iScience, Volume *26*

Supplemental information

Improving the extraction of ancient *Yersinia pestis* genomes from the dental pulp

Pierre Clavel, Lexane Louis, Clio Der Sarkissian, Catherine Thèves, Claudia Gillet, Lorelei Chauvey, Gaétan Tressières, Stéphanie Schiavinato, Laure Calvière-Tonasso, Norbert Telmon, Benoît Clavel, Richard Jonvel, Stéfan Tzortzis, Laetitia Bouniol, Jean-Marc Fémolant, Jennifer Klunk, Hendrik Poinar, Michel Signoli, Caroline Costedoat, Maria A. Spyrou, Andaine Seguin-Orlando, and Ludovic Orlando

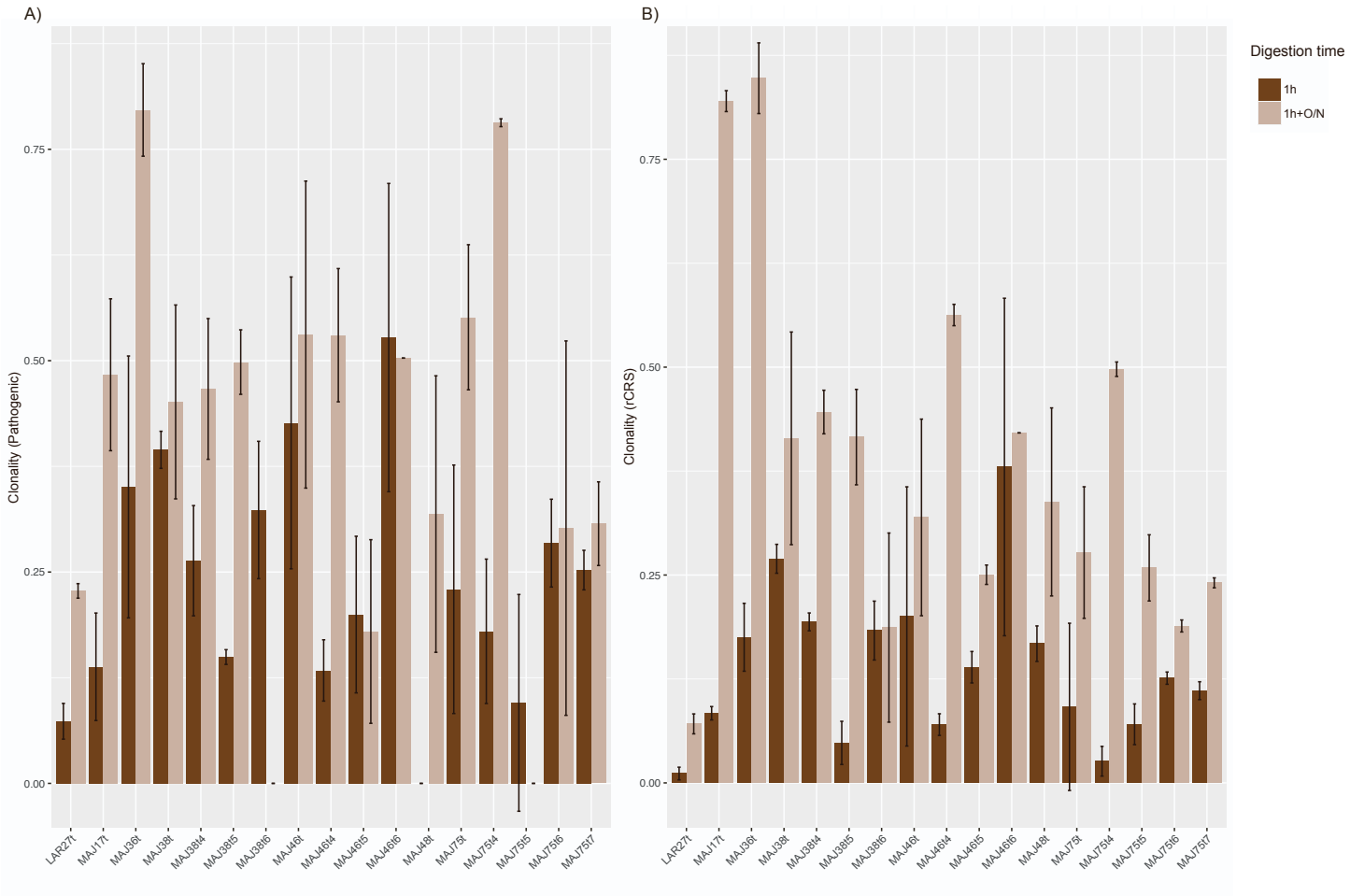

Figure S1. Clonality of unique high-quality alignments. (A, B) One hour of pre-digestion wash (1h, dark brown) versus complete pellet digestion overnight (1h+O/N, light brown). Analyses are repeated 10 times, following random sampling of 4,557 collapsed reads in each experimental condition for normalizing sequencing efforts. Data are represented as mean ± SEM. Related to Figure 3.

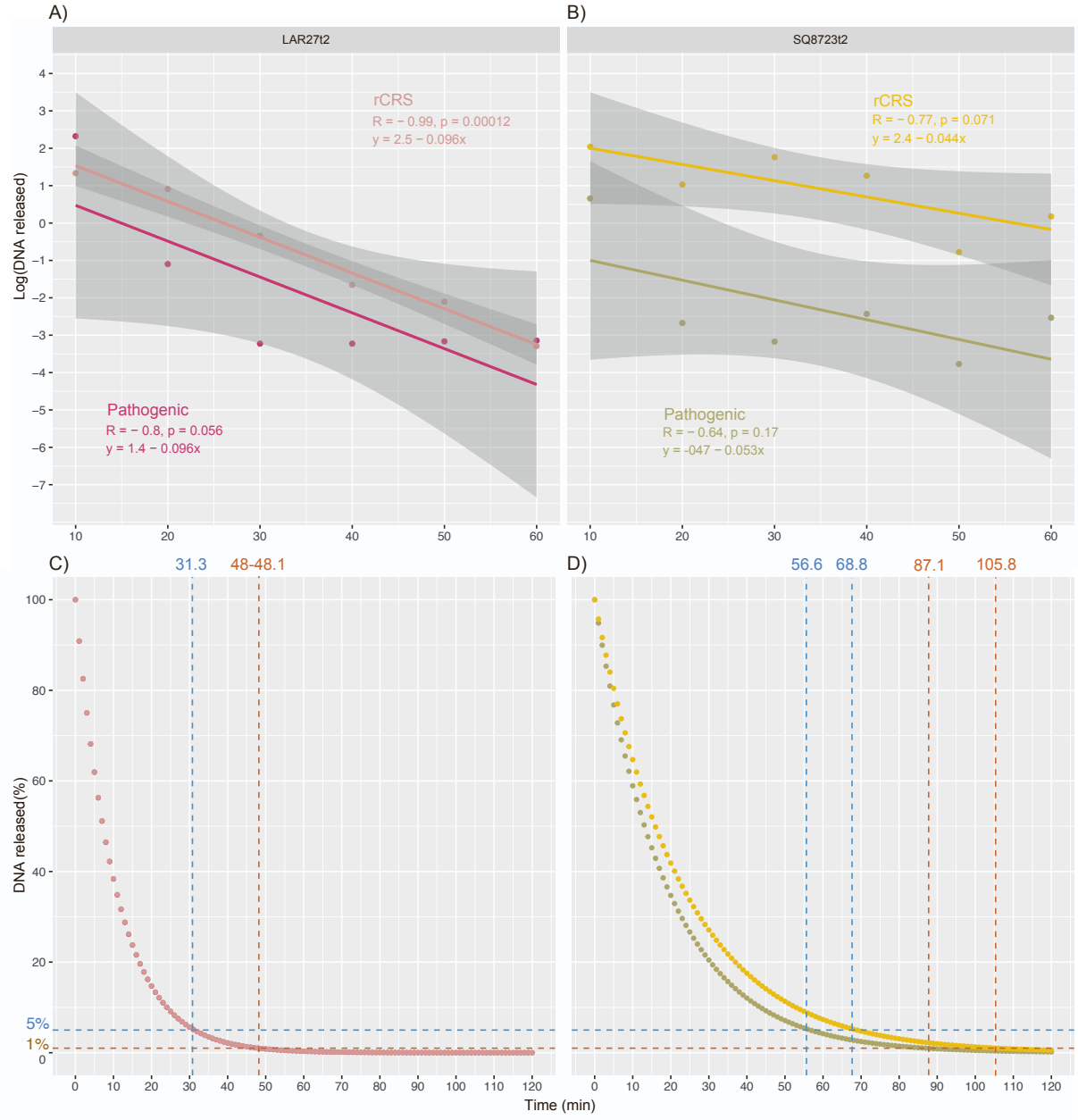

Figure S2. Kinetics of DNA release during pre-digestion wash. (A, B) Percentage of DNA released following pre-digestion wash of incremental time (10 min to 1 hour). The results are shown for the LAR27t2 **(A)** and SQ8723t2 **(B)** teeth. High-quality sequence alignments are aligned against the pathogen genome (CO92 and plasmids, purple and green) or the human mtDNA genome (rCRS, pink and yellow). **(B, C)** Prediction for the DNA fraction remaining to be released following incremental pre-digestion wash. Predictions are based on the linear models shown in panels A and B for *Y. pestis*

(purple and green) and human mtDNA (pink and yellow). The dashed lines indicate in blue and orange the times when respectively 95 % and 99 % of the DNA has been released. Related to Figure 4.

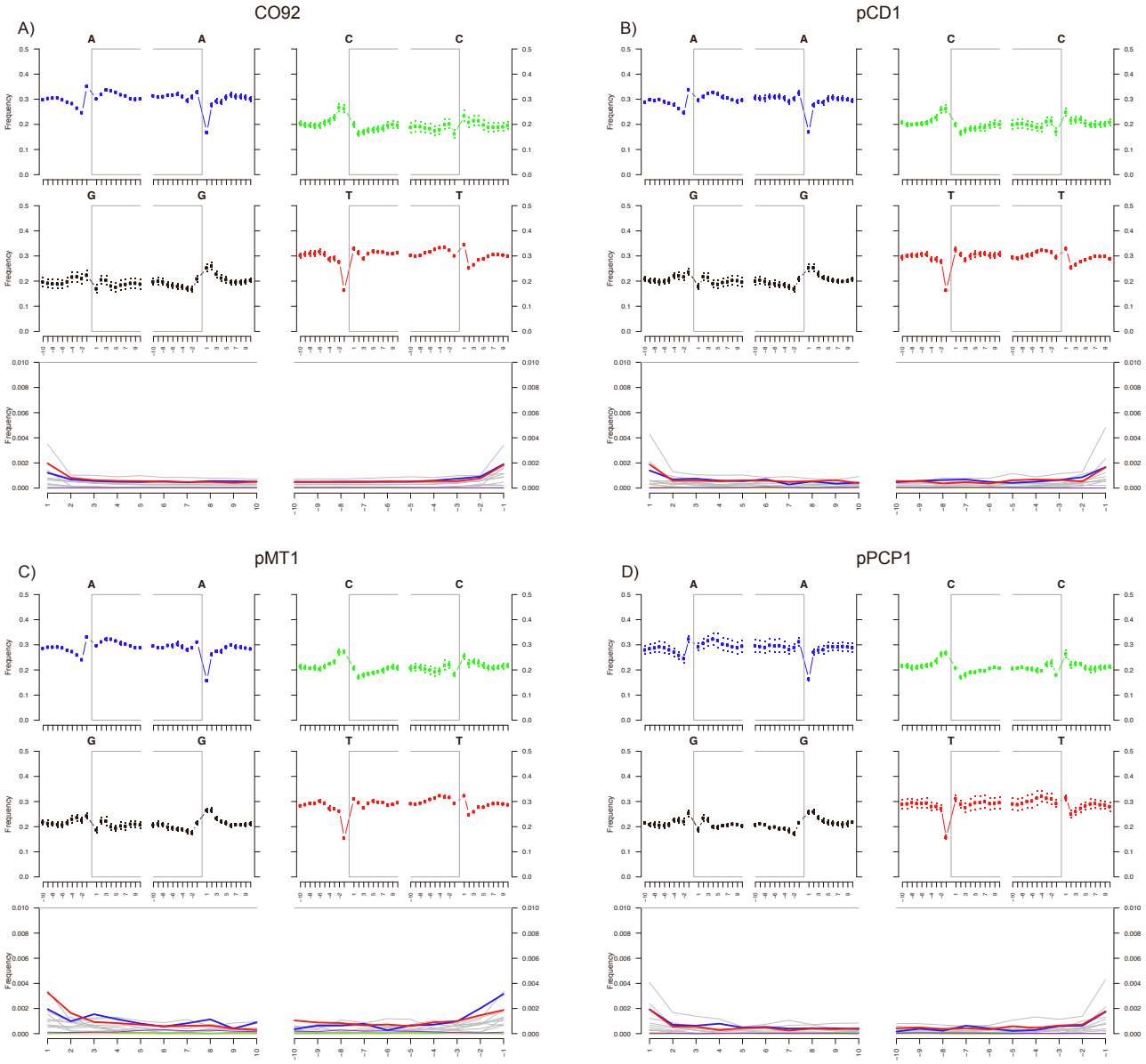

Figure S3. Base composition and nucleotide mis-incorporation profiles. Typical post-mortem DNA damage patterns for AMIENS268 using mapDamage2 with default parameters, excepting that the -m and -y options were turned to 10 and 0.01, respectively[13]. The base composition of the ten positions preceding and following read starts is provided (-10 to -1 and 1 to 10, respectively). The base composition of the ten positions preceding and following read ends is provided (-10 to -1 and 1 to 10, respectively). Rates and within-read positions of nucleotide mis-incorporation after USER treatment are

indicated for **(A)** the CO92 chromosome and the plasmids **(B)** pCD1, **(C)** pMT1 and **(D)** pPCP1. Related to Figure 7.

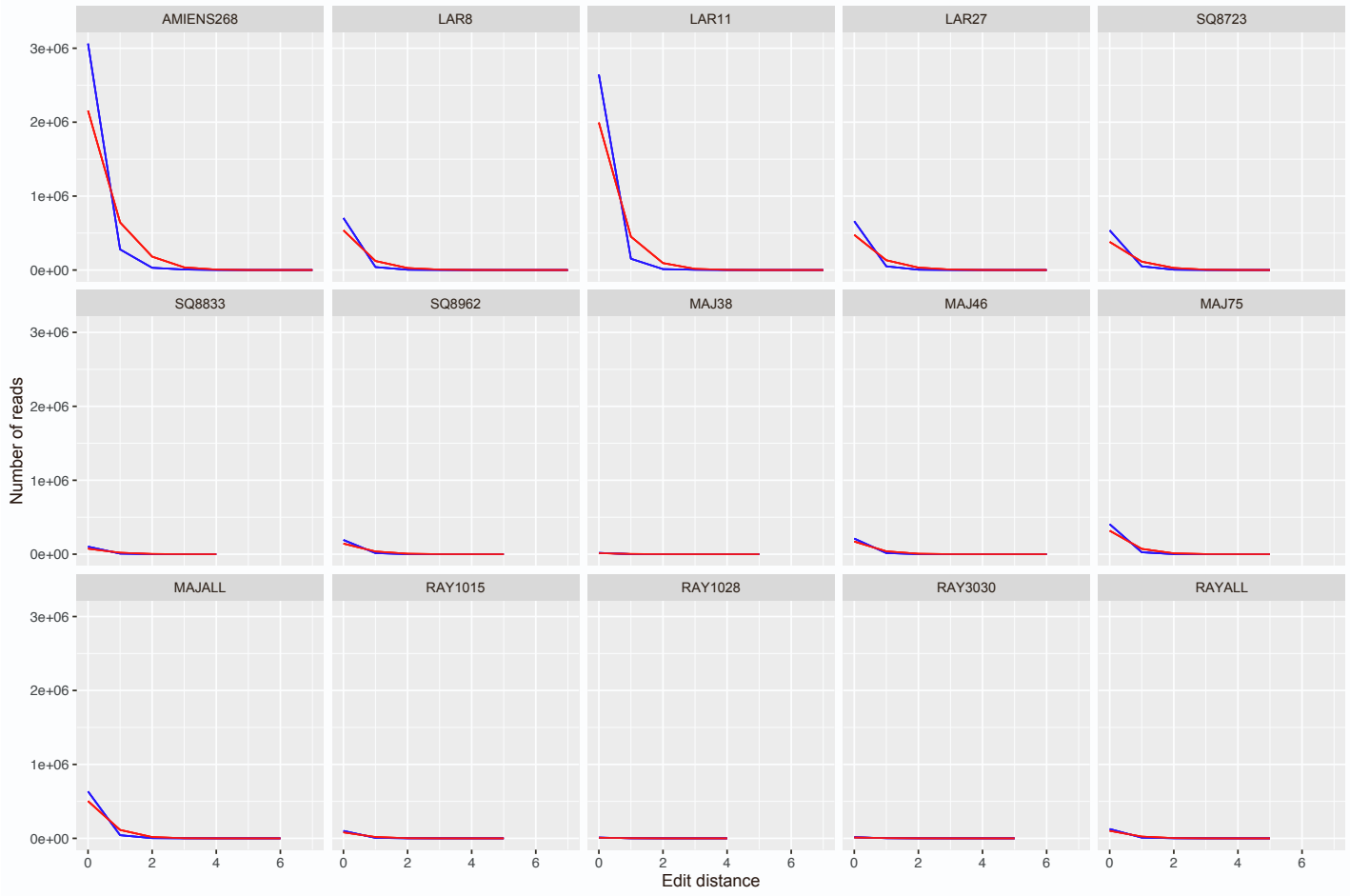

Figure S4. Edit distance distributions. Calculations were based on reads mapping against the CO92 *Y. pestis* genome (blue) and the *Y. pseudotuberculosis* reference genome (red), and restricted to the samples investigated in this study. Edit distance distributions confirmed the genetic proximity to *Y. pestis* relative to the close outgroup *Y. pseudotuberculosis*. The 'ALL' suffix refers to the merging of all sequences of the three samples from La Major (MAJ38, MAJ46 and MAJ75) and Les Rayettes (RAY1015, RAY1028 and RAY3030) sites. Related to Figure 7.

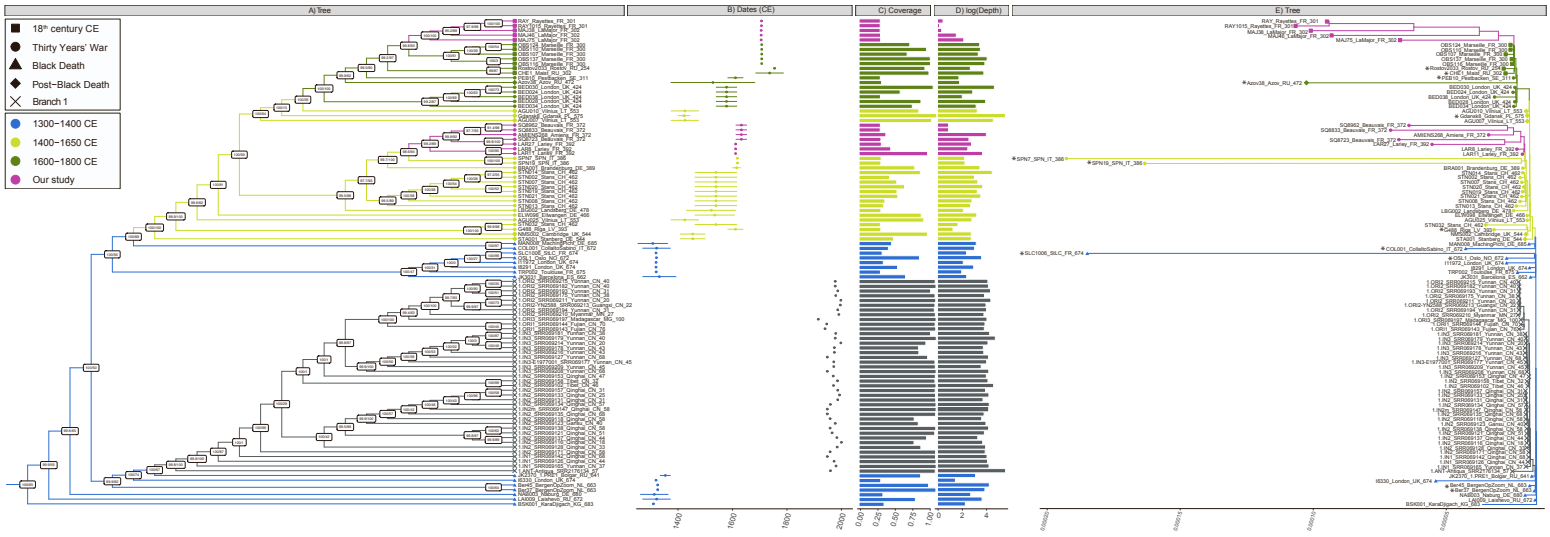

Figure S5. Phylogenetic reconstruction. (A) Maximum Likelihood tree (IQ-TREE), considering 29,609 genome-wide variant positions known to provide reliable phylogenetic signal. This analysis included 185 previously published genomes, but the tree shown is restricted to second and third pandemic genomes, disregarding branch lengths, for clarity. No branch lengths are reported to increase readability and node supports are estimated from SH-aLRT calculations (left) and ultrafast bootstrap approximation (right). **(B)** Sample dates, with error margins representing the lower and upper boundaries provided by archaeological contexts. **(C)** and **(D)** represent the coverage and the logarithm of the average sequencing depth-of-coverage achieved for the different CO92 genomes considered. **(E)** Maximum Likelihood tree (IQ-TREE), considering the entire genome with branch length (non-USER treated samples are indicated with a star). The colors illustrated follow the coloring scheme from Spyrou et al., 2022[14] and the tip shapes indicate the historical context of archaeological samples investigated with respect to Figure 1. Related to Figure 5.

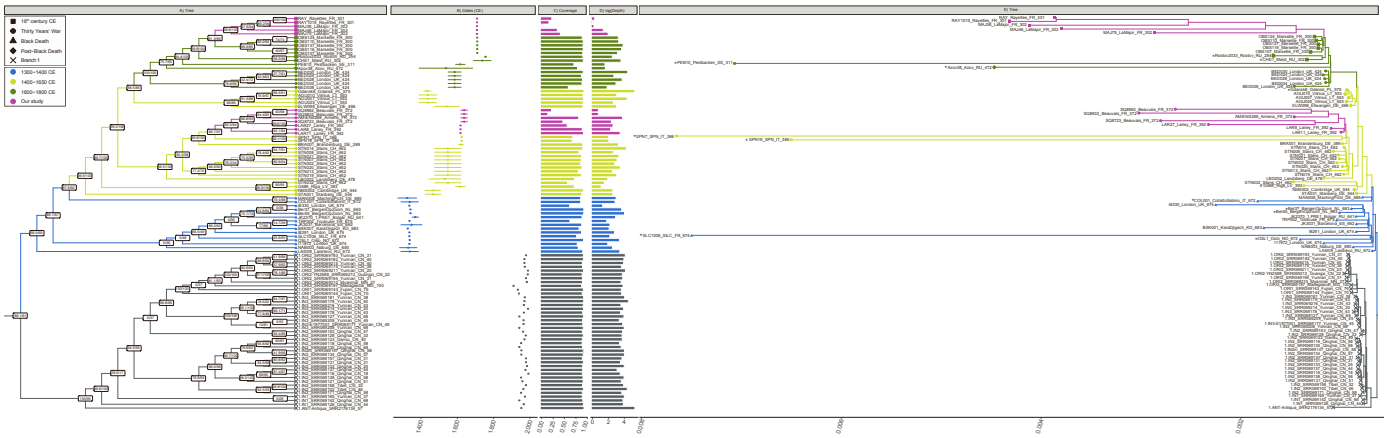

Figure S6. Phylogenetic reconstruction. (A) Maximum Likelihood tree (IQ-TREE), considering the entire genome (CO92). This analysis included 185 previously published genomes, but the tree shown is restricted to second and third pandemic genomes, disregarding branch lengths, for clarity. No branch lengths are reported to increase readability and node supports are estimated from SH-aLRT calculations (left) and ultrafast bootstrap approximation (right). **(B)** Sample dates CE, with error margins representing the lower and upper boundaries provided by archaeological contexts. **(C)** and **(D)** represent the coverage and the logarithm of the average sequencing depth-of-coverage achieved for the different CO92 genomes considered. **(E)** Maximum Likelihood tree (IQ-TREE), considering the entire genome with branch length (non-USER treated samples are indicated with a star). The colors illustrated follow the coloring scheme from Spyrou et al., 2022[14] and the tip shapes indicate the historical context of archaeological samples investigated with respect to Figure 1. Related to Figure 5.

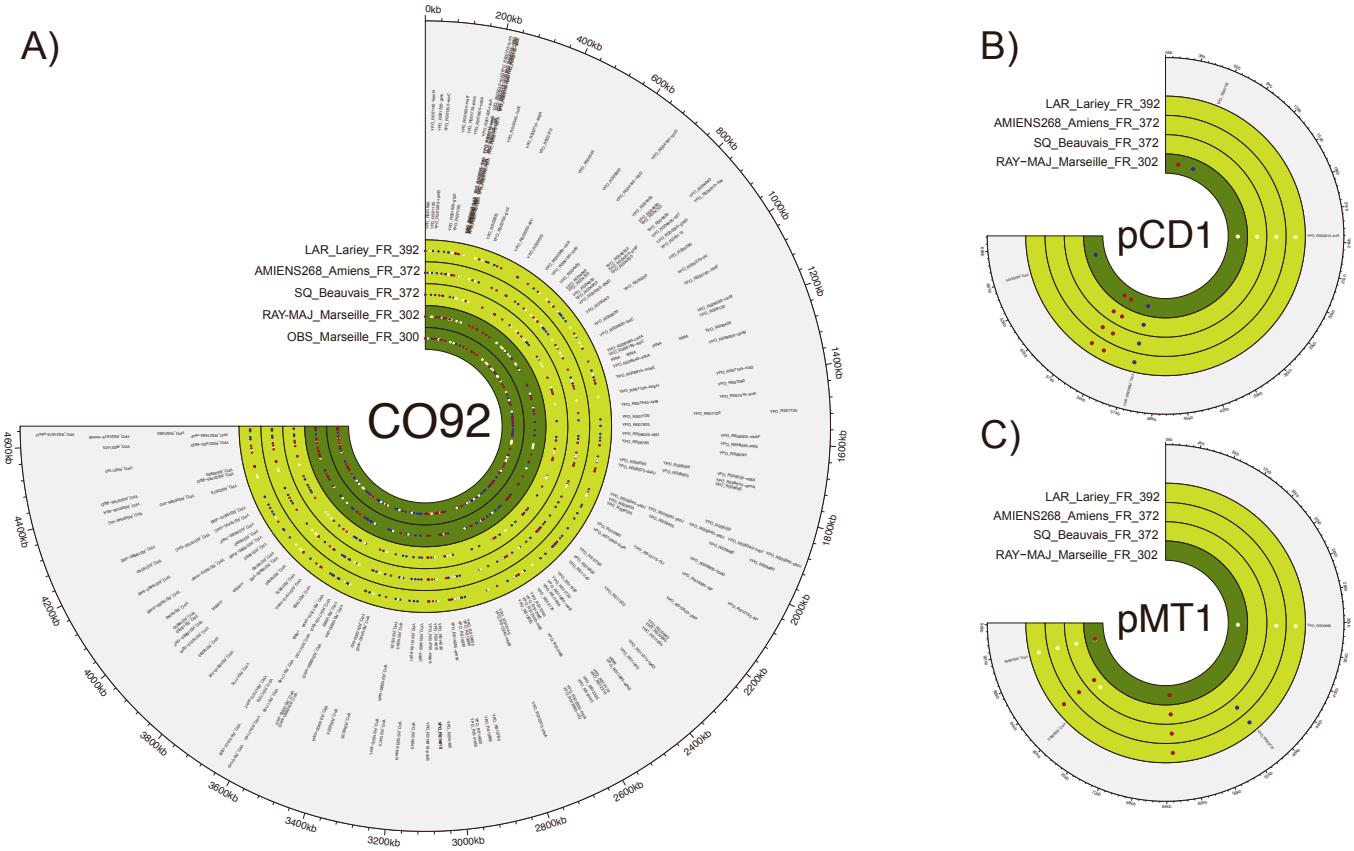

Figure S7. SNP variants. SNP variation was detected using snpToolkit on samples from the 17th c. CE (yellow) and the 18th c. CE (green) on the circular chromosome **(A)**, the pCD1 **(B)** and pMT1 **(C)** plasmids. Non-synonymous SNPs are indicated by blue dots, synonymous SNPs by white dots and intergenic SNPs by red dots. Circular plots were made using the R package circlize. Related to Figure 7.