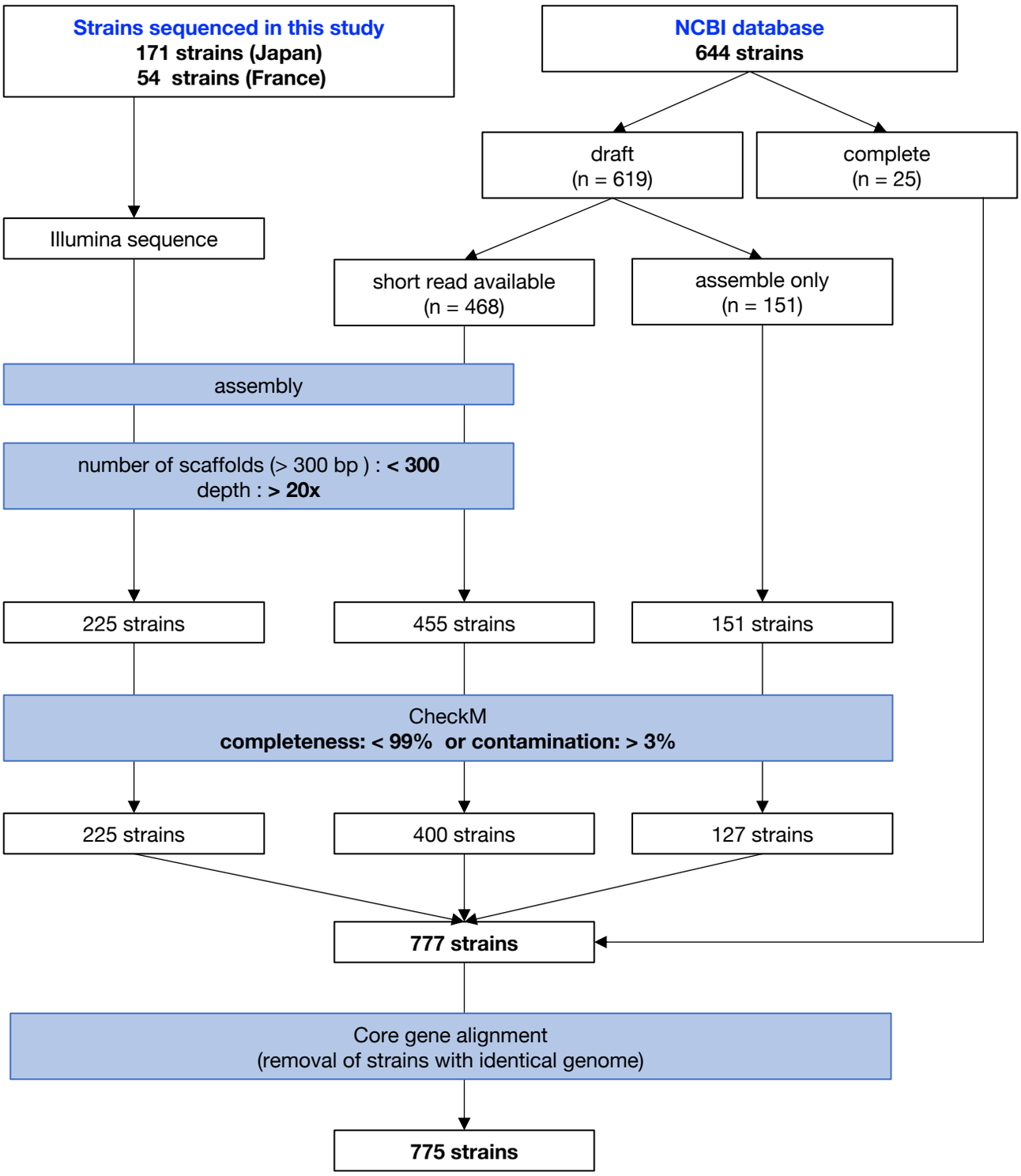

Fig. S1 Processes of genome sequencing and quality filtering to obtain the genome dataset analyzed in this study.

We sequenced 225 strains collected in Japan and France in this study. The genome sequence data of 644 strains were obtained from the NCBI database. When short read data were available, short read data were used for genome assembly using Platanus_B. Low-quality draft genomes were removed by two-step filtering. Two genomes having core gene sequences identical to the other genomes were also removed from the final set.

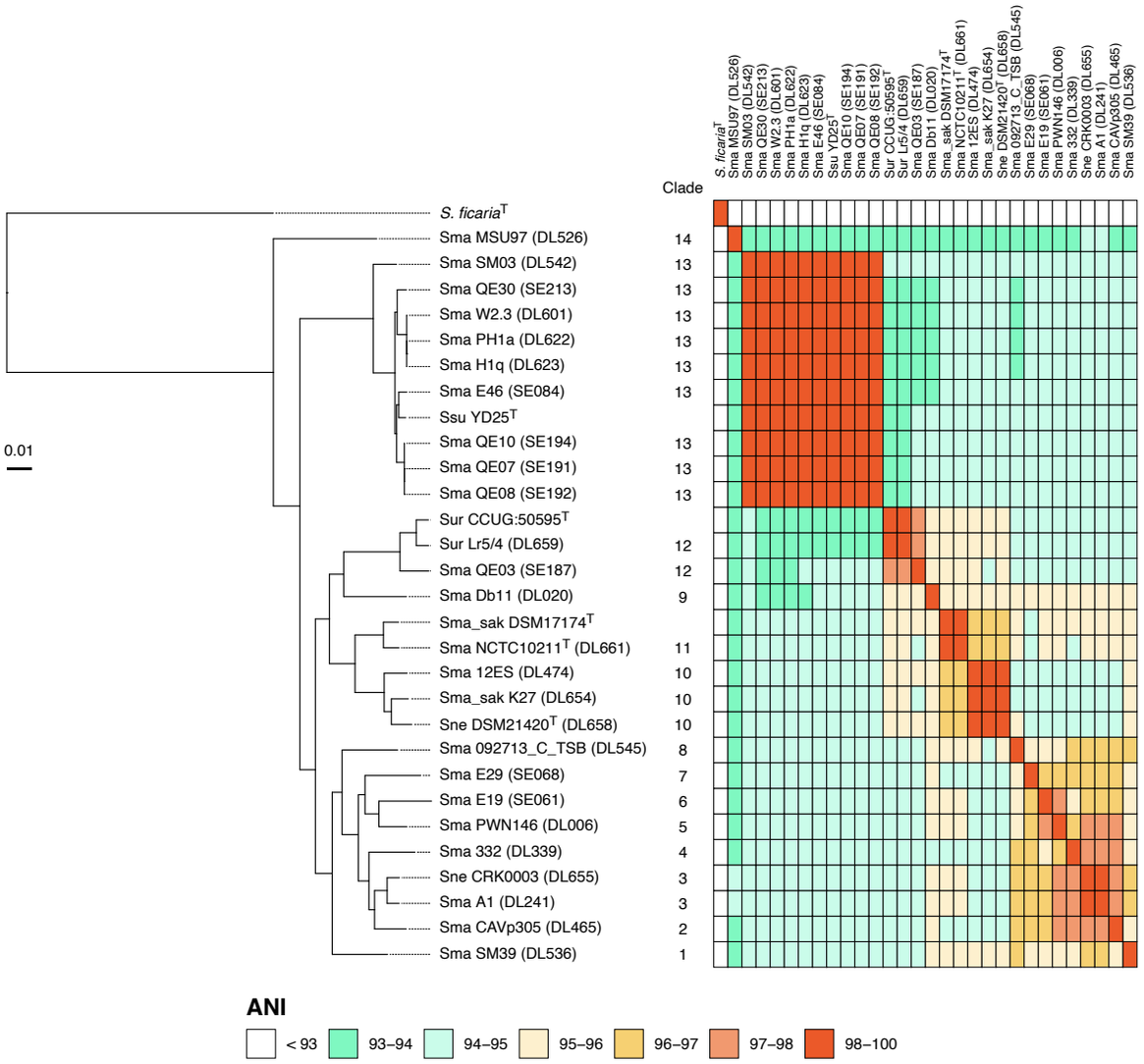

Fig. S2 Genetic relatedness of the strains belonging to clades 13 and 14 and the two strains showing apparently incorrect species names to the type strains of the Sma complex and its closely related species.

Nine clade 13 strains, one clade 14 strain, and two strains that were deposited with apparently incorrect species/subspecies names (strain IDs; DL654 and DL655) were analyzed for their genetic relatedness to the type strains of the Sma complex and its closely related species by constructing an ML phylogenetic tree and calculating ANI scores. From each of the other clades (other than clades 13 and 14), one strain was selected and included in the analyses. The ML tree was constructed based on the 561,515 SNP sites identified in the core genes, which were identified using Roary (n = 2,794). The type strain of *S. ficaria* was used as an outgroup. The ANI matrix shown on the right-hand side was obtained by all-to-all ANI analysis using PYANI in ANIb mode, and cells were coloured according to ANI values. Sma_sak; *S. marcescens* subsp. *sakuensis*, Sma; *S. marcescens*, Sne; *S. nematodiphila*, Sur; *S. ureilytica*, Ssu; *S. surfactantfaciens*. The strain IDs used in this study are shown in parentheses. Note that the type strains of Sma_sak, Sur, and Ssu were not included in our dataset described in the main text because their genome sequences were not available when we constructed the dataset. Therefore, their strain IDs are missing.

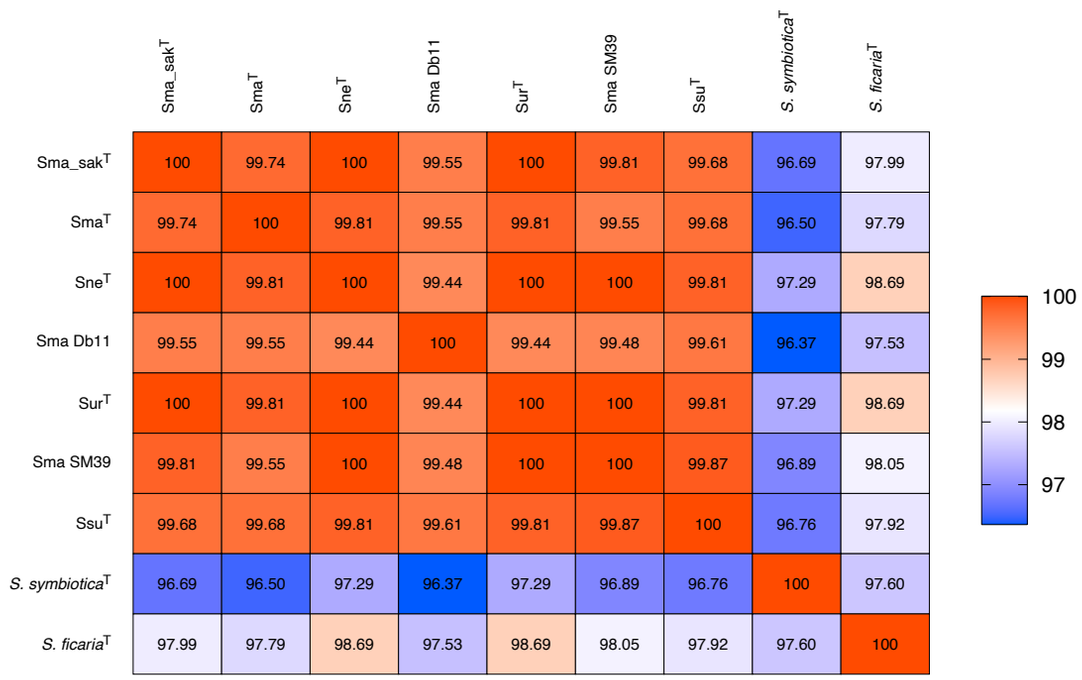

Fig. S3 16S rRNA sequence identity matrix among the Sma complex and its closely related *Serratia* species. For each strain, we generated the consensus sequences of 16S rRNA genes and aligned them with Clustal Omega to create a percent identity matrix. Cells are coloured according to identity values. Sma_sak; *S. marcescens* subsp. *sakuensis*, Sma; *S. marcescens*, Sne; *S. nematodiphila*, Sur; *S. ureilytica*, Ssu; *S. surfactantfaciens*. Strains other than Db11 and SM39 are the type strains of each species and/or subspecies.

A

Cluster size	1	2	3	4	5	6	7
Numbers of SNP_{10} clusters	515	71	22	6	3	1	1

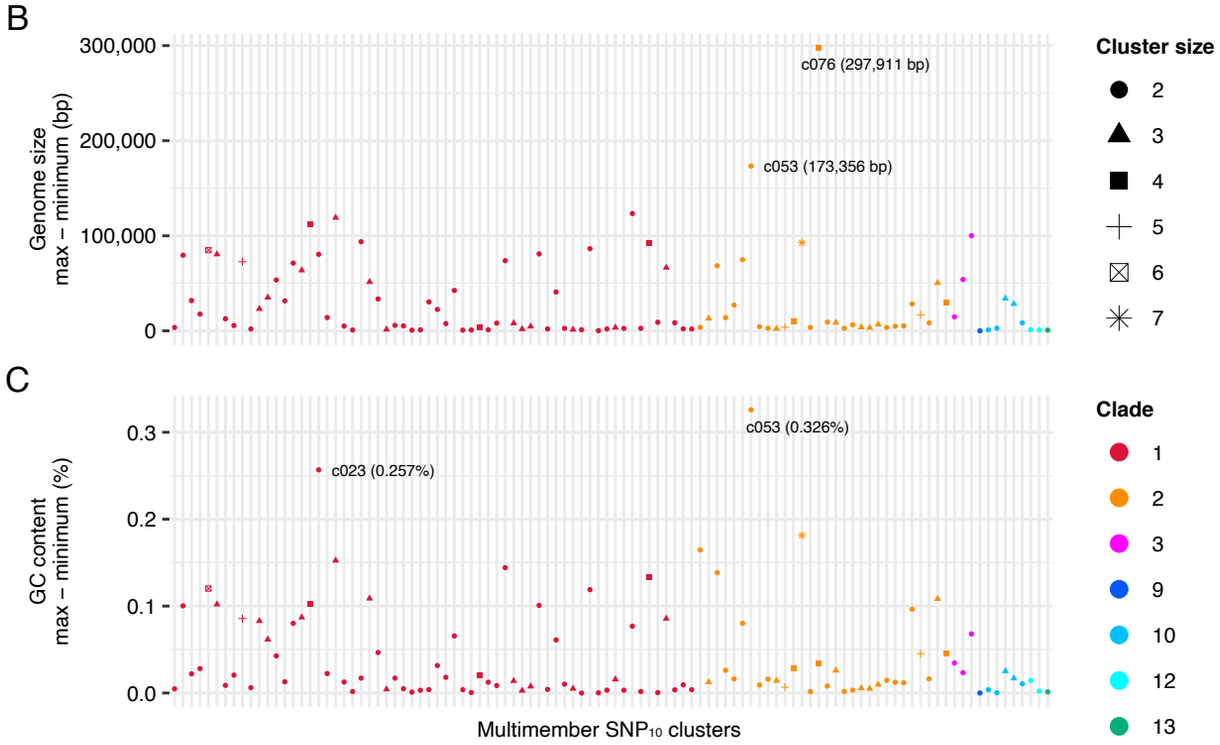

Fig. S4 Genome size and GC content variations within each SNP_{10} cluster.

A. Distribution of cluster sizes (numbers of members belonging to the same cluster).

B and C. Intracluster differences in genome size (B) and GC content (C). Each mark indicates intracluster differences (max - minimum) in each multimember cluster. Cluster sizes and clades of each cluster are indicated by different shapes and colours. The cluster IDs and intracluster differences of the SNP_{10} clusters exhibiting notable differences (genome size; >150 kb, GC content; >0.2%) are shown.

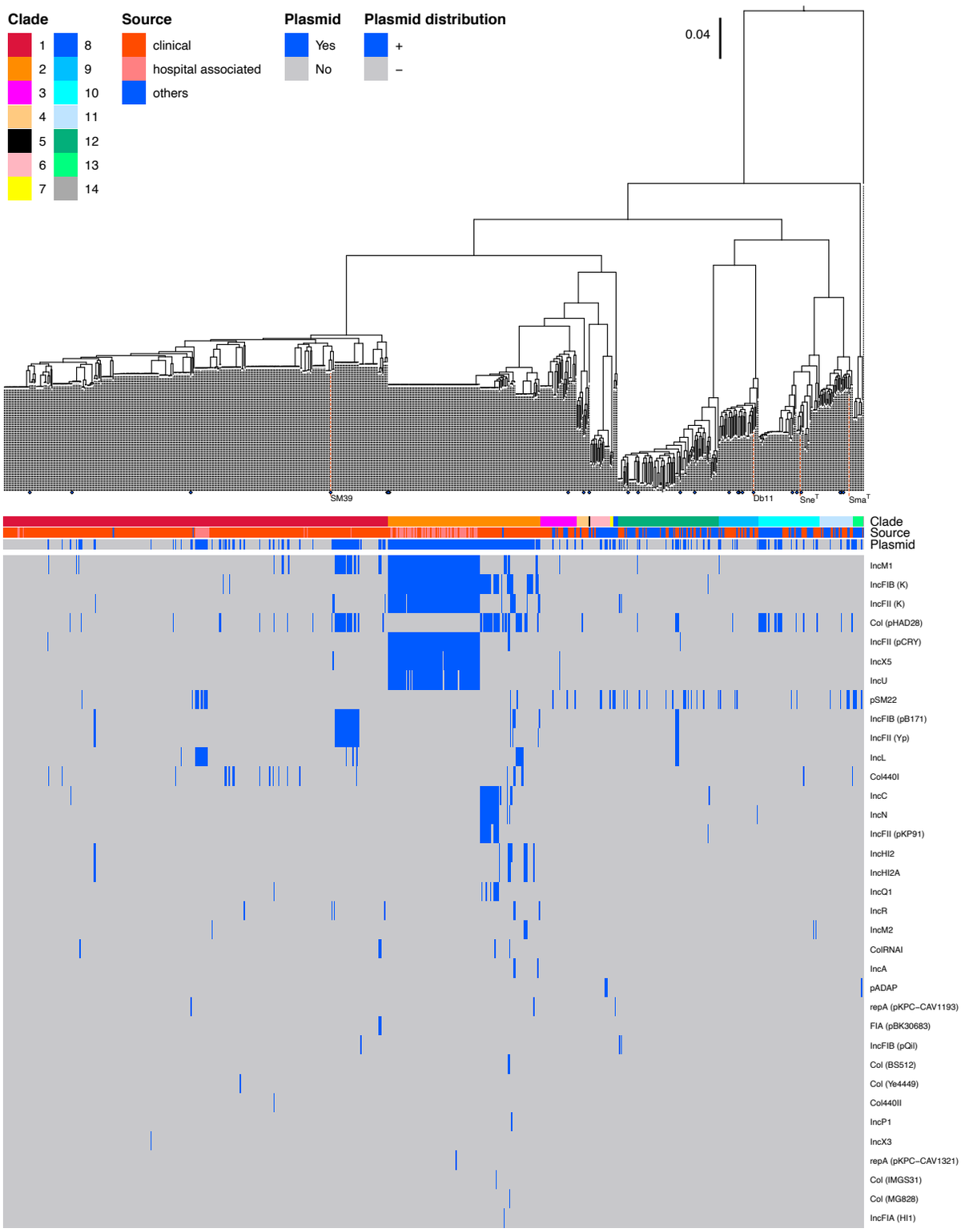

Fig. S5 Distribution of plasmid replicons in the Sma complex.

The distribution of 35 plasmid replicons identified by PlasmidFinder is shown. The same ML tree and strain information on the clade and isolation source presented in Fig. 4 are shown here.

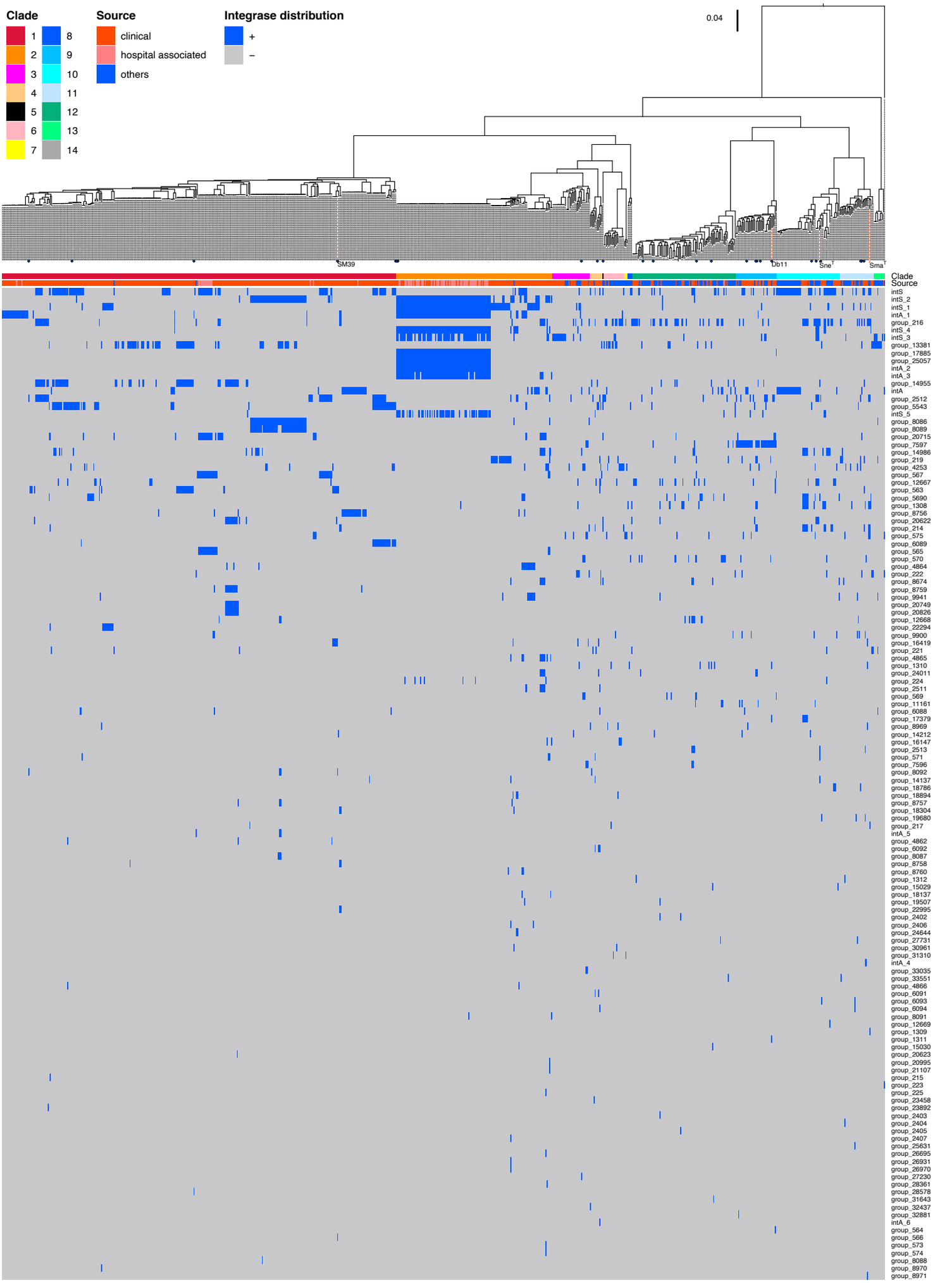
Clade
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Source
clinical
hospital associated
others
Integrase distribution
+
−
0.04
SM39
Db11
Sne
Sma
Clade
Source
intS
intS_2
intS_1
intA_1
group_216
intS_4
intS_3
group_13381
group_17885
group_25057
intA_2
intA_3
group_14955
intA
group_2512
group_5543
intS_5
group_8086
group_8089
group_20715
group_7597
group_14986
group_219
group_4253
group_567
group_12667
group_563
group_5690
group_1308
group_8756
group_20622
group_214
group_575
group_6089
group_565
group_570
group_4864
group_222
group_8674
group_8759
group_9941
group_20749
group_20826
group_12668
group_22294
group_9900
group_16419
group_221
group_4865
group_1310
group_24011
group_224
group_2511
group_569
group_11161
group_6088
group_17379
group_8969
group_14212
group_16147
group_2513
group_571
group_7596
group_8092
group_14137
group_18786
group_18894
group_8757
group_18304
group_19680
group_217
intA_5
group_4862
group_6092
group_8087
group_8758
group_8760
group_1312
group_15029
group_18137
group_19507
group_22995
group_2402
group_2406
group_24644
group_27731
group_30961
group_31310
intA_4
group_33035
group_33551
group_4866
group_6091
group_6093
group_6094
group_8091
group_12669
group_1309
group_1311
group_15030
group_20623
group_20995
group_21107
group_215
group_223
group_225
group_23458
group_23892
group_2403
group_2404
group_2405
group_2407
group_25631
group_26695
group_26931
group_26970
group_27230
group_28361
group_28578
group_31643
group_32437
group_32881
intA_6
group_564
group_566
group_573
group_574
group_8088
group_8970
group_8971

Fig. S6 Distribution of integrase genes in the Sma complex.

The distribution of 130 integrase genes (gene groups) identified in the pangenome data of the Sma complex is shown with the gene group names in the pangenome data generated by Roary. The same ML tree and strain information on the clade and isolation source presented in Fig. 4 are shown here.

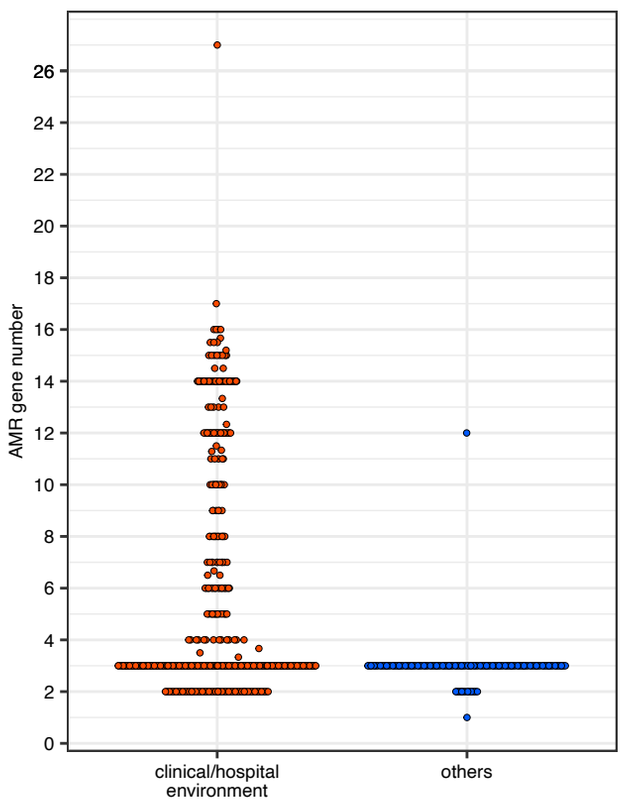

Fig. S7 Comparison of the numbers of AMR genes between the SNP_{10} clusters from clinical/hospital environment sources and those from the other sources.

The numbers of AMR genes in the multimember SNP_{10} clusters were calculated as the average number of the strains belonging to the same SNP_{10} cluster.

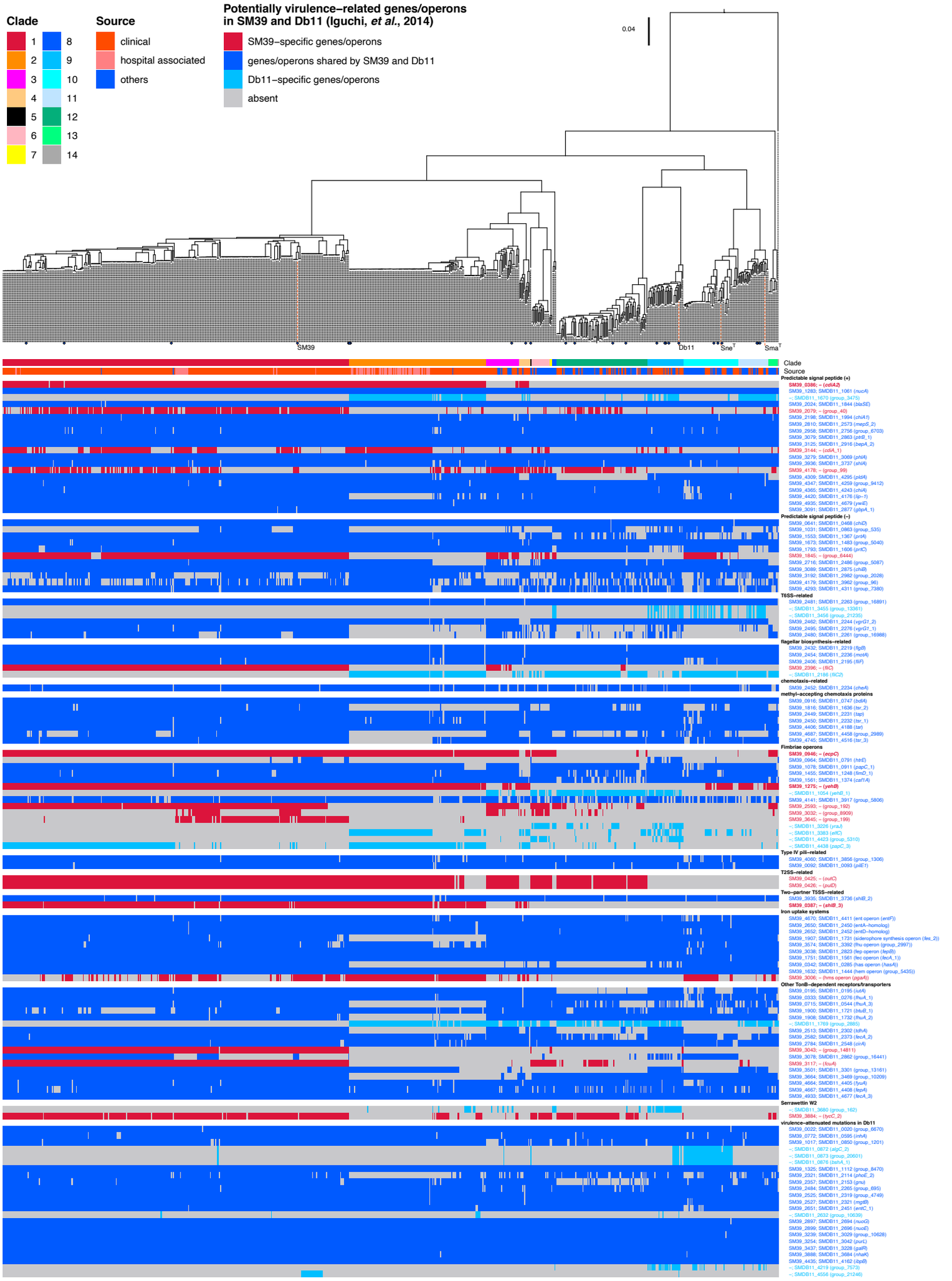
Clade
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Source
clinical
hospital associated
others
Potentially virulence–related genes/operons in SM39 and Db11 (Iguchi, et al., 2014)
SM39–specific genes/operons
genes/operons shared by SM39 and Db11
Db11–specific genes/operons
absent
0.04
SM39
Db11
SneT
SmaT
Clade
Source
Predictable signal peptide (+)
SM39_0386; – (cdiA2)
SM39_1283; SMDB11_1061 (nucA)
–; SMDB11_1670 (group_3475)
SM39_2024; SMDB11_1844 (blaSE)
SM39_2079; – (group_40)
SM39_2198; SMDB11_1994 (chiA1)
SM39_2810; SMDB11_2573 (mepS_2)
SM39_2958; SMDB11_2756 (group_6703)
SM39_3079; SMDB11_2863 (pbtB_1)
SM39_3125; SMDB11_2916 (bepA_2)
SM39_3144; – (cdiA_1)
SM39_3279; SMDB11_3069 (phlA)
SM39_3936; SMDB11_3737 (shlA)
SM39_4178; – (group_99)
SM39_4309; SMDB11_4295 (plsA)
SM39_4347; SMDB11_4259 (group_9412)
SM39_4365; SMDB11_4243 (chiA)
SM39_4420; SMDB11_4176 (ilp–1)
SM39_4935; SMDB11_4679 (yenE)
SM39_3091; SMDB11_2877 (bepA_1)
Predictable signal peptide (–)
SM39_0641; SMDB11_0468 (chiD)
SM39_1031; SMDB11_0863 (group_535)
SM39_1553; SMDB11_1367 (prtA)
SM39_1673; SMDB11_1483 (group_5040)
SM39_1793; SMDB11_1606 (prtC)
SM39_1845; – (group_6444)
SM39_2716; SMDB11_2486 (group_5087)
SM39_3089; SMDB11_2875 (chiB)
SM39_3192; SMDB11_2982 (group_2028)
SM39_4179; SMDB11_3962 (group_96)
SM39_4293; SMDB11_4311 (group_7380)
T6SS–related
SM39_2481; SMDB11_2263 (group_16891)
–; SMDB11_3455 (group_13381)
–; SMDB11_3456 (group_21235)
SM39_2462; SMDB11_2244 (vgrG1_2)
SM39_2495; SMDB11_2276 (vgrG1_1)
SM39_2480; SMDB11_2261 (group_16988)
flagellar biosynthesis–related
SM39_2432; SMDB11_2219 (flgB)
SM39_2454; SMDB11_2236 (motA)
SM39_2406; SMDB11_2195 (fliF)
SM39_2396; – (fliC)
–; SMDB11_2186 (fliC2)
chemotaxis–related
SM39_2452; SMDB11_2234 (cheA)
methyl–accepting chemotaxis proteins
SM39_0916; SMDB11_0747 (bdlA)
SM39_1816; SMDB11_1636 (tsr_2)
SM39_2449; SMDB11_2231 (tap)
SM39_2450; SMDB11_2232 (tsr_1)
SM39_4406; SMDB11_4188 (tar)
SM39_4687; SMDB11_4458 (group_2989)
SM39_4745; SMDB11_4516 (tsr_3)
Fimbriae operons
SM39_0946; – (ecpC)
SM39_0964; SMDB11_0791 (htrE)
SM39_1078; SMDB11_0911 (papC_1)
SM39_1455; SMDB11_1248 (fimD_1)
SM39_1561; SMDB11_1374 (caf1A)
SM39_1275; – (yehB)
–; SMDB11_1054 (ycfE_1)
SM39_4141; SMDB11_3917 (group_5806)
SM39_2593; – (group_192)
SM39_3032; – (group_8909)
SM39_3645; – (group_199)
–; SMDB11_3226 (yraJ)
–; SMDB11_3383 (sfmC)
–; SMDB11_4423 (group_5310)
–; SMDB11_4438 (papC_3)
Type IV pili–related
SM39_4060; SMDB11_3856 (group_1306)
SM39_0092; SMDB11_0093 (pilE1)
T2SS–related
SM39_0425; – (outC)
SM39_0426; – (pulD)
Two–partner TSSS–related
SM39_3935; SMDB11_3736 (shlB_2)
SM39_0387; – (shlB_3)
Iron uptake systems
SM39_4670; SMDB11_4411 (ent operon (entF))
SM39_2650; SMDB11_2450 (entA–homolog)
SM39_2652; SMDB11_2452 (entD–homolog)
SM39_1907; SMDB11_1731 (siderophore synthesis operon (fes_2))
SM39_3574; SMDB11_3392 (fhu operon (group_2997))
SM39_3038; SMDB11_2823 (fep operon (fepB))
SM39_1751; SMDB11_1561 (fec operon (fecA_1))
SM39_0342; SMDB11_0285 (has operon (hasA))
SM39_1632; SMDB11_1444 (hem operon (group_5435))
SM39_3006; – (hma operon (zgpA))
Other TonB–dependent receptors/transporters
SM39_0195; SMDB11_0195 (iutA)
SM39_0333; SMDB11_0276 (fhuA_1)
SM39_0715; SMDB11_0544 (fhuA_3)
SM39_1900; SMDB11_1721 (btuB_1)
SM39_1908; SMDB11_1732 (fhuA_2)
–; SMDB11_1769 (group_2885)
SM39_2513; SMDB11_2302 (tbfA)
SM39_2582; SMDB11_2373 (fecA_2)
SM39_2784; SMDB11_2548 (cirA)
SM39_3043; – (group_14811)
SM39_3078; SMDB11_2862 (group_16441)
SM39_3117; – (fcuA)
SM39_3501; SMDB11_3301 (group_13161)
SM39_3664; SMDB11_3469 (group_10209)
SM39_4664; SMDB11_4405 (fyuA)
SM39_4667; SMDB11_4408 (fepA)
SM39_4933; SMDB11_4677 (fecA_3)
Serrawettin W2
–; SMDB11_3680 (group_162)
SM39_3884; – (tycC_2)
virulence–attenuated mutations in Db11
SM39_0022; SMDB11_0020 (group_6670)
SM39_0772; SMDB11_0595 (ishA)
SM39_1017; SMDB11_0850 (group_1201)
–; SMDB11_0872 (algC_2)
–; SMDB11_0873 (group_20601)
–; SMDB11_0876 (bshA_1)
SM39_1325; SMDB11_1112 (group_8470)
SM39_2321; SMDB11_2114 (phoE_2)
SM39_2357; SMDB11_2153 (gnu)
SM39_2484; SMDB11_2265 (group_695)
SM39_2525; SMDB11_2319 (group_4749)
SM39_2527; SMDB11_2321 (mglB)
SM39_2651; SMDB11_2451 (entC_1)
–; SMDB11_2632 (group_10639)
SM39_2897; SMDB11_2694 (nuoG)
SM39_2899; SMDB11_2696 (nuoE)
SM39_3239; SMDB11_3029 (group_10628)
SM39_3254; SMDB11_3042 (purL)
SM39_3437; SMDB11_3228 (galR)
SM39_3888; SMDB11_3684 (nhaK)
SM39_4435; SMDB11_4162 (dpiB)
–; SMDB11_4219 (group_7573)
–; SMDB11_4556 (group_21246)

Fig. S8 Distribution of potentially virulence-related genes/operons described for the SM39 and/or Db11 genomes.

The distribution of potentially virulence-related genes/operons that were described for the SM39 and/or Db11 genomes by Iguchi *et al.* [Ref. 15 in the main text] is shown. The same ML tree and strain information on the clade and isolation source presented in Fig. 4 are shown here. As for operons (or clusters of genes), representative genes indicated in each operon were analyzed. Gene IDs in the SM39 and Db11 genomes are also indicated with their gene group names in the pangenome data in parentheses. The genes that were present in only SM39, both SM39 and Db11, and only Db11 are indicated by red, blue, and light blue, respectively. The four genes that showed a biased distribution to clades 1 and 2 are indicated in bold.

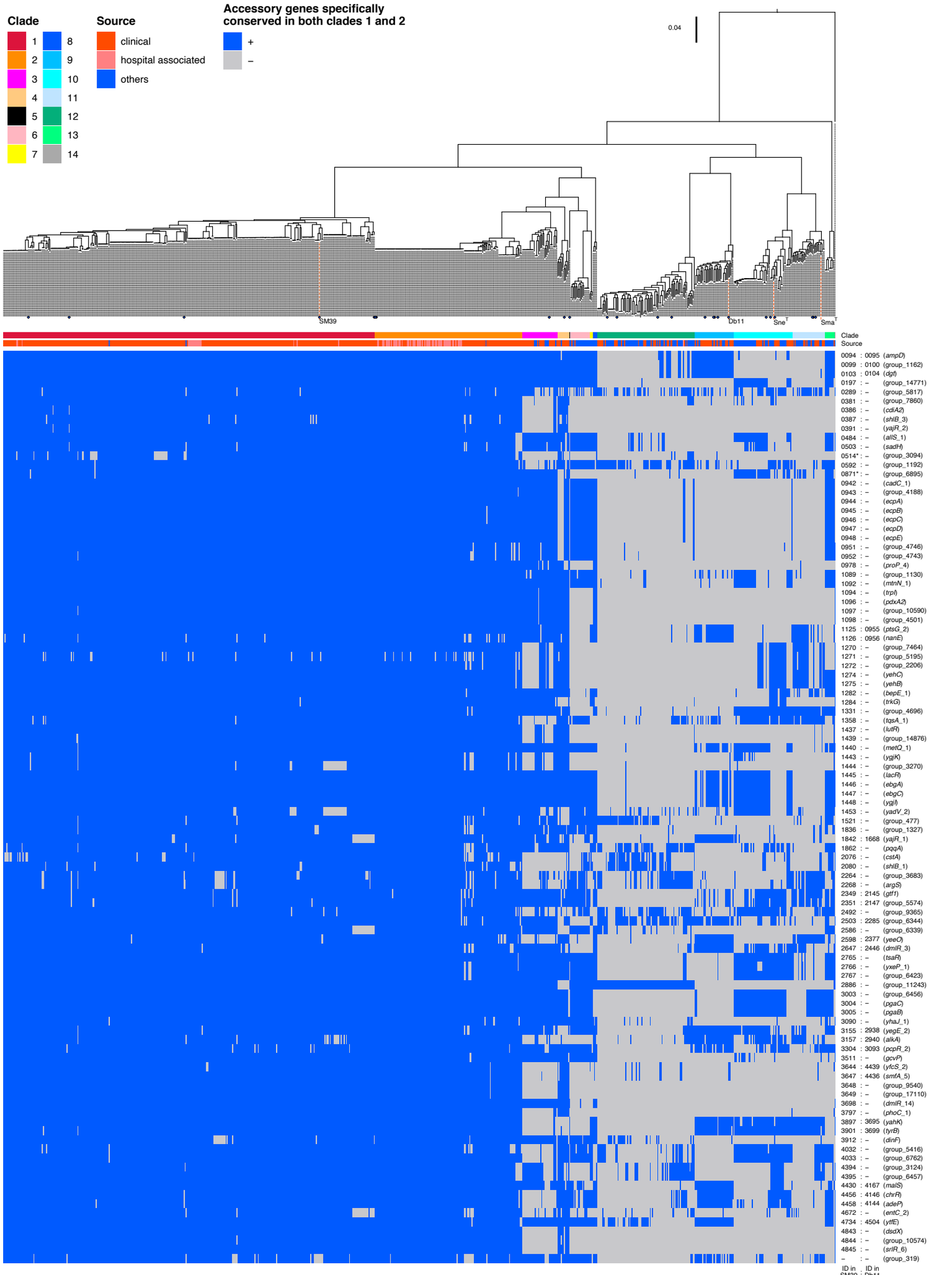
Clade
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Source
clinical
hospital associated
others
Accessory genes specifically conserved in both clades 1 and 2
+
–
0.04
SM39
Db11
Sne T
Sma T
Clade
Source
0094 : 0095 (ampD)
0099 : 0100 (group_1162)
0103 : 0104 (dgt)
0197 : – (group_14771)
0289 : – (group_5817)
0381 : – (group_7860)
0386 : – (cdiA2)
0387 : – (shlB_3)
0391 : – (yajR_2)
0484 : – (allS_1)
0503 : – (sadH)
0514* : – (group_3094)
0592 : – (group_1192)
0871* : – (group_6895)
0942 : – (cadC_1)
0943 : – (group_4188)
0944 : – (ecpA)
0945 : – (ecpB)
0946 : – (ecpC)
0947 : – (ecpD)
0948 : – (ecpE)
0951 : – (group_4746)
0952 : – (group_4743)
0978 : – (proP_4)
1089 : – (group_1130)
1092 : – (mtnN_1)
1094 : – (trpI)
1096 : – (pdxA2)
1097 : – (group_10590)
1098 : – (group_4501)
1125 : 0955 (ptsG_2)
1126 : 0956 (nanE)
1270 : – (group_7464)
1271 : – (group_5195)
1272 : – (group_2206)
1274 : – (yehC)
1275 : – (yehB)
1282 : – (bepE_1)
1284 : – (trkG)
1331 : – (group_4696)
1358 : – (tqsA_1)
1437 : – (lutR)
1439 : – (group_14876)
1440 : – (metQ_1)
1443 : – (ygjK)
1444 : – (group_3270)
1445 : – (lacR)
1446 : – (ebgA)
1447 : – (ebgC)
1448 : – (ygjI)
1453 : – (yadV_2)
1521 : – (group_477)
1836 : – (group_1327)
1842 : 1668 (yajR_1)
1862 : – (pqqA)
2076 : – (cstA)
2080 : – (shlB_1)
2264 : – (group_3683)
2268 : – (argS)
2349 : 2145 (gtf1)
2351 : 2147 (group_5574)
2492 : – (group_9365)
2503 : 2285 (group_6344)
2586 : – (group_6339)
2598 : 2377 (yeeO)
2647 : 2446 (dmlR_3)
2765 : – (tsaR)
2766 : – (yxeP_1)
2767 : – (group_6423)
2886 : – (group_11243)
3003 : – (group_6456)
3004 : – (pgaC)
3005 : – (pgaB)
3090 : – (yhaJ_1)
3155 : 2938 (yegE_2)
3157 : 2940 (alkA)
3304 : 3093 (pcpR_2)
3511 : – (gcvP)
3644 : 4439 (yfcS_2)
3647 : 4436 (smfA_5)
3648 : – (group_9540)
3649 : – (group_17110)
3698 : – (dmlR_14)
3797 : – (phoC_1)
3897 : 3695 (yahK)
3901 : 3699 (tyrB)
3912 : – (dinF)
4032 : – (group_5416)
4033 : – (group_6762)
4394 : – (group_3124)
4395 : – (group_6457)
4430 : 4167 (malS)
4456 : 4146 (chrR)
4458 : 4144 (adeP)
4672 : – (entC_2)
4734 : 4504 (ytfE)
4843 : – (dsdX)
4844 : – (group_10574)
4845 : – (srlR_6)
– : – (group_319)
ID in SM39 : ID in Db11

Fig. S9 Distribution of the accessary genes specifically conserved in clades 1 and 2.

Among the 287 accessary genes that are highly conserved in both clades 1 and 2 (present in >90% SNP_{10} clusters) but significantly less frequently present in the other clades (Bonferroni $p < 0.01$), the top 100 genes (Bonferroni $p < 9.99 \times 10^{-44}$) are shown according to the order in the SM39 genome. The same ML tree and strain information on the clade and isolation source presented in Fig. 4 are shown here. Gene IDs in SM39 and Db11 and their gene group names in the pangenome data are indicated. The list of the 287 genes is available in Table S11. The two gene IDs indicated by asterisks are missing numbers in the genome annotation data of SM39.