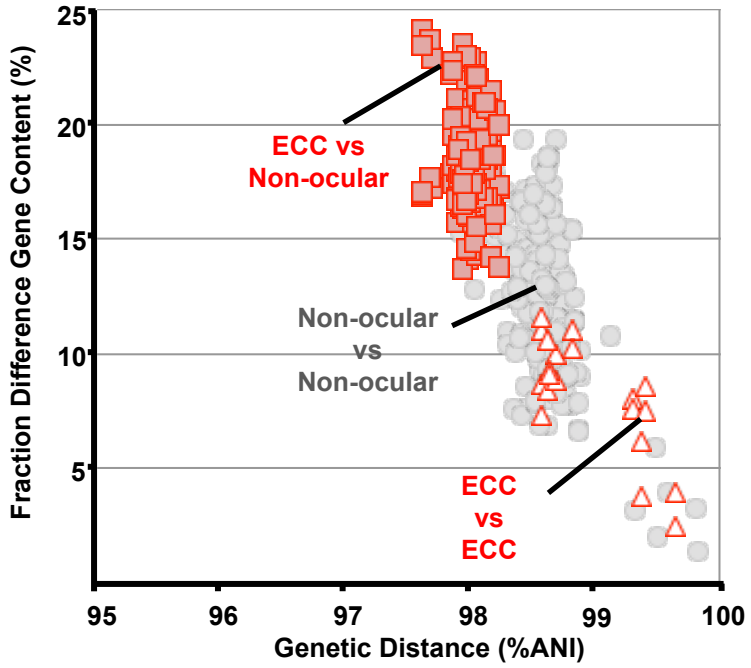
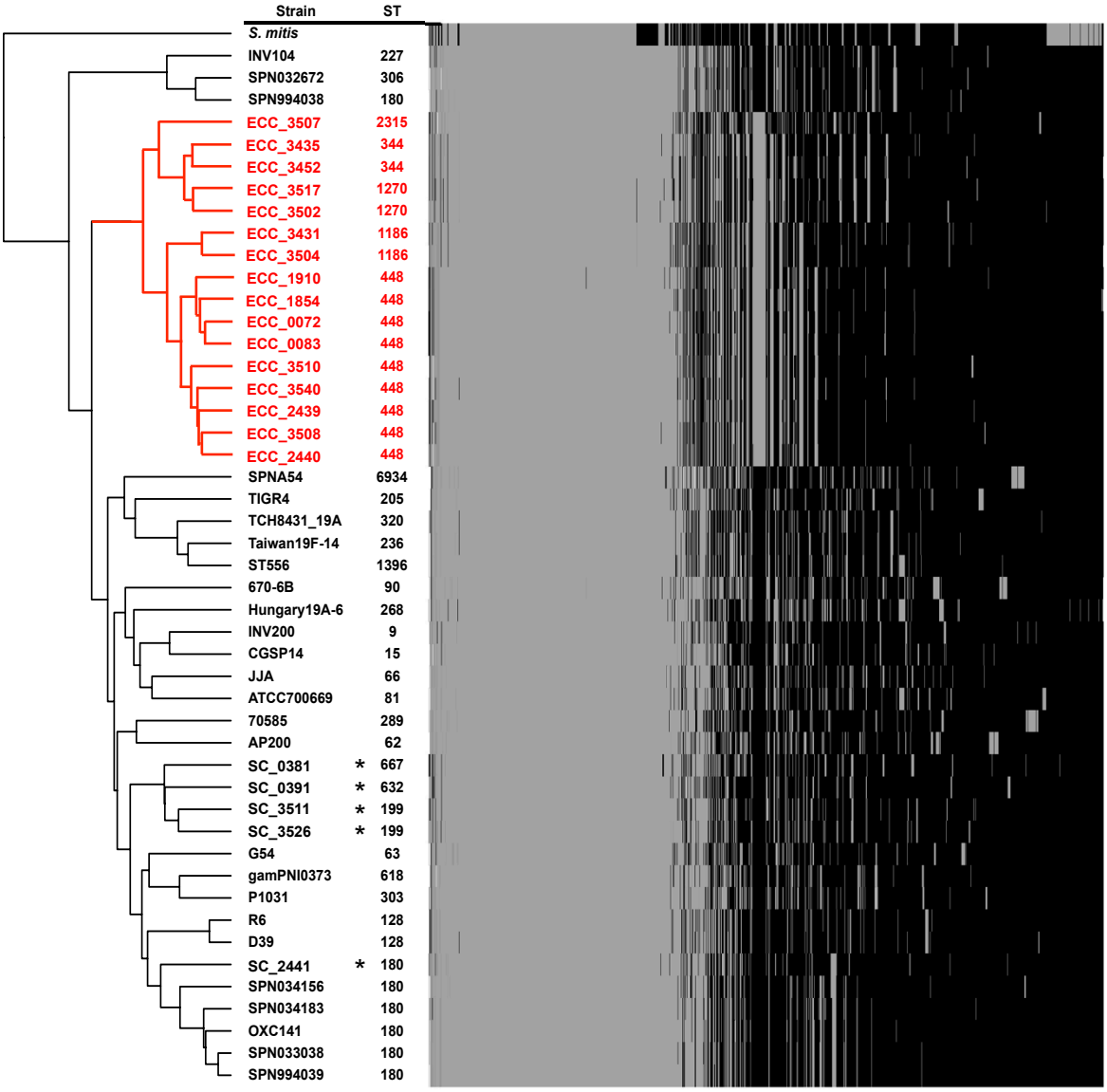
Supplementary Figure-1 GILMORE



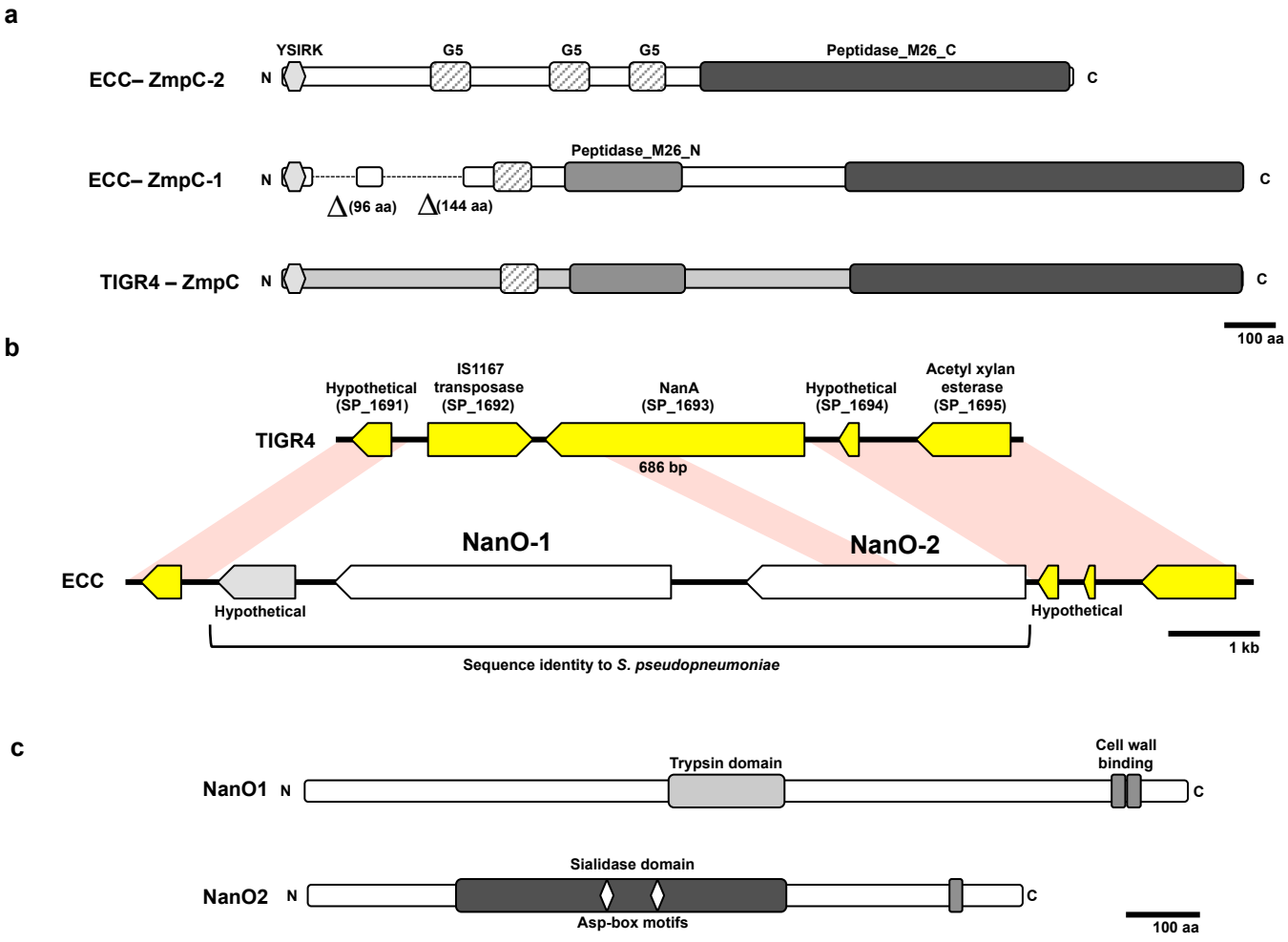
Supplementary Figure 1. Average nucleotide identity (ANI) analysis. Each point represents a pairwise comparison of two genomes.

Supplementary Figure-2 GILMORE



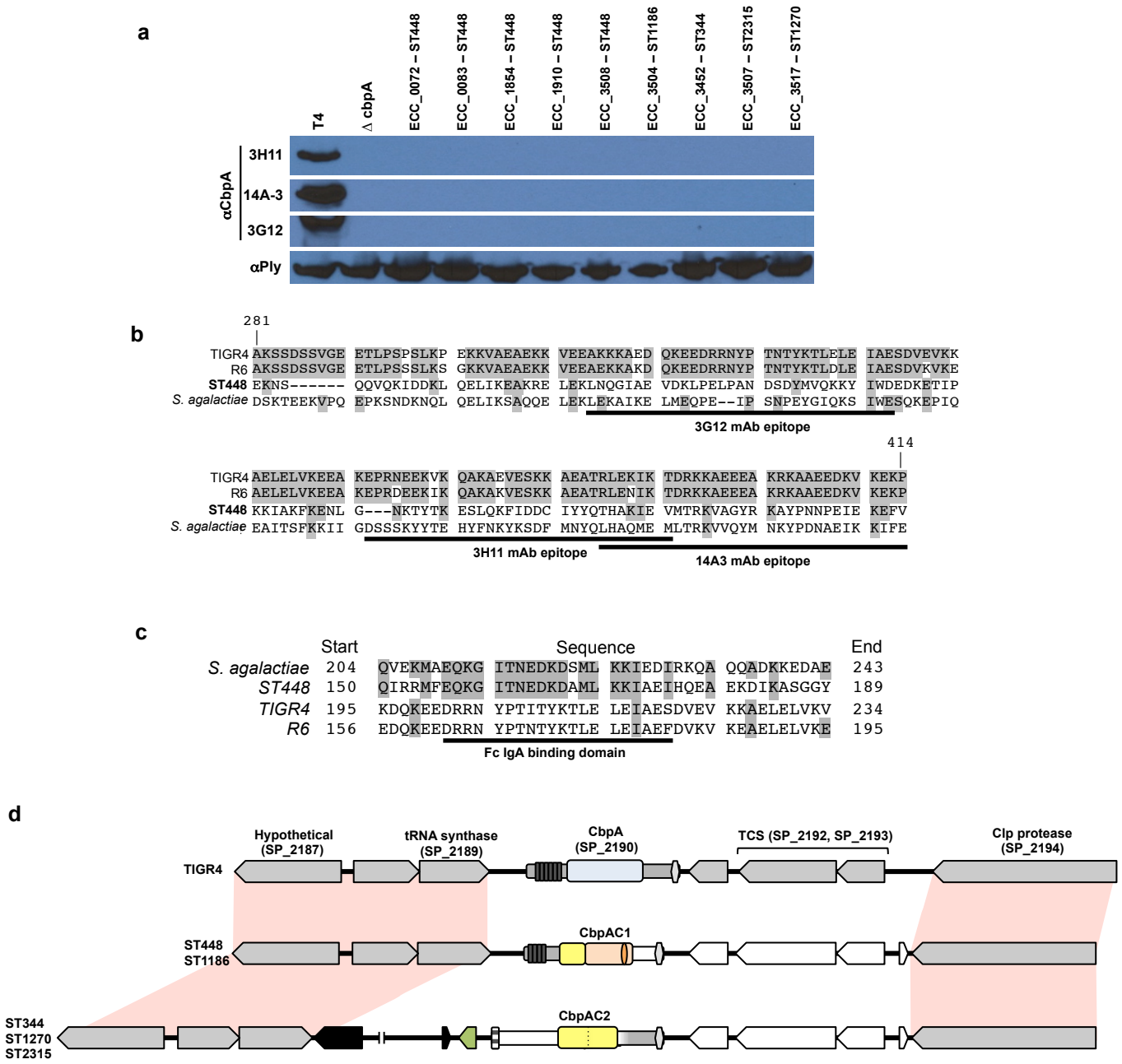
Supplementary Figure 2. Phylogenetic reconstruction based upon the pattern of gene presence and absence substantiates the divergence between ECC and non-ocular *S. pneumoniae*. Gene presence (gray line) and absence (black line) was identified across each of the genomes. Strains that are rarer causes of conjunctivitis are denoted with an asterisk.

Supplementary Figure-3 GILMORE



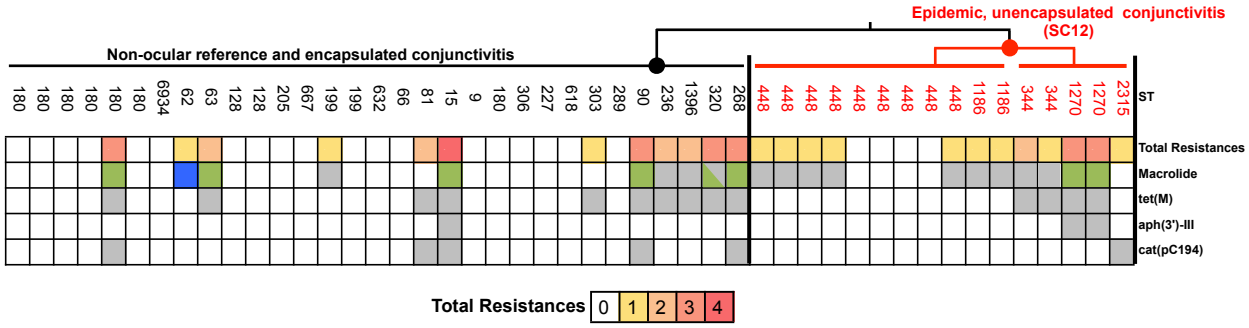
Supplementary Figure 3. ECC genomes encode unique virulence factors. Illustration of proteins unique to ECC genomes. Regions of shared identity between protein or gene sequences are shown with red highlight. (**a**) Domains predicted within the ZmpC1 and ZmpC2 proteins of ECC strains and ZmpC of TIGR4. Two deletions in ECC ZmpC1 compared to ZmpC of TIGR4 are shown (Δ). ZmpC comparison between conjunctivitis isolate and TIGR4 adapted from Menon *et al.*[1]. (**b**) Nucleotide synteny at the *nanA* locus in the non-ocular reference TIGR4 and the NanO1/NanO2 sialidases encoded only within ECC strains, likely acquired by recombination with *S. pseudopneumoniae*. (**c**) Predicted domains encoded within the ECC sialidases NanO1 and NanO2.

Supplementary Figure-4 GILMORE



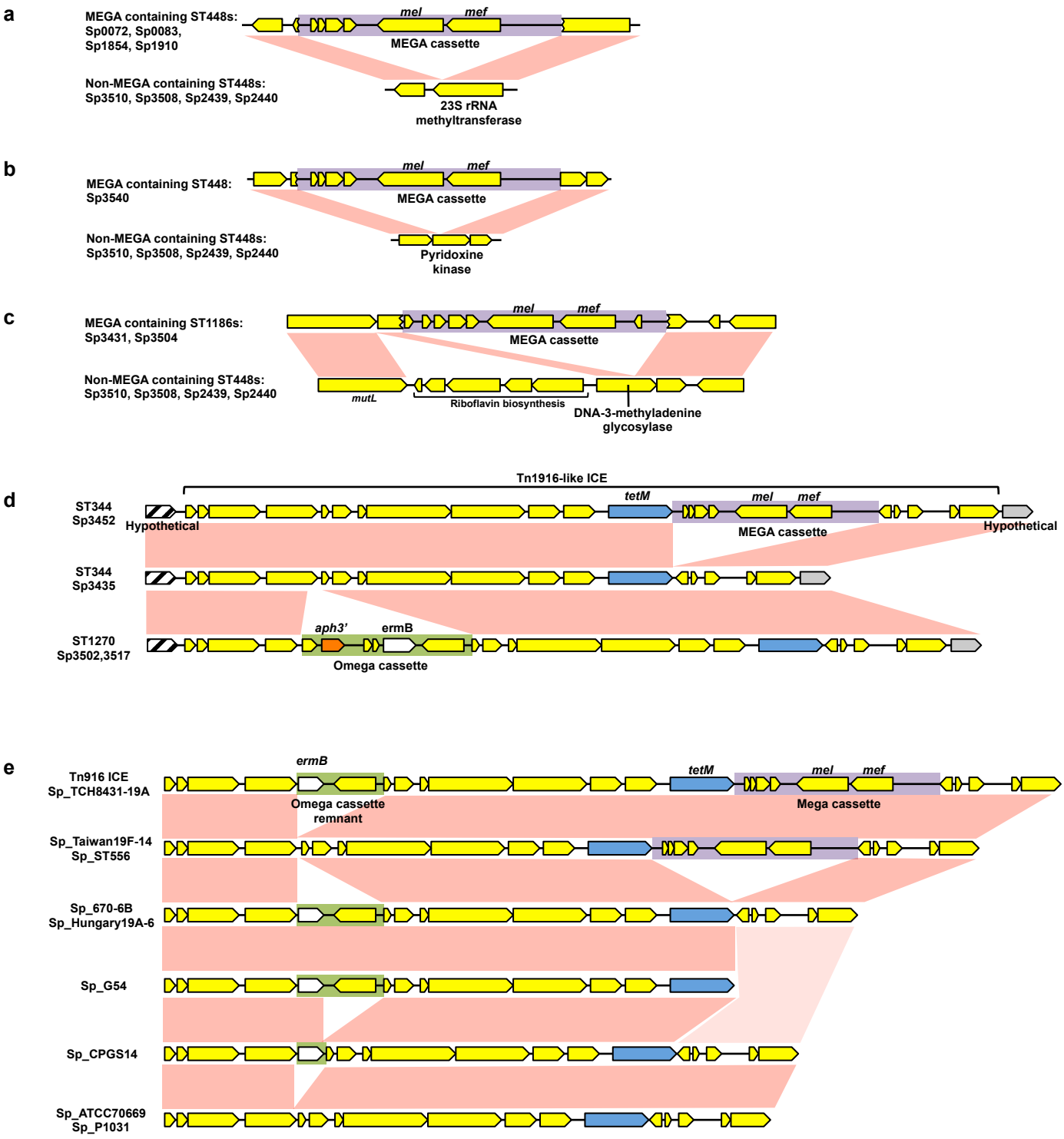
Supplementary Figure 4. ECC genomes carry an atypical virulence factor at the CbpA locus. (**a**) Western blot of total bacterial extracts, probed with three monoclonal antibodies (clones: 3H11, 14A-3, 3G12 see Mann *et al*.[2]) that recognize highly conserved epitopes within CbpA (see panel B). T4, TIGR4; delta-*cbpA*, TIGR4 with CbpA deleted. Pneumolysin served as positive control for protein loading (anti-Ply). See Supplementary Figure 8 for full blots. (**b**) CbpA sequence of *S. pneumoniae* strains TIGR4 and R6, ST448, and *S. agalactiae* strain A909 Beta-antigen binding protein. Peptide sequences in agreement with the TIGR4 reference are highlighted. The anti-CbpA monoclonal epitopes from (a) are denoted. (**c**) Sequence of the IgA-Fc binding domain of Beta-antigen across *S. agalactiae*, ST448 and *S. pneumoniae* strains TIGR4 and R6. Peptide sequences in agreement with the *S. agalacticae* A909 reference are highlighted in gray. (**d**) Genome synteny surrounding CbpA-C1 and CbpA-C2, including the TCS system controlling CbpA expression, as compared to TIGR4 reference. CbpA, CbpA-C1, CbpA-C2 illustrated as described in Fig. 5. Sequence associated with *S. pneumoniae* (gray), *S. agalactiae* (white), *S. pseudopneumoniae* (black), and sequence that could not be assigned to an organism (green) are indicated. Nucleotide sequence with identity is shown in pink. Loss of nucleotide synteny in the ECC genomes compared to TIGR4 occurs exactly at the end of SP_2189 and SP_2194, and is shared at the exact location in all ECC genomes.

Supplementary Figure-5 GILMORE



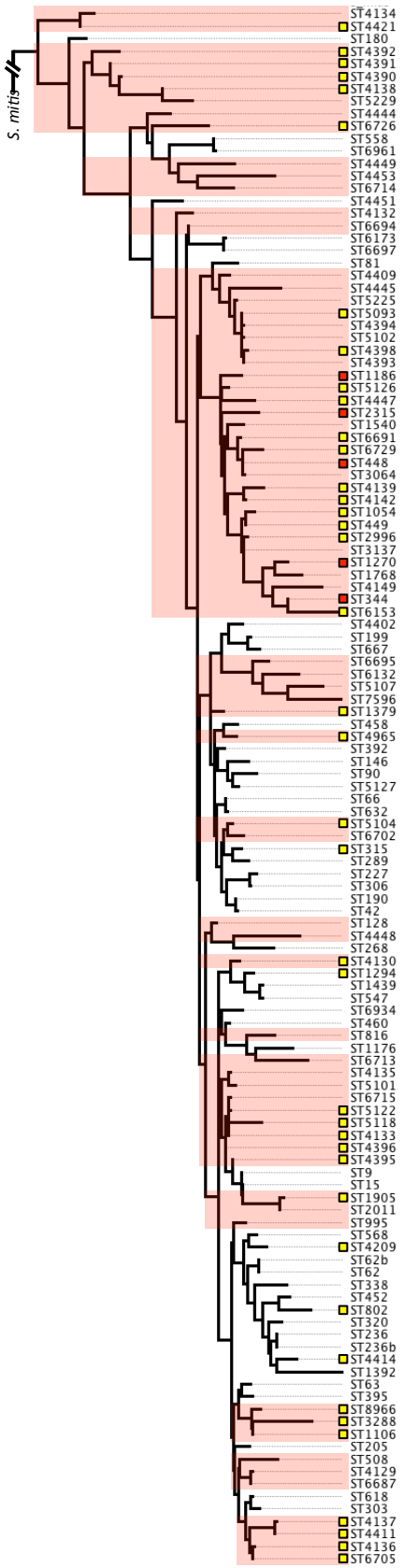
Supplementary Figure 5. Summary of resistance genes. Each box represents a strain. Specific organization of the strains within the tree can be found in Figure 3 and Supplementary Data 2. Macrolide resistances: *mef/mel* (gray), *ermB* (green), and *ermA* (blue).

Supplementary Figure-6 GILMORE



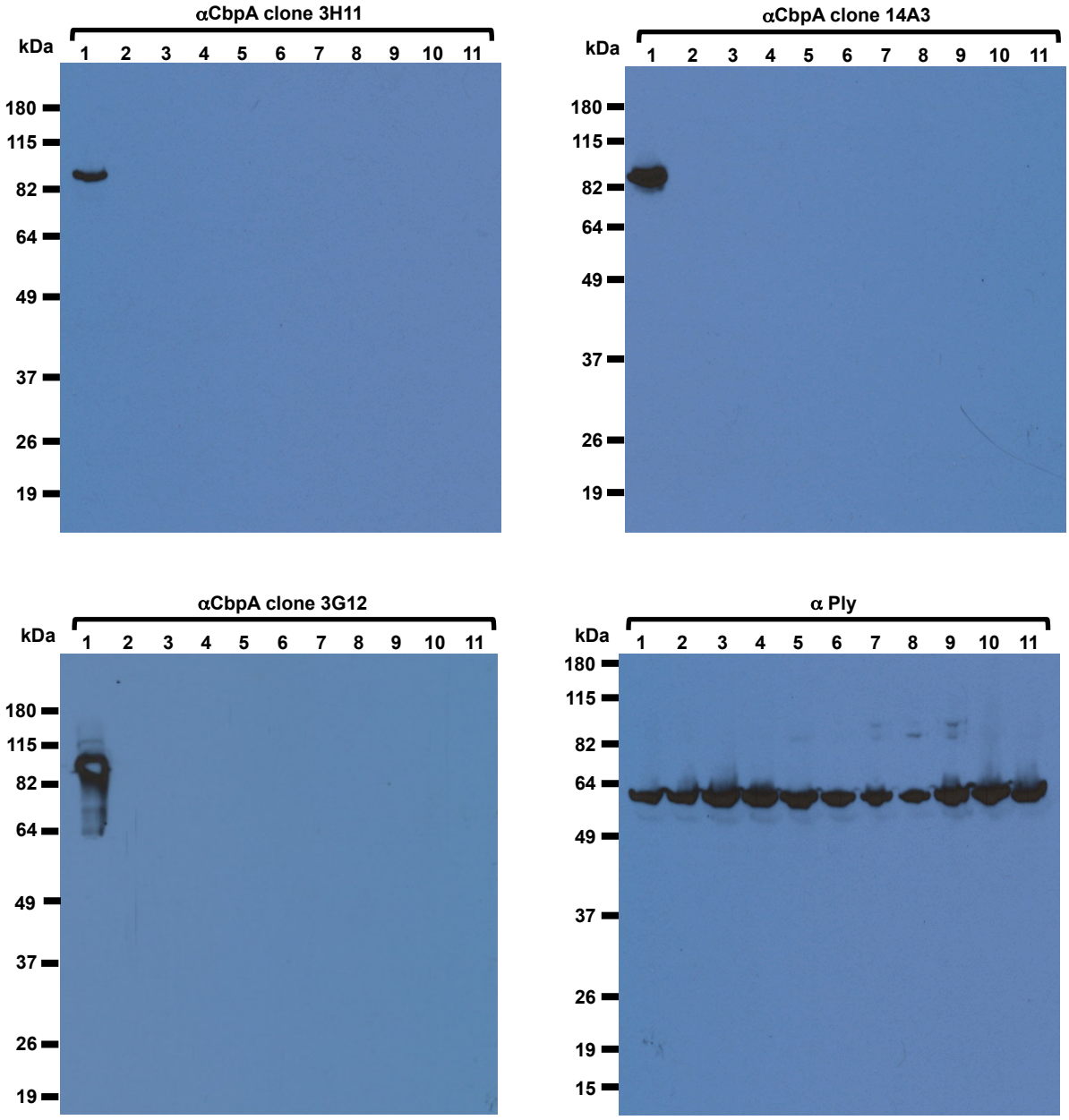
Supplementary Figure 6. Structure of macrolide resistance elements in ECC isolates. Structure of the resistance elements within: ST448 genomes ECC_0071, ECC_0083, ECC_1854, ECC_1910 (**a**) or ECC_3510 (**b**) as compared to non-MEGA containing ST448s; (**c**) ST1186 genomes compared to TIGR4 as a reference; (**d**) ST344 and ST1270 genomes; (**e**) non-ocular references strains. Regions of shared sequence are shown in pink.

Supplementary Figure-7 GILMORE



Supplementary Figure 7. MLST-based phylogenetic relationships among *S. pneumoniae* strains. Included are sequence types represented in Figure 2 (from conjunctivitis and other infections) and sequence types found in two recent surveys of nasopharyngeal colonization[3,4]. Unencapsulated strains are highlighted in pink. ECC members are denoted with a red box. A diverse set of genomes selected from two recent nasopharyngeal carriage surveys[3,4] for additional analyses are denoted with a yellow box and can be found in Supplementary Data 7.

Supplementary Figure-8 GILMORE



Supplementary Figure 8. Western blots for canonical CbpA. Western blot of total bacterial extracts, probed with monoclonal antibodies against CbpA (clones: 3H11, 14A-3, 3G12 see Mann *et al.*[2]) and pneumolysin (anti-Ply). Lanes 1 through 11 are identical to those shown in Supplementary Figure 4. Molecular weight markers are denoted.

Supplementary References

1 Menon, B. B. & Govindarajan, B. Identification of an atypical zinc metalloproteinase, ZmpC, from an epidemic conjunctivitis-causing strain of *Streptococcus pneumoniae*. *Microb Pathog* **56**, 40-46 (2013).

2 Mann, B. *et al.* Broadly Protective Protein-Based Pneumococcal Vaccine Composed of Pneumolysin Toxoid-CbpA Peptide Recombinant Fusion Protein. *J Infect Dis* (2013).

3 Chewapreecha, C. *et al.* Dense genomic sampling identifies highways of pneumococcal recombination. *Nat Genet* (2014).

4 Croucher, N. J. *et al.* Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nat Genet* **45**, 656-663 (2013).