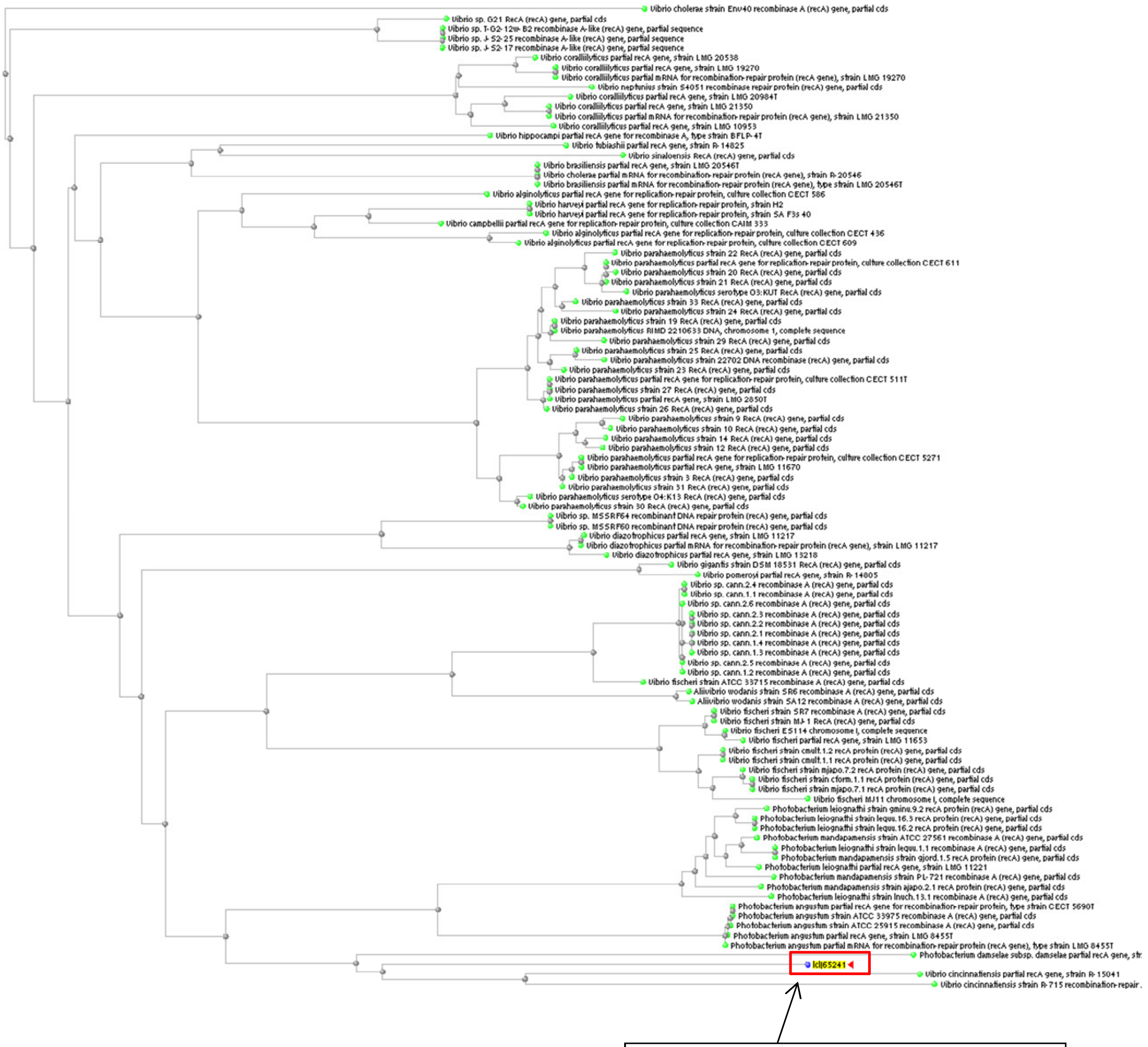


Query ID lcl|65241 represents the nucleotide sequence of allele *recA107* and shows that this allele does not share identical *recA* sequence to those of the other *Vibrio* species from NCBI database.

FIG S1 Neighbour-Joining tree of *recA* nucleotide sequences of sixteen *Vibrios* species and related species such as *Photobacterium* species from NCBI database. Query ID lcl|65241 (representing the nucleotide sequence of allele *recA107*) represents a distinct lineage and presumably originates from a previously unidentified species.

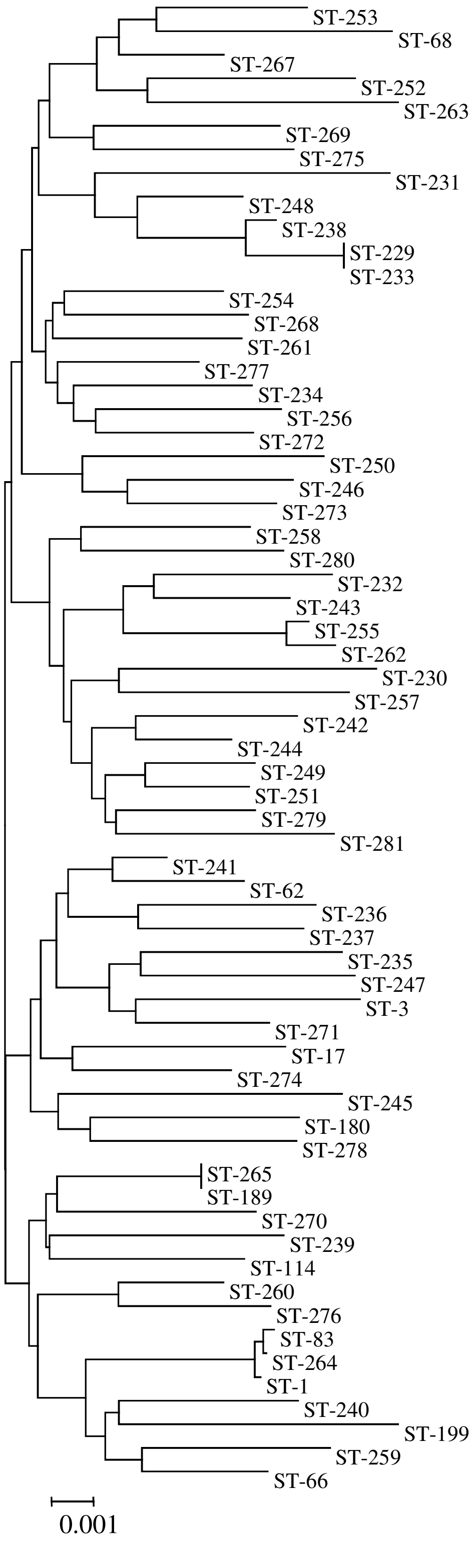
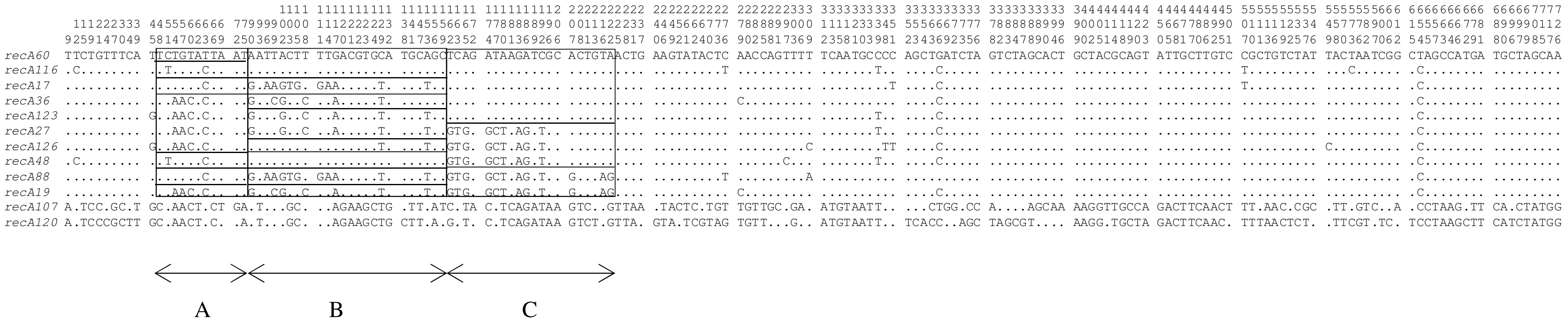


FIG S2 Neighbour-Joining tree of 63 STs representing the concatenated sequences of six housekeeping genes (with *recA* removed) of 102 *V. parahaemolyticus* isolates from multiple sources in Thailand. The branching pattern of this tree does not consist of the two main lineages A and B (as well as clades I and II) that are clearly differentiated in the phylogenetic tree of the concatenated sequences of all seven housekeeping genes (Fig. 3).

FIG S3

(A)



```
                                       1111 1111111111 1111111111 1111111112 2222222222 2222222222 2222222333 3333333333 3333333333 3333333333 3444444444 4444444445 5555555555 5555555666 6666666666 666667777
          111222333 4455566666 7799990000 1112222223 3445556667 7788888990 0011122233 4445666777 7888899000 1111233345 5556667777 7888888999 9000111122 5667788990 0111112334 4577789001 1555666778 899990112
         9259147049 5814702369 2503692358 1470123492 8173692352 4701369266 7813625817 0692124036 9025817369 2358103981 2343692356 8234789046 9025148903 0581706251 7013692576 9803627062 5457346291 806798576
recA60   TTCTGTTTCA TTCTGTATTA ATAATTACTT TTGACGTGCA TGCAGCTCAG ATAAGATCGC ACTGTAACTG AAGTATACTC AACCAGTTTT TCAATGCCCC AGCTGATCTA GTCTAGCACT GCTACGCAGT ATTGCTTGTC CGCTGTCTAT TACTAATCGG CTAGCCATGA TGCTAGCAA
recA116  .C........ ..T....C.. .......... .......... .......... .......... .......... .........T .......... .......T.. ....C..... .......... .......... .......... T......... ...C...... .C........ .........
recA17   .......... .......C.. ..G.AAGTG. GAA.....T. ...T...... .......... .......... .......... .......... .........T ....C..... .......... .......... .......... T......... .......... .C........ .........
recA36   .......... ...AAC.C.. ..G..CG..C ..A.....T. .......... .......... .......... .......... C......... .......... ....C..... .......... .......... .......... .......... .......... .C........ .........
recA123  .......... G..AAC.C.. ..G...G..C ..A.....T. ...T...... .......... .......... .......... .......... .......T.. ....C..... .......... .......... .......... .......... .......... .C........ .........
recA27   .......... ...AAC.C.. ..G...G..C ..A.....T. ...T..GTG. GCT.AG.T.. .......... .......... .......... .......T.. ....C..... .......... .......... .......... .......... .......... .C........ .........
recA126  .......... G..AAC.C.. .......... ........T. ...T..GTG. GCT.AG.T.. .......... .......... .........C ........TT ....C..... .......... .......... .......... .......... C......... .C........ .........
recA48   .C........ ..T....C.. .......... .......... ......GTG. GCT.AG.T.. .......... .......... ......C... .......T.. ....C..... .......... .......... .......... .......... .......... .C........ .........
recA88   .......... .......C.. ..G.AAGTG. GAA.....T. ...T..GTG. GCT.AG.T.. G...AG.... .........T .........A .......... .......... .......... .......... .......... .......... .......... .C........ .........
recA19   .......... ...AAC.C.. ..G..CG..C ..A.....T. ...T..GTG. GCT.AG.T.. G...AG.... .......... C......... .......... ....C..... .......... .......... .......... .......... .......... .C........ .........
recA107  A.TCC.GC.T GC.AACT.CT GA.T...GC. ..AGAAGCTG .TT.ATC.TA C.TCAGATAA GTC..GTTAA .TACTC.TGT TGTTGC.GA. ATGTAATT.. ...CTGG.CC A....AGCAA AAGGTTGCCA GACTTCAACT TT.AAC.CGC .TT.GTC..A .CCTAAG.TT CA.CTATGG
recA120  A.TCCCGCTT GC.AACT.C. .A.T...GC. ..AGAAGCTG CTT.A.G.T. C.TCAGATAA GTCT.GTTA. GTA.TCGTAG TGTT...G.. ATGTAATT.. TCACC..AGC TAGCGT.... AAGG.TGCTA GACTTCAAC. TTTAACTCT. .TTCGT.TC. TCCTAAGCTT CATCTATGG
```

(B)

Alleles	Recombinant segments			Mosaic designations	Isolates	ST	Sources
recA60	🟦 A1	🟥 B1	🟨 C1	A1,B1,C1	VP222, VP224	276	seafood
recA116	🟨 A2	🟥 B1	🟨 C1	A2,B1,C1	VP6	230	frozen shrimp
recA17	🟩 A3	🟦 B2	🟨 C1	A3,B2,C1	VP136	199	human carriage
recA36	🟥 A4	🟩 B3*	🟨 C1	A4,B3,C1	VP154	68	human carriage
recA123	🟥 A4	🟨 B4	🟨 C1	A4,B4,C1	VP138, VP162, VP170, VP178, VP180, VP182, VP184, VP188	255, 262	clinical, human carriage
recA27	🟥 A4	🟨 B4	🟩 C2	A4,B4,C2	VP24, VP176, VP166, VP172, VP194	235, 364, 83	clinical, frozen shrimp
recA126	🟥 A4	🟪 B5	🟩 C2	A4,B5,C2	VP156	260	human carriage
recA48	🟨 A2	🟥 B1	🟩 C2	A2,B1,C2	VP174	263	clinical
recA88	🟩 A3	🟦 B2	🟥 C3	A3,B2,C3	VP140	62	human carriage
recA19	🟥 A4	🟩 B3	🟥 C3	A4,B3,C3	VP132, VP158, VP164	3, 17	clinical, human carriage

FIG S3 (A) Distribution of polymorphic nucleotide sites among a sample of *recA* alleles. The numbers written vertically above the sequences represent the positions of polymorphic nucleotide sites. The dots represent sites where the nucleotides match those of the top sequence (*recA*60). The boxes indicate regions of sequence identity that represent proposed recombinant segments. (B) Schematic representation of recombinant fragments A, B and C among *recA* alleles corresponding to the nucleotide sequences shown in Fig. S3A. Different nucleotide sequences in recombinant segments A, B, and C are represented by A1-4, B1-5, and C1-3, respectively. Segments B3 and B3* differ at only a single nucleotide site. Mosaic designations are represented by different combination of A, B and C. Mosaic alleles have been formed by from one to three separate intragenic recombination events. The figure shows the *recA60, recA116, recA17, recA36, recA123, recA27, recA126, recA48, recA88,* and *recA19* alleles have complex combinations of the internal segments A, B, and C that are indicative of multiple intragenic recombination events.

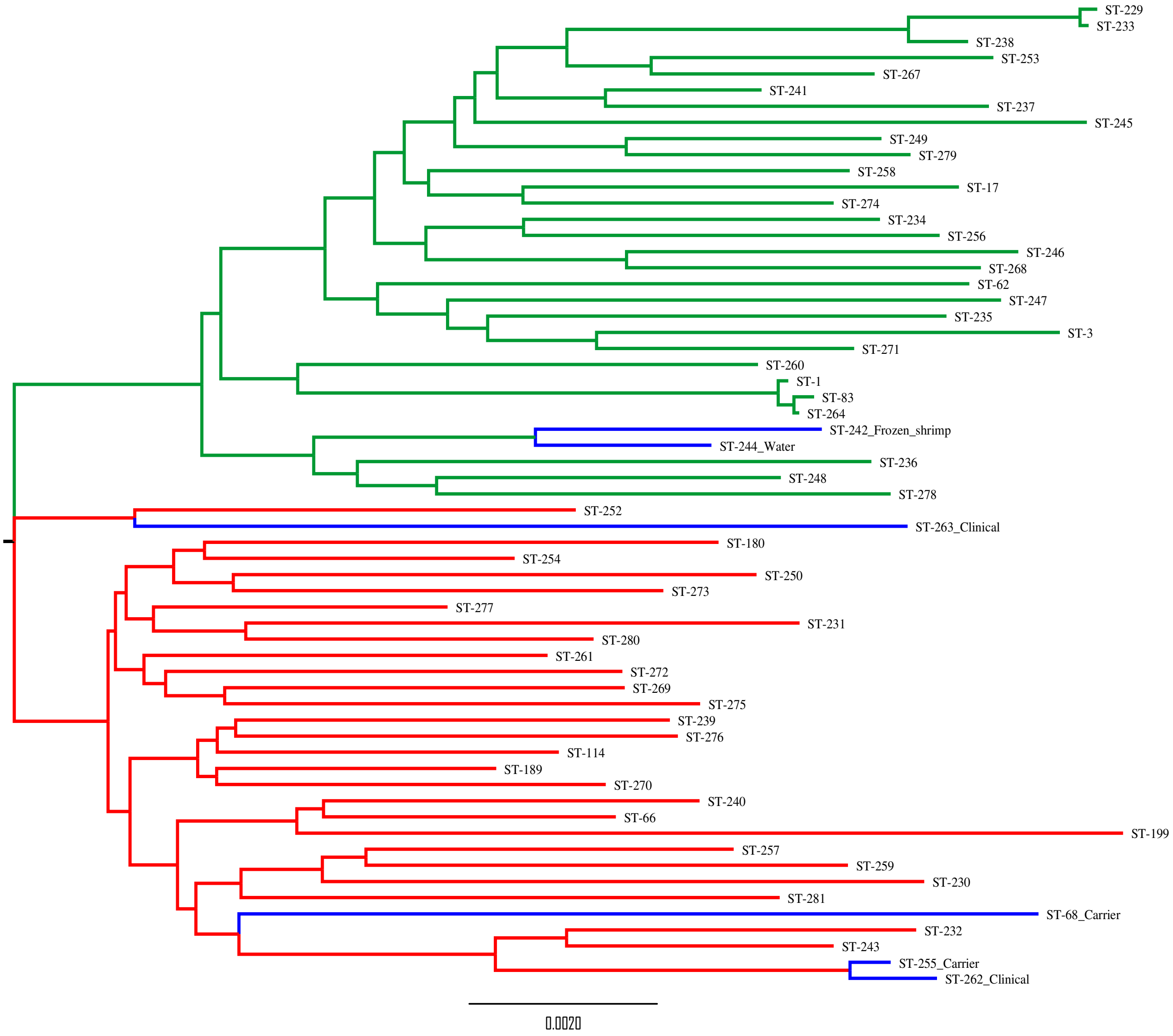


FIG S4 Neighbour-Joining tree representing BAPS clusters and admixture STs. The lineages representing different STs are coloured according to the BAPS cluster classification. Red and green colours represent STs in distinct population clusters whereas blue represent admixture STs. The population structure was obtained using the admixture model where K = 2.

Population clustering of *V. parahaemolyticus* 61 STs

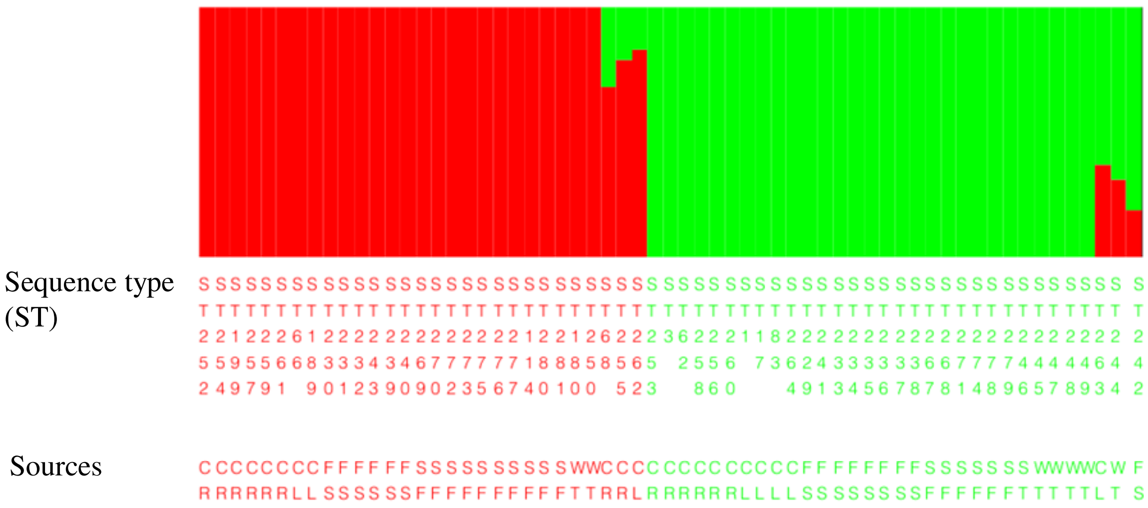


FIG S5 Bayesian clustering analysis inferred by BAPS for 61 STs indicate two distinct clusters (red and green). The colours represent different population clusters and the coloured segment represents the fraction of genotype belonging to each cluster. The population structure was obtained using an admixture analysis model where K = 2. Each individual ST and the source of isolates are represented by the horizontal scale. CL represents clinical isolates, CR represents carrier isolates, SF represents seafood isolates, FS represents frozen shrimp isolates, ST represents shrimp tissue isolates, and WT represent water isolates.

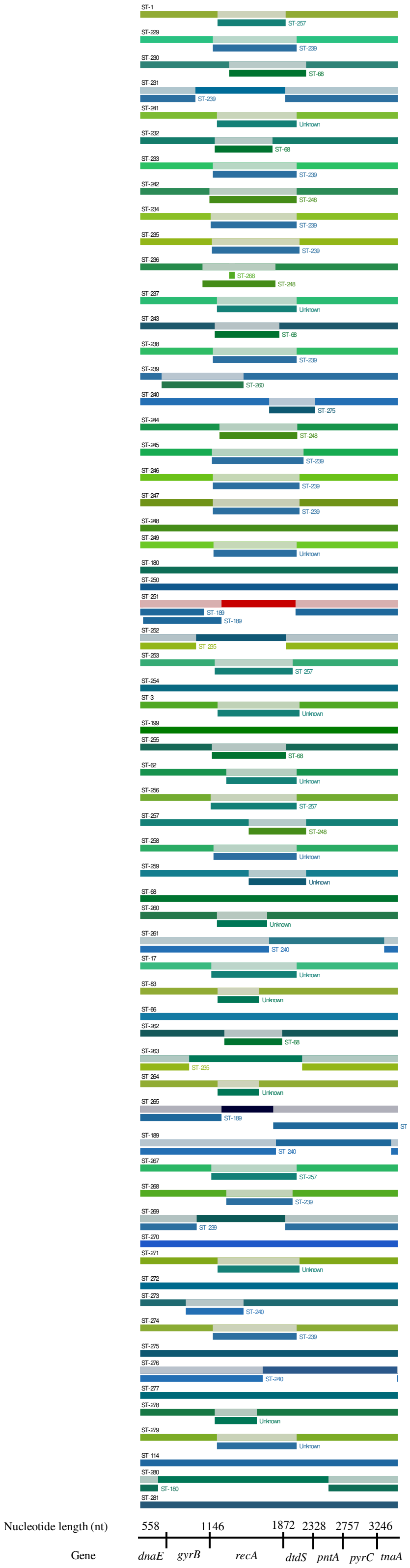


FIG S6 Recombination events among 63 STs predicted by RDP3 analysis. Thirteen unique recombination events were identified from 175 recombination signals detected among 63 STs of *V. parahaemolyticus*. The figure shows the estimated recombination breakpoints in the nucleotide sequences of 63 STs. Predominant recombination events occurring at the *recA* locus are represented.

TABLE S1 Nucleotide sequences of PCR and sequencing primers designed for DNA amplification and sequencing of seven housekeeping gene fragments used for MLST of *V. parahaemolyticus.* Primers -F1 and -R1 represent PCR primers; primers -F2 and -R2 represent sequencing primers.

Gene	**Primer**	**PCR/Sequencing**	**Base position**	**Sequence (5'-3')**
dnaE	dnaE-F1	PCR	639-656	CGA GAT TCG TGT TGC GAT
	dnaE-R1	PCR	1414-1397	CTA GCG TCA TAC CCG GAT
	dnaE-F2	Sequencing	735-752	AAT GTG TGA GCT GTT TGC
	dnaE-R2	Sequencing	1329-1312	ACG GAT TAC CGC TTT CGC
gyrB	gyrB-F1	PCR	582-599	GTT CTT GAA CTC AGG CGT
	gyrB-R1	PCR	1339-1322	GTG GTA GGA TTG CCT GAT
	gyrB-F2	Sequencing	655-671	GAA GGT GGT ATT CAA GC
	gyrB-R2	Sequencing	1281-1264	GTC ACC CTC CAC AAT GTA
recA	recA-F1	PCR	75-92	CAT GCG CCT TGG TGA TAA
	recA-R1	PCR	1006-989	CAG GTG CTT CTG GTT GAG
	recA-F2	Sequencing	111-128	AAC CAT TTC AAC GGG TTC
	recA-R2	Sequencing	877-860	TGT AGC TGT ACC AAG CAC
dtdS	dtdS-F1	PCR	47-64	GGA TGA CCG AAG TAG ACA
	dtdS-R1	PCR	618-601	AGC AAG CTC TAG ACG GTA
	dtdS-F2	Sequencing	75-92	TGG CCA TAA CGA CAT TCT
	dtdS-R2	Sequencing	571-554	GAG CAC CAA CGT GTT TAG
pntA	pntA-F1	PCR	600-617	TGA CGT TCG TCC AGA AGT
	pntA-R1	PCR	1275-1258	TAC CGA TGC AAT CCA AGC
	pntA-F2	Sequencing	671-688	AAG ACT CTG GTT CTG GTG
	pntA-R2	Sequencing	1158-1141	TTG AGG CTG AGC CGA TAC
pyrC	pyrC-F1	PCR	194-211	GCG AAC AAT TCG AAC CTC
	pyrC-R1	PCR	874-857	TTG CGA ACG CTT CCA AGT
	pyrC-F2	Sequencing	269-286	CAA CCG GTA AAA TTG TCG
	pyrC-R2	Sequencing	798-781	AGT GTA AGA ACC GGC AGA
tnaA	tnaA-F1	PCR	548-565	TCT GTG CCA TCA TCA CGA
	tnaA-R1	PCR	1147-1130	CCT CCA GAT ACA ACG CAT
	tnaA-F2	Sequencing	591-608	CCA ACC GGT ATC GAT GGA
	tnaA-R2	Sequencing	1083-1066	TAT TTT CGC CGC ATC AAC

TABLE S2 Nucleotide sequences of additional PCR and sequencing primers designed for amplification and sequencing of seven housekeeping gene fragments of the genes which could not be amplified or sequenced by the primers presented in Table S1.

Gene	Primer	PCR/Sequencing	Base position	Sequence (5'-3')
dnaE	dnaE-F3	PCR	677-694	AAG ATC CAC GCC GAC CAA
	dnaE-R3	PCR	1470-1453	TTC CTC ATC GGC CTC ATA
recA	recA-F3	PCR	54-71	GCA ATT CGG TAA AGG CTC
	recA-F4	PCR	32-49	CTG CGC TAG GTC ARA TTG
	recA-R3	PCR	925-908	GCA GGT AGT TAC AAG CGT
	recA-R4	PCR	1032-1015	TTC TTG CTC AGG CTT CTC
dtdS	dtdS-F3	PCR	23-40	AGC TAA AGC CTG AAS AAG
	dtdS-R3	PCR	645-628	TAC TGC RCG YGT TAC GCC
	dtdS-R5	Sequencing	598-580	CGT TTA CGT CTG TGA TAA C
pntA	pntA-F3	PCR	578-595	TTG GCG CTA TCG TTC GTG
	pntA-F4	PCR	625-642	CAA GTT GAG TCG ATG GGT
	pntA-R3	PCR	1245-1228	GGC TGC AAC AAG ACC AAT
	pntA-R4	PCR	1249-1232	CAA CGG CTG CAA CAA GAC
pyrC	pyrC-F7	Sequencing	232-249	GAT AAC ACC ACG CCA GAA
tnaA	tnaA-F3	PCR	567-584	AGT GAC GTG TAA CAG CTC
	tnaA-F3	Sequencing	623-640	TGT ACG AAA TTG CCA CCA
	tnaA-R3	PCR	1128-1111	ACA CAA GGC TTG TGC TGG

TABLE S3 Amino acid sequence types (aaSTs) correspond to nucleotide sequence types (mlstSTs) of *V. parahaemolyticus* 348 STs from MLST database (http://pubmlst.org/vparahaemolyticus/)

aaST	mlstSTs
1	19, **269**, 299
2	**1,**4,5,9,10,13,16,**17**,18,22,23,24,25,26,87,98,102,109,110,122,130,132,133,153,157,158,167,170,183,217,219,227,228,**234,239,241,243,248,249,254,267**,**271**,293,300,309,322,330,348
3	191
4	209, **277**, 321
5	82,**83,264**
6	112
7	2,**3**,21,27,71,72,159,176,192,196,201,211,220,**235**,266,295,305,343
8	**66**
9	**68**
10	146, 151, 155, 203, **260**
11	63,140,208,323
12	6,7
13	14
14	**256**
15	**258**
16	288
17	12,20,74,76,85,113,116,121,124,166,172,202,210,301,310
18	311
19	86
20	64,328,332
21	126
22	15, **255**
23	73
24	28,135
25	148,152
26	342
27	165
28	163
29	317
30	88, **189, 265,** 345
31	282
32	308

TABLE S3 (continued)

aaST	**mlstSTs**
33	**229,233**
34	8,30,31,37,46,47,49,50,53,55,56,57,60,**62**,67,70,79,80,84,91,92,97,100,101,104,105,106,115,117,119,120,123,125,131,134,139,141,142,147,156,161,164,171,173,174,178,**180,**188,194,195,198,200,207,213,214,215,216,218,223,224,226,**242,244,245,246,247,251,253,259,263,270,273,275,280**,289,291,304,312,315,316,319,325,326,327,329,335,336,341
35	48, 225, **268**
36	75,78,181,284,340
37	205
38	**232**
39	36,38,39,40,41,42,51,52,59,107,111,129,137,138,187,193,**199,237**
40	54, **230**
41	96,143,185,283,287,200,000
42	169
43	128
44	**238**
45	90,108,168,318
46	11,221
47	**257**
48	204
49	**250**
50	285
51	297
52	29,99,136,145,154,179,190,**240,272,276**,337
53	94, 162, 197, **278,** 320
54	286
55	32,33,34,77,95,306,307,324,334
56	127, 186, 206, **236, 262**
57	**261**
58	89
59	150
60	43,44,81,314,346
61	313
62	61
63	149
64	294
65	339

TABLE S3 (continued)

aaST	mlstSTs
66	45
67	93,144,177,296,333,300
68	69,222
69	290,298
70	58
71	184
72	103
73	160
74	182
75	35
76	**252**
77	65
78	118
79	303
80	**114**
81	212, **274**
82	175
83	**231**
84	**281**
85	**279**
86	331,344
87	338

Numbers in bold represent STs that were found in Thai *V. parahaemolyticus* isolates in the present study.