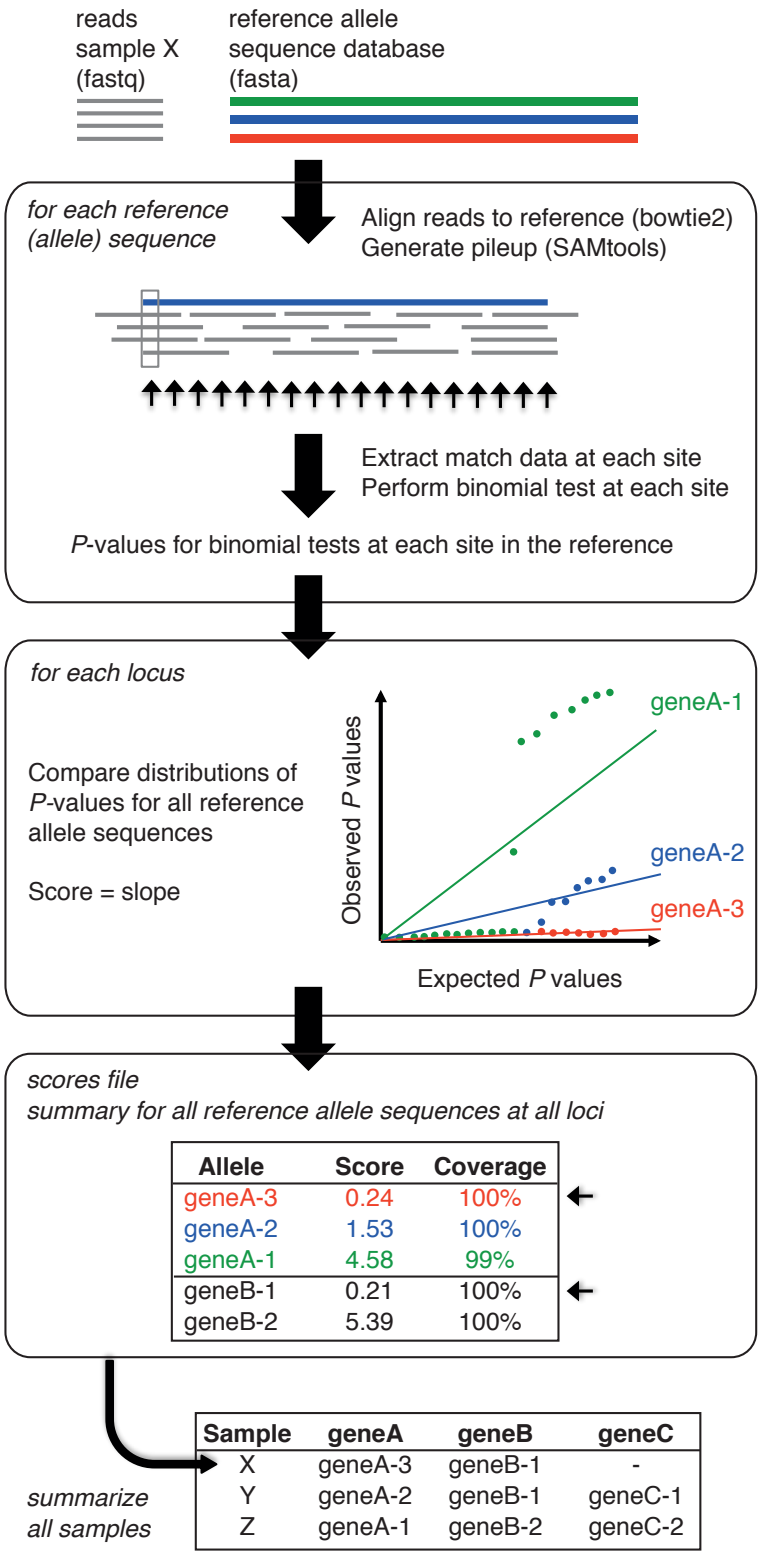
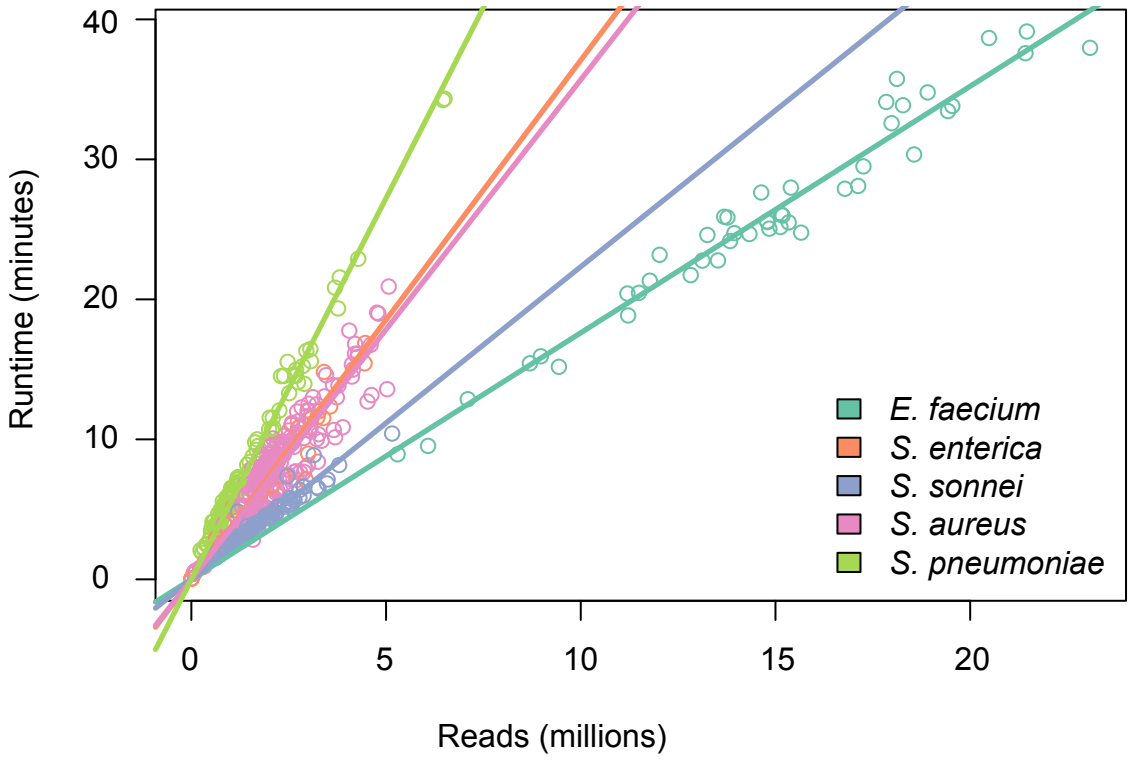
Supplementary Figures

Supplementary Figure 1: Summary of SRST2 approach



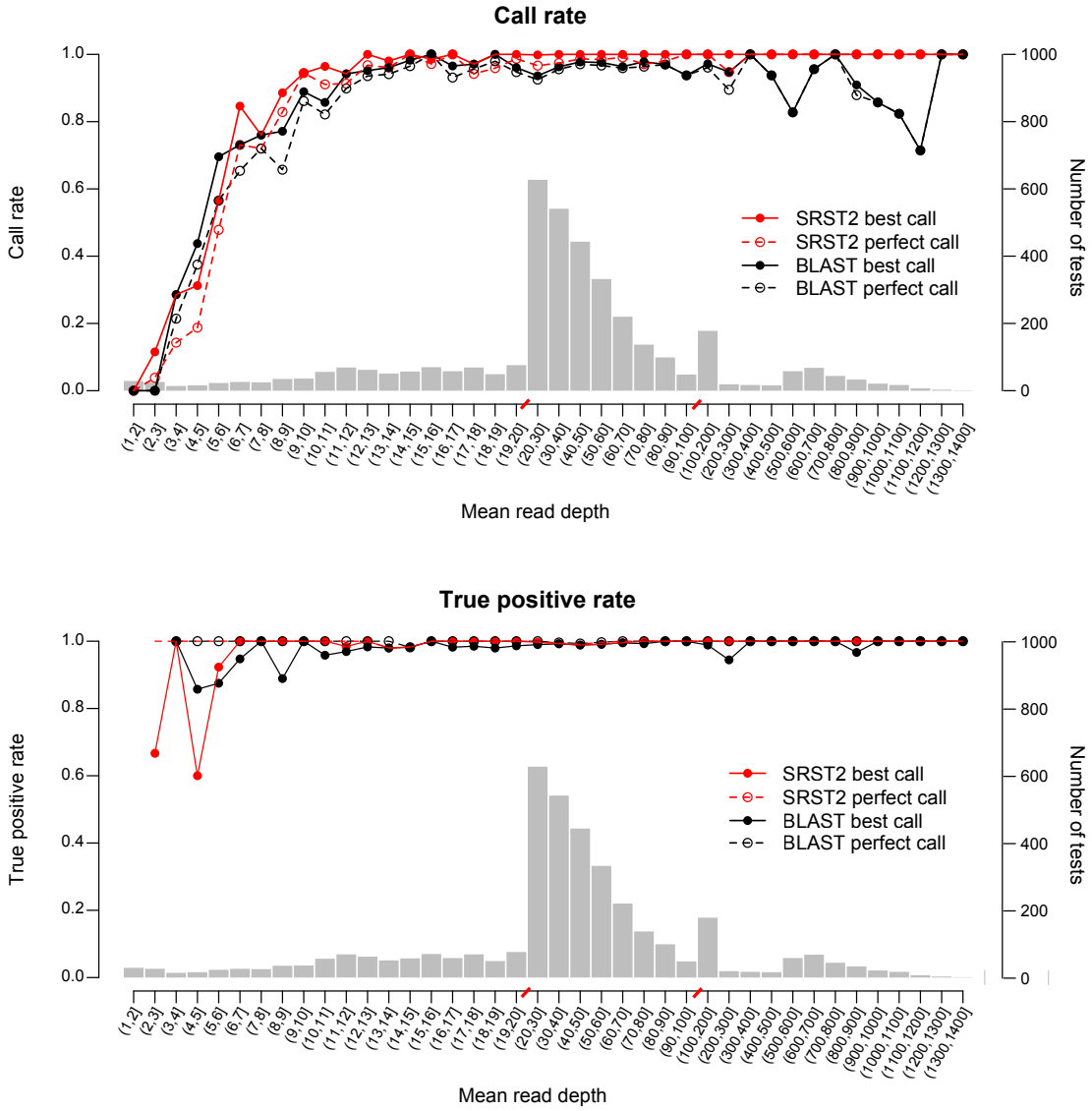
Supplementary Figure 2: Run times for MLST analysis with SRST2

Lines are linear regression of runtime on reads, calculated separately for each species from public datasets (details in **Supplementary Table 1**).



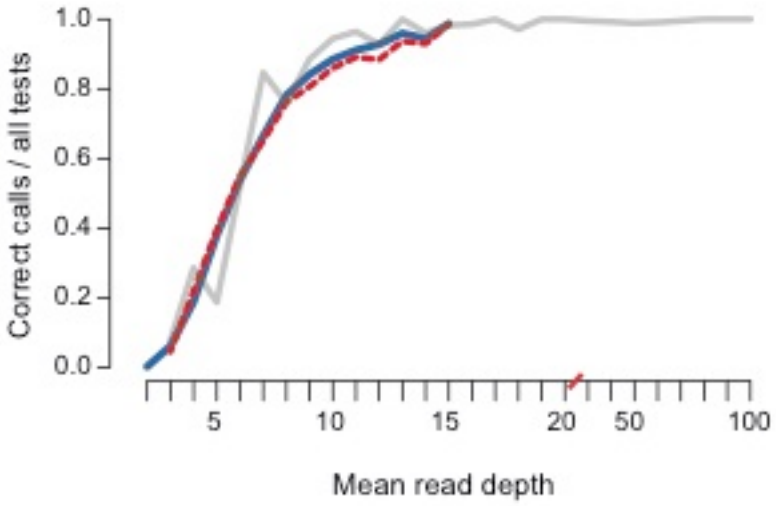
Supplementary Figure 3: Accuracy of allele calling using SRST2 vs assembly and BLAST

MLST analysis of public data from 5 species (N=543 genomes, 3801 loci, details Supplementary Table 1). Tests were grouped by read depth and accuracy rates (left y-axis, correct allele calls as a proportion of tests), calculated at each depth (x-axis, red slashes indicate scale change). Grey bars, number of tests at each depth (right y-axis); Lines, accuracy of allele calling. (**A**) Call rate (total allele calls / 3801). (**B**) True positive rate (correct allele calls / total allele calls).



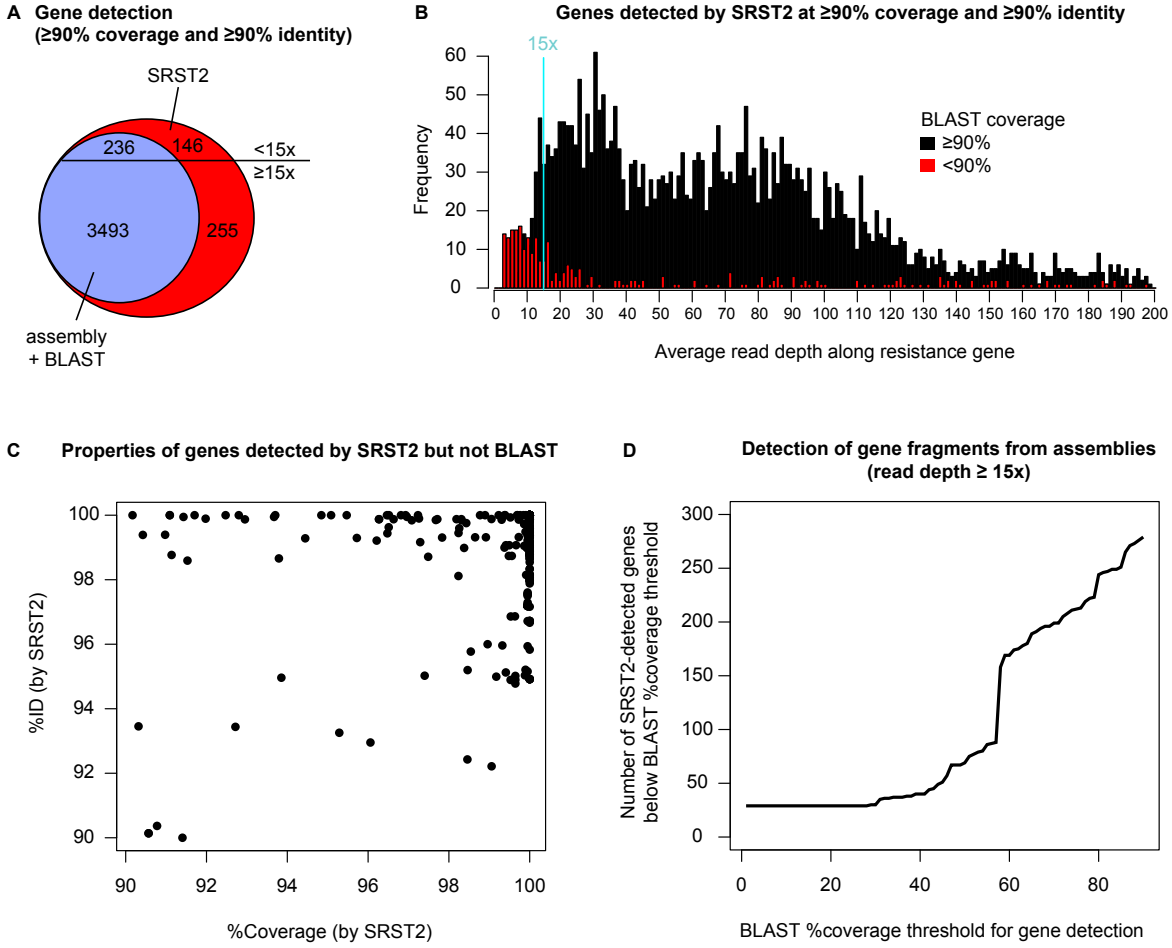
Supplementary Figure 4: Accuracy of SRST2 allele calling at low read depths and with expanded MLST database size

MLST analysis of public *S. aureus* data. (N=10 read sets; each sampled 100 times to different depths; details in Methods). Tests were grouped by read depth and accuracy rates (y-axis, correct allele calls as a proportion of all tests), calculated at each depth (x-axis, red slashes indicate scale change from 1x to 10x). Red, real *S. aureus* MLST database; blue, expanded *S. aureus* MLST database (see Methods); grey, unsampled data from 5 species mapped to real databases (as shown in **Fig. 1, S1**).



Supplementary Figure 5: Resistance gene detection

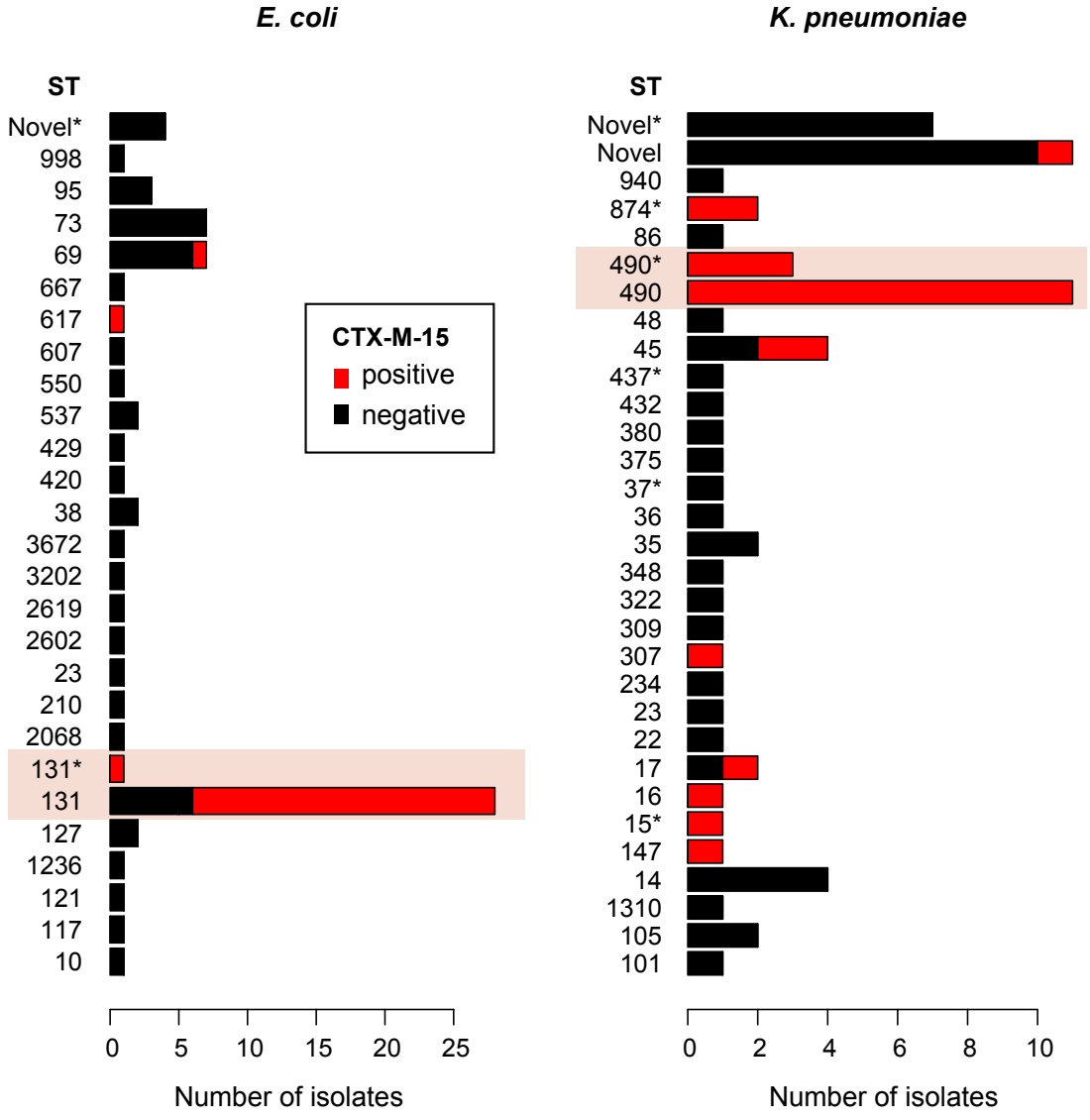
(**A**) Venn diagram of antimicrobial resistance genes detected by SRST2 and assembly+BLAST, where the threshold for 'detection' of a gene is ≥90% coverage and ≥90% identity with a reference allele. No genes were detected by assembly+BLAST but not SRST2. (**B**) Distribution of average read depths per gene, calculated by SRST2 from mapped reads, for all genes detected by SRST2. (**C**) Coverage and nucleotide identity (%ID), as calculated by SRST2, for all genes detected by SRST2 but not by assembly+BLAST. (**D**) Impact of lowering the coverage threshold for detection of genes by BLAST (for those genes with ≥15x read depth).



Supplementary Figure 6: SRST2 analysis of sequence types and beta-lactamase CTX-M-15 genes amongst hospital isolates

Rates of isolation of different sequence types (STs), coloured by CTX-M-15 status, as determined by SRST2 run with default parameters on a public data set of strains from a single hospital. In each species, a single known ST dominates the population (highlighted) and is also the dominant source CTX-M-15 genes.

A '*' next to an ST indicates a match to the closest defined ST; i.e., that for all 7 loci the closest known allele is the one belonging to that ST, however at ≥1 these loci there is an imprecise match (SNP or indel) compared to the known allele sequence. 'Novel' indicates a novel sequence type resulting from a combination of known alleles, with precise matches at all loci ('NF' in SRST2 output); 'Novel*' indicates a novel combination of alleles, with ≥1 of those alleles being novel itself (i.e. with no exact match in the MLST database) ('NF*' in SRST2 output).



Supplementary Table 1

(A) Validation data

Species	Citation	N	Population	Sequencing Centre	Average read depth	Read length
Staphylococcus aureus	Holden et al, *Genome Res* 2013, 23(4):653-64	134	Clonal, ST22	Sanger, UK	24x	55
Staphylococcus aureus	Castillo-Ramirez et al, *Genome Biol* 2012, 13(12):R126	128	Clonal, ST239	Sanger, UK	60x	65
Streptococcus pneumoniae	Croucher et al, *Science* 2011, 331(6016):430-4	113	Clonal, ST81	Sanger, UK	30x	55
Salmonella enterica Typhimurium	Okoro et al, *Nat Genet* 2012, 44(11):1215-21	44	Clonal, ST313	Sanger, UK	34x	76
Shigella (E. coli)	Holt et al, *Nat Genet* 2012 44(9):1056-9	81	Clonal, *S. sonnei*	Sanger, UK	25x	55
Enterococcus faecium	Howden, 2013 [14]	43	Diverse, dominated by ST203, ST17	Melbourne, Australia	658x	101
Listeria monocytogenes	This paper	231	Diverse	Melbourne, Australia	36x	152

(B) Demonstration data (public)

Species	Citation	N	Average read depth	Read length
Enterococcus faecium (Fig 2a-c)	Howden 2013 [14]	43	658	101
Hospital outbreak investigations (Fig 2d-e)	Reuter, 2013 [5]	20	36x	151
K. pneumoniae, E. coli	Stoesser 2013 [6]	69, 74	34x	101