variance, explained by the individual principal components. (C) First two PCs for strains from clade C (P<0.001, according to Tracy-Widom statistics). The first two PCs reveal a separation between strains sampled from Trois Bassins as opposed to most other locations. (D) Variance, explained by the individual PCs for clade C.