Supporting Information

Ogura et al. 10.1073/pnas.0903585106

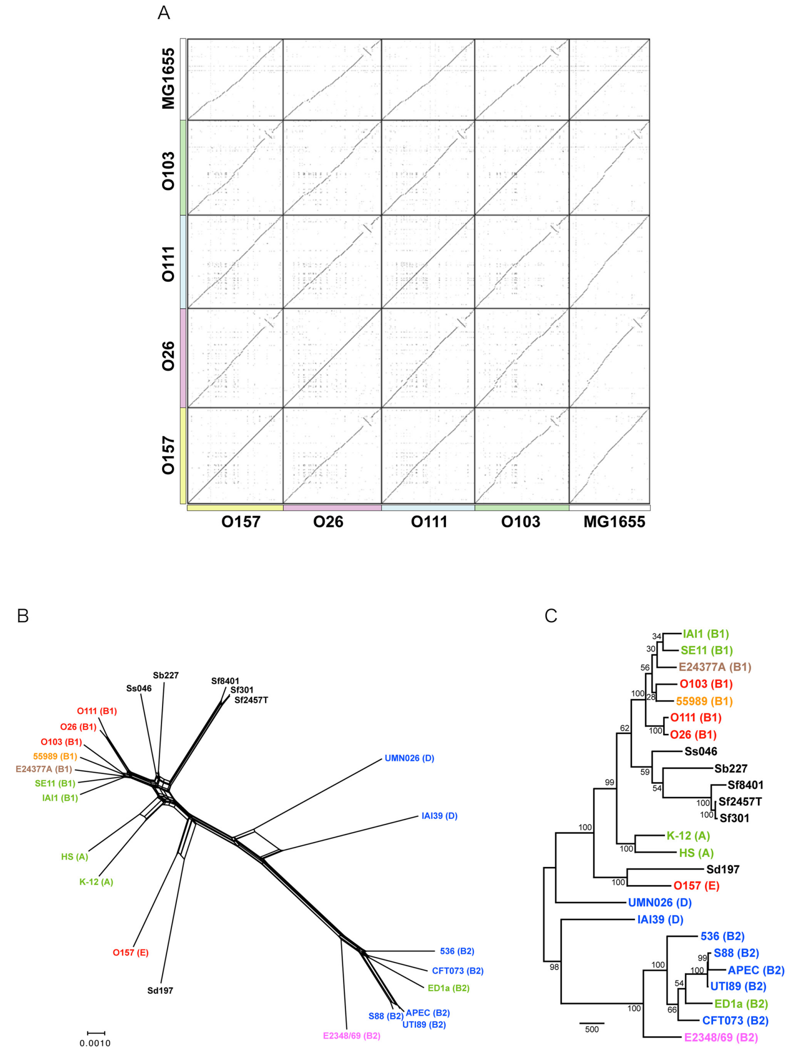

Fig. S1. Whole genome comparison of the EHEC and other *E. coli/Shigella* strains. (*A*) Dot-plot matrices of the chromosome sequences of the four EHEC strains and of K-12 MG1655. (*B*) The splits tree constructed by SplitsTree 4 by using the concatenated nucleotide sequences of the 926 orthologous CDSs from the 25 sequenced strains. The scale bar represents the number of substitutions per site. The *E. coli* phylogroup (A, B1, B2, D, or E) of each strain is indicated in brackets, and pathotypes of the strains are indicated by different colors. (*C*) The maximum parsimony tree constructed by the MEGA4 software based on the concatenated nucleotide sequences of the 345 selected orthologous CDSs. The 345 CDSs were predicted as unlikely recombinogenic genes by the PHI-test (cut off value: $P \geq 0.05$). The scale bar represents the number of substitutions. The *E. coli* phylogroup and pathotype of each strain are indicated, as in (*B*). Locus_tags of the 926 orthologous CDSs in K-12 MG1655 are listed below (the 345 selected CDSs are indicated by asterisks): b0003*, b0004, b0006, b0010*, b0011, b0014*, b0019, b0025, b0027*, b0028*, b0046*, b0050, b0051*, b0054, b0059, b0061, b0063, b0065, b0067, b0071, b0072, b0080*, b0083*, b0084, b0086, b0087, b0088, b0090*, b0091, b0092, b0093*, b0096*, b0098, b0102*, b0103*, b0104, b0110*, b0111*, b0114, b0121*, b0126*, b0128*, b0131*, b0134, b0146*, b0153, b0154, b0155, b0157*, b0159*, b0161, b0163*, b0166*, b0167, b0171*, b0177, b0178, b0181*, b0182*, b0183, b0185, b0192, b0195, b0196*, b0199*, b0208*, b0212, b0220*, b0224, b0237, b0238*, b0239, b0240*, b0243, b0379*, b0388*, b0390, b0401*, b0405, b0406*, b0413*, b0417, b0421, b0425, b0433*, b0437, b0438, b0440, b0441, b0443*, b0444, b0451, b0465, b0467, b0468*, b0472*, b0475*, b0477, b0489, b0522, b0524, b0566*, b0577, b0581, b0589, b0594, b0595, b0597*, b0599, b0605*, b0624*, b0631, b0634, b0635, b0636, b0640, b0641, b0643*, b0653, b0658*, b0659*, b0661, b0676, b0677*, b0678, b0680, b0683*, b0686, b0707, b0708, b0710, b0713, b0714, b0720, b0722*, b0724*, b0726, b0727*, b0728, b0729, b0735*, b0736*, b0738*, b0753, b0754, b0756, b0757, b0758, b0761, b0763, b0764*, b0765, b0767*, b0773*, b0775, b0776, b0778, b0779, b0782*, b0783*, b0785, b0791, b0793, b0800, b0809, b0810, b0811*, b0817*, b0819*, b0827, b0831, b0840, b0847*, b0848*, b0850*, b0852*, b0859, b0861*, b0862*, b0863, b0866*, b0867, b0870, b0871, b0876, b0877, b0881, b0885, b0886, b0887, b0892, b0893, b0898, b0910*, b0911*, b0914, b0918, b0919*, b0920, b0922, b0925, b0927*, b0928*, b0937*, b0947, b0948, b0949, b0950, b0961*, b0964*, b0970*, b0975, b0979, b0995*, b0999*, b1000*, b1034, b1035*, b1045*, b1047, b1059, b1062*, b1064*, b1066*, b1067, b1069, b1072, b1086, b1088, b1091*, b1092, b1098, b1099, b1101*, b1104*, b1106*, b1119, b1123, b1124, b1126, b1129*, b1130*, b1131, b1134, b1136, b1174*, b1175*, b1180*, b1183*, b1188, b1203*, b1205*, b1208*, b1211, b1212*

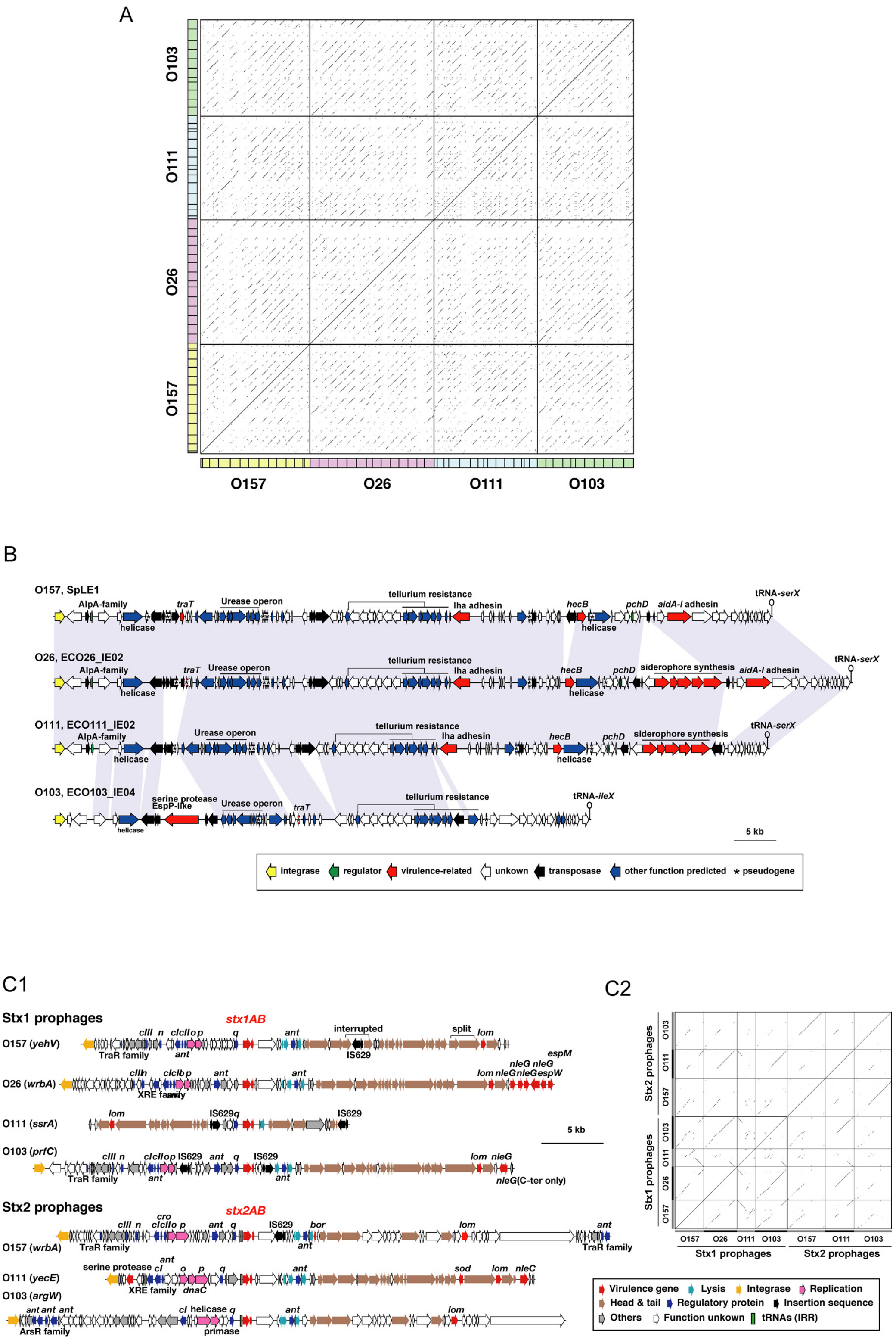

Fig. S2. Lambdoid PPs and SpLE1-like elements found in the four EHECs and their genomic comparison. (*A*) Dot-plot matrices of the concatenated sequences of lambdoid PPs found in the 4 EHECs. (*B*) The gene organization of SpLE1-like elements identified in the four EHEC genomes. Homologous regions are indicated by purple shading and integration sites are also indicated. (*C*) The gene organization (*C1*) and dot-plot matrices of the nucleotide sequences (*C2*) of the Stx1- and Stx2-transducing phages identified in the four EHECs are shown. Integration sites of each PP and IE are also indicated.

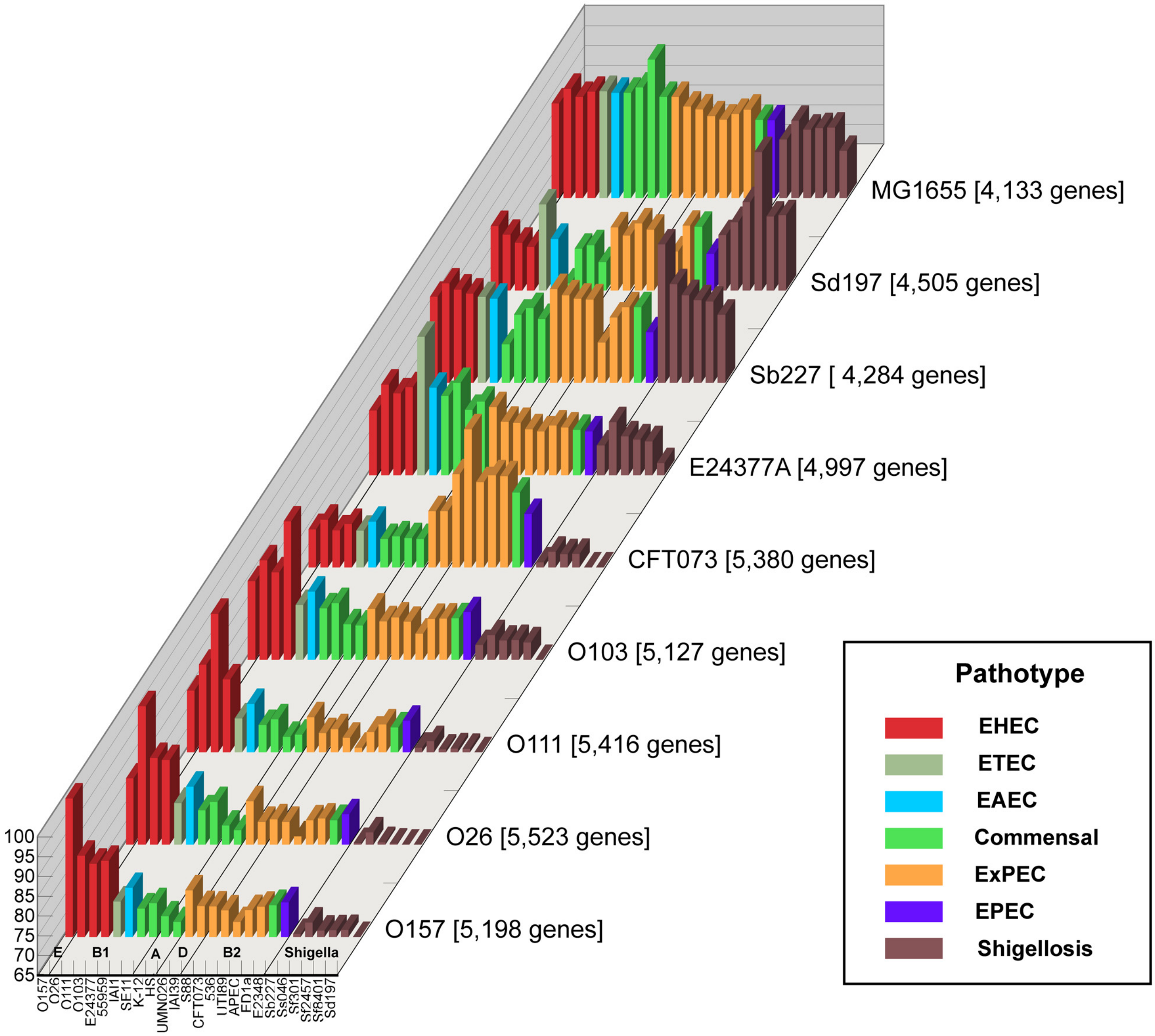

Fig. S3. Conservation of the EHEC genes in the 25 fully sequenced *E. coli/Shigella* strains. The percentage conservation of O157, O26, O111, and O103 CDSs (excluding pseudogenes) in each sequenced strain was determined by 1-way TBLASTN analysis (threshold of ≥90% amino acid sequence identity and ≥60% aligned length coverage of a query sequence). The results obtained for the 5 strains representing other pathotypes or phylogroups [CFT073 (extra-intestinal pathogenic, phylogroup B2), E24377A (enterotoxigenic, B1), K-12 MG1655 (nonpathogenic, A), Sb227 (*S. boydii*), and Sd197 (*S. dysenteriae*)] are also shown.

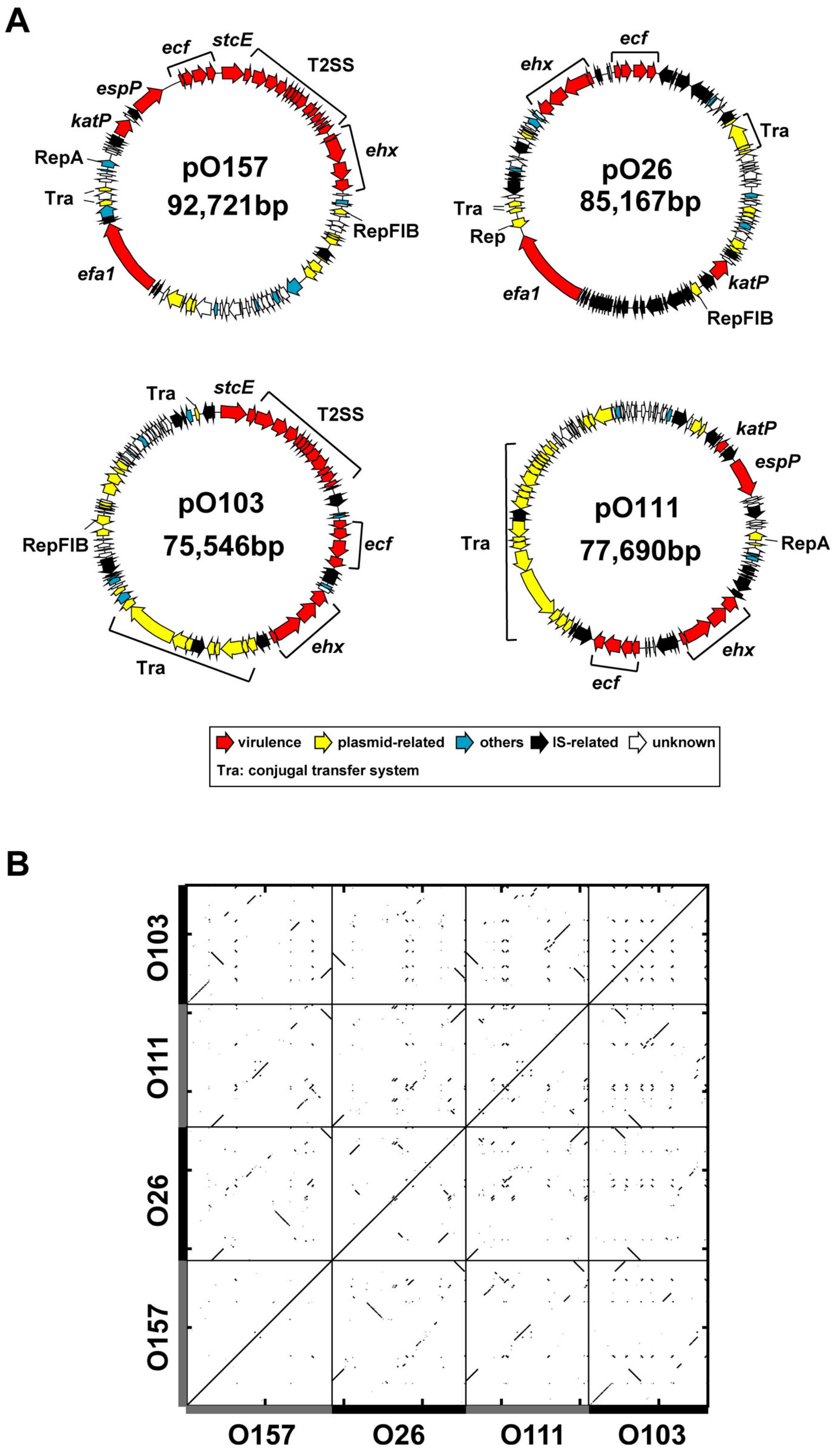

Fig. S4. Comparison of the four EHEC virulence plasmids. Circular maps (*A*) and dot-plot matrices of the nucleotide sequences (*B*) of the virulence plasmids found in the four EHECs.

Other Supporting Information Files

- Table S1
- Table S2
- Table S3
- Table S4
- Table S5
- Table S6