Supplementary Material for:

Homopolymer tail-mediated ligation PCR: a streamlined and highly efficient method for DNA cloning and library construction

David W. Lazinski and Andrew Camilli

Howard Hughes Medical Institute and Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, MA, USA

BioTechniques 54:25-34 (January 2013) doi 10.2144/000113981

Keywords: molecular cloning; annealing-assisted ligation; DNA capture; massively-parallel sequencing

Supplemental Materials and Methods

DNA fragmentation and homopolymer tail addition

Genomic DNA was prepared from *Vibrio cholerae* strain E7946 using the Blood and Tissue kit (Qiagen, Valencia, CA). One microgram of E7946 DNA was brought to a total volume of 100 µL in pure water and put through a Performa DTR Gel Filtration Cartridge (Edge Biosystems, Gaithersburg, MD) according to the manufacturer's instructions to remove any small molecules. The eluate was placed in a 2-mL microfuge tube in a prechilled (4°C) high-intensity cuphorn sonifier (Branson) for 2 min at 100% intensity using a 10 s on and 5 s off duty cycle and sheared to a size range, where on a molar basis, the majority of DNA was 100–800 bp. The DNA ends (100 ng, 10 ng, 1 ng, 0.1 ng, or 0.01 ng of sheared template) were made blunt and 5′ ends phosphorylated by treatment with the Quick Blunting kit (New England BioLabs, Ipswich, MA) according to the manufacturer's directions, then heat-inactivated at 75°C for 30 min. Small molecules were removed using a Performa DTR Gel Filtration Cartridge. A deoxycytosine (dC) homopolymer tail of 20 nucleotide average length was added to the 3′ ends of DNA by treatment with 0.5 µL terminal deoxynucleotidyl transferase (Promega, Madison, WI), 450 µM 2′-deoxycytidine 5′-triphosphate (dCTP), 50 µM 2′,3′-dideoxycytidine 5′-triphosphate (ddCTP) and 1x TdT reaction buffer (Promega) at 37°C for 1 h. Alternatively, a homopolymer tail of 2-amino deoxyadenosine 5′-triphosphate was added to 100 ng of blunted template DNA using a 29:1 molar ratio of 2-amino dATP:ddATP (Trilink Biotechnologies, San Diego, CA) as above. The reactions were heat-inactivated at 75°C for 30 min and small molecules were removed using a Performa DTR Gel Filtration Cartridge.

Oligonucleotide ligation

The oligonucleotide olj623 (5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCGGGGGGG-3′; leftward Illumina sequence underlined) or olj682 (5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCGGGTTTTTTT-3′) was ligated to the 5′ ends of the *V. cholerae* dC or 2-amino dA homopolymer-tailed DNA samples, respectively, using T4 DNA ligase (New England Biolabs) as follows: 24 µL of DNA, 2 µL of 30 µM oligonucleotide, 3 µL 10x T4 DNA ligase buffer and 1 µL T4 DNA ligase and incubated at room temperature for 1 h. Small molecules were removed using a Performa DTR Gel Filtration Cartridge.

Amplification by PCR

The dC and 2-amino dA homopolymer tailed gDNA samples were amplified in single PCR reactions as follows: 28 µL DNA, 2 µL of 10 mM dNTPs, 5 µL 10x EasyA reaction buffer (Stratagene, Agilent Technologies, Santa Clara, CA), 1 µL 30 µM forward primer, 1 µL 30 µM reverse primer, 12 µL water and 1 µL Easy A enzyme (Stratagene). For the dC-tailed sample, the primers were olj623 (above) and a barcoded primer of the form 5′-CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGGGGGGGGGGGGG-3′; where NNNNNN designates the reverse complement of the barcode used and the rightward Illumina sequence is underlined) and the amplification conditions were as follows: 95°C for 2 min; 12–36 cycles of 95°C for 30 s, 65°C for 30 s, 72°C for 120 s; 72°C for 120 s; 4°C hold. For the 2-amino dA-tailed sample, the primers were olj682 (above) and olj684 (5′-CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTTTTTTTTTTTTTTTTTTTTTTT-3′)and the amplification conditions were as follows: 95°C for 2 min; 15 cycles of 95°C for 30 s, one of seven different annealing temperatures (48°C, 52.1°C, 54.7°C, 58.3°C, 62.4°C, 65.7°C, or 68.4°C) for 30 s, 72°C for 120 s; 72°C for 120 s; 4°C hold.

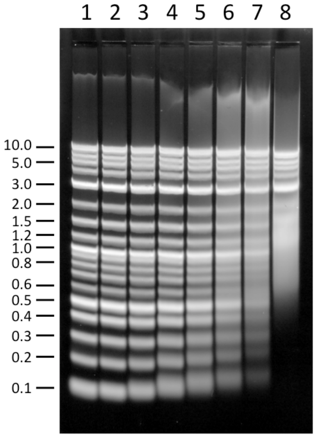
DNA sequencing and data analysis

The DNA was sequenced for 51 bases from a single-end using a HiSeq2000 (Illumina, San Diego, CA) with the custom sequencing primer: olj628 (5′-ACACTCTTTCCCTACACGACGCTCGGGGGGG-3′). The resulting sequencing reads were manipulated and mapped to the reference genome using Genomics Workbench (CLCbio, Cambridge, MA). Reads were filtered for quality using the "trim sequences" feature with the following input parameters: trim using quality scores limit = 0.001, maximum number of ambiguities = 0, discard reads below length = 51. Homopolymer sequences were not specifically trimmed, however, we observed that Illumina software assigns lower quality scores to bases present in homopolymer stretches. Hence, homopolymer tails were removed upon filtering for quality. Next, filtered or unfiltered reads were used with the reference genome (GenBank accession nos. NC_002505, NC_002506, AB374228, and NC_003028.3) as input for the "map reads to reference" feature with mismatch cost = 1, limit = 6, fast ungapped alignment unchecked, insertion cost = 2, deletion cost = 2, and global alignment checked. Finally, the single nucleotide polymorphism (SNP) and deletion/insertion polymorphism (DIP) detection features were used with *V. cholerae* reference assemblies and default input parameters except that the minimum coverage was set to 4 for both SNP and DIP detection while the minimum SNP variant frequency was set to 90% and the minimum DIP variant frequency was set to 70%. Coverage level distribution and coverage versus GC content were determined using the "create detailed mapping report" function. Unique 5′ end counting was done on the Tufts University Galaxy server using the *hopcount* script as previously described (2).

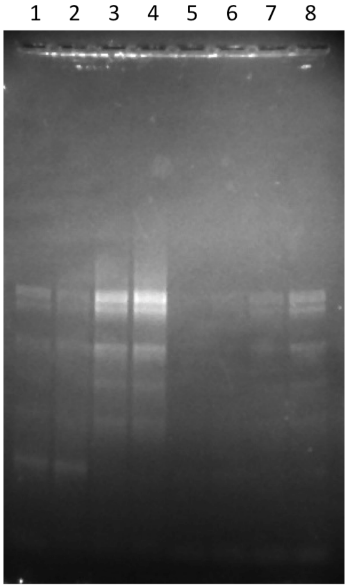
References

1. **Boulé, J.-B., F. Rougeon, and C. Papanicolaou.** 2001. Terminal deoxynucleotidyl transferase indiscriminately incorporates ribonucleotides and deoxyribonucleotides. J. Biol. Chem. 276(33):31388–31393.

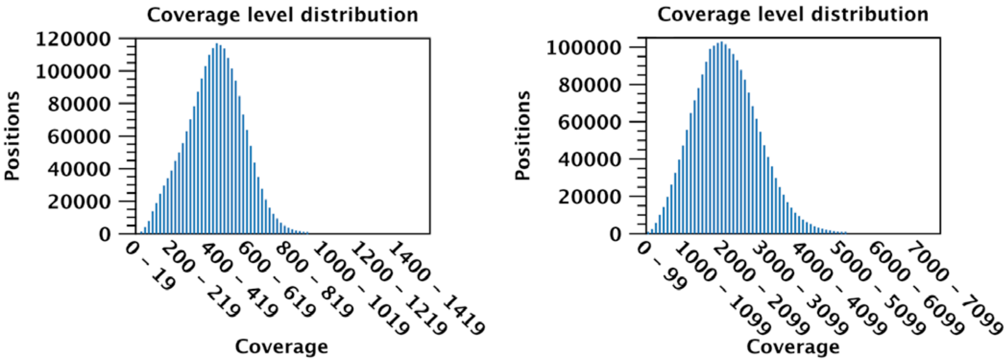
2. **Klein, B.A., E.L. Tenorio, D.W. Lazinski, A. Camilli, M.J. Duncan, and L.T. Hu.** 2012. Identification of essential genes of the periodontal pathogen *Porphyromonas gingivalis*. BMC Genomics *13*:578.



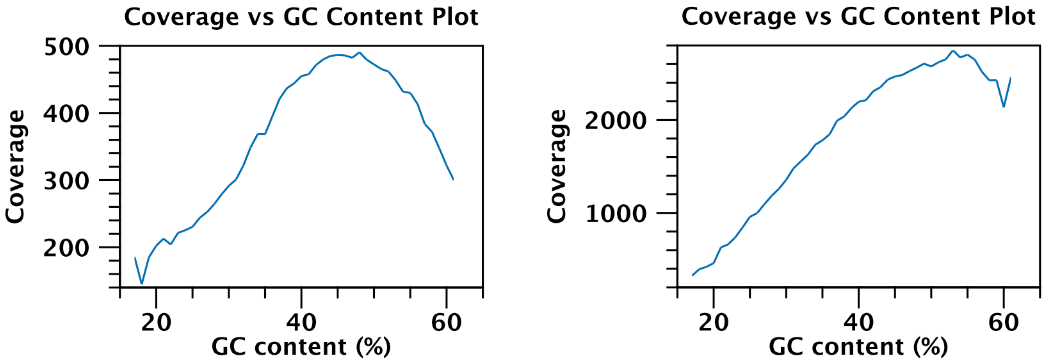
Supplementary Figure S1. Titration of ddCTP with dCTP in tailing reactions resolved with 2.0% agarose electrophoresis. All tailing reactions contain 1μg 2-log DNA ladder (NEB) and were set up as described in the “Materials and methods” section, except that concentrations of ddCTP to dCTP were varied as follows: lane 1: 0 μM ddCTP, 0 μM dCTP; lane 2: 250 μM ddCTP, 250 μM dCTP; lane 3: 125 μM ddCTP, 375 μM dCTP; lane 4: 67.5 μM ddCTP, 432.5 μM dCTP; lane 5: 33.75 μM ddCTP, 466.25 μM dCTP; lane 6: 16.88 μM ddCTP, 483 μM dCTP; lane 7: 8.44 μM ddCTP, 491.56 μM dCTP; lane 8: 0 μM ddCTP, 500 μM dCTP. The sizes of bands in lane 1, from bottom to top, are 100 bp, 200 bp, 300 bp, 400 bp, 500 bp, 600 bp, 700 bp, 800 bp, 900 bp, 1000 bp, 1200 bp, 1500 bp, 2000 bp, 3000 bp, 4000 bp, 6000 bp, 8000 bp, and 10,000 bp. Note in lane 8 the complete absence of bands smaller than 400 bp and the smearing of bands greater than this size. This indicates that, in the absence of ddCTP, TdT added tails to the 100-bp band of sufficient length that the resulting molecules ran with apparent sizes in excess of 400 bp. Furthermore, since all of the 100-, 200-, and 300-bp bands were absent from lane 8, we conclude that TdT quantitatively added these long tails to essentially all molecules. Note in lane 7 that, relative to lane 1, the mobility of the center of the 100-bp band was shifted by ~60 bp. Thus, when ddCTP represented only 1/64 of the total amount of deoxynucleotide triphosphate precursor, the apparent mobility was shifted by ~64 bp. To a first approximation, this indicates that the affinity of TdT is similar for ddCTP and dCTP; this is consistent with previous observations (1).



Supplementary Figure S2. Evaluation of the efficiency of HTML-PCR reactions promoted by different lengths of dC/dG pairing during the ligation step. One nanogram of pUC19 DNA was linearized with the blunt-cleaving restriction endonuclease BstU I, C-tailed, and then ligated to oligonucleotides that were identical except for the length of their 3´ oligo dG stretches. In lanes 1 and 5, the oligonucleotide contained four dG residues at its 3´ terminus, five dG residues in lanes 2 and 6, six dG residues in lanes 3 and 7, and seven dG residues in lanes 4 and 8. Following ligation, the eight substrates were amplified in PCR reactions with common primesr for either 30 cycles (lanes 1–4) or 20 cycles (lanes 5–8) and the products were resolved by 2.0% agarose electrophoresis. Of the 20-cycle reactions, lane 8 showed the greatest intensity of products; hence, we conclude that the added stability of seven dC/dG pairs enhanced the efficiency of ligation.



Supplementary Figure S3. Comparison of the coverage level distributions of libraries generated by HTML-PCR (A) or Nextera (B). Note that five-fold more of the Nextera sample was sequenced relative to the HTML-PCR sample; thus, respective x axes are scaled accordingly.



Supplementary Figure S4. Comparison of the coverage (as a function of GC content) of libraries generated by HTML-PCR (A) or Nextera (B). Note that five-fold more of the Nextera sample was sequenced relative to the HTML-PCR sample; thus, respective y axes are scaled accordingly.

Supplementary Table S1. Single Nucleotide Polymorphisms (SNPs) in the E7946 Sequence.

Reference Genome used for mapping	Reference Position	Variation Type	Reference nucleotide*	SNP Variation*	Frequency (%)**	Counts	Coverage	Annotations for site	Amino Acid Change
NC_002505	42082	SNP	A	C	100	107	107	Gene: Gene, CDS: sun protein	Asp407Glu
NC_002505	125122	SNP	A	C	100	261	261	Gene: Gene, CDS: GGDEF family protein	Asn451His
NC_002505	162660	SNP	N	C	100	207	207	Gene: Gene, CDS: hypothetical protein	Xxx330Gly
NC_002505	172531	SNP	K	T	100	224	224	Gene: Gene, CDS: peptide ABC transporter,_permease protein	Xxx178Leu
NC_002505	262291	SNP	G	T	94.4	17	18	Gene: Gene, CDS: rfbT-related protein	Cys65Phe
NC_002505	263544	SNP	T	C	100	5	5	Gene: Gene, CDS: rfbT protein	
NC_002505	263545	SNP	A	C	100	8	8	Gene: Gene, CDS: rfbT protein	Thr4Pro
NC_002505	364992	SNP	K	G	99.1	211	213	Gene: Gene, CDS: hypothetical protein	Xxx69Gly
NC_002505	401063	SNP	A	G	100	227	227	Gene: Gene, CDS: phosphoadenosine phosphosulfate reductase	
NC_002505	424629	SNP	G	T	100	170	170	Gene: Gene, CDS: regulatory protein CsrD	
NC_002505	424772	SNP	G	T	98.6	207	210	Gene: Gene, CDS: regulatory protein CsrD	Arg504Leu
NC_002505	479266	SNP	G	T	91.7	11	12		
NC_002505	575609	SNP	A	G	100	228	228	Gene: recA, CDS: *recA*	Tyr363Cys
NC_002505	581125	SNP	T	G	98.6	70	71	Gene: tRNA-Arg-3, tRNA: tRNA-Arg-3	
NC_002505	663766	SNP	T	A	99.5	217	218	Gene: Gene, CDS: hypothetical protein	Leu82Phe
NC_002505	697746	SNP	A	G	100	288	288	Gene: Gene, CDS: c-di-GMP phosphodiesterase A-related protein	Val555Ala
NC_002505	854836	SNP	K	G	100	168	168	Gene: Gene, CDS: citrate lyase,_beta subunit	Xxx91Gly
NC_002505	906942	SNP	A	G	100	195	195	Gene: Gene, CDS: hypothetical protein	
NC_002505	929971	SNP	T	C	100	88	88	Gene: Gene, CDS: Slt family transglycosylase	
NC_002505	1011632	SNP	W	A	99.6	233	234	Gene: Gene, CDS: D-alanyl-D-alanine carboxypeptidase	Xxx194Asp
NC_002505	1183050	SNP	A	G	100	200	200	Gene: Gene, CDS: 8-amino-7-oxononanoate synthase	Arg20Gly
NC_002505	1237259	SNP	C	G	97.9	187	191	Gene: Gene, CDS: hypothetical protein	Glu192Gln
NC_002505	1237260	SNP	G	C	98.9	188	190	Gene: Gene, CDS: hypothetical protein	His191Gln
NC_002505	1335598	SNP	A	G	100	157	157	Gene: Gene, CDS: hypothetical protein	Asn107Ser
NC_002505	1417919	SNP	T	G	99.4	158	159	Gene: Gene, CDS: hypothetical protein	
NC_002505	1542173	SNP	C	T	99.4	180	181	Gene: Gene, CDS: sensor histidine kinase/response regulator	Ser241Leu
NC_002505	1561424	SNP	K	G	100	174	174	Gene: Gene, CDS: RTX toxin RtxA	Xxx3773Asp
NC_002505	1561926	SNP	S	C	100	164	164	Gene: Gene, CDS: RTX toxin RtxA	Xxx3940Ala
NC_002505	1561942	SNP	S	G	100	151	151	Gene: Gene, CDS: RTX toxin RtxA	
NC_002505	1561949	SNP	W	A	100	158	158	Gene: Gene, CDS: RTX toxin RtxA	Xxx3948Asn
NC_002505	1587146	SNP	K	G	100	4	4	Gene: Gene, CDS: ATP-dependent protease LA-related protein	
NC_002505	1587147	SNP	Y	T	100	4	4	Gene: Gene, CDS: ATP-dependent protease LA-related protein	
NC_002505	1587148	SNP	Y	C	100	6	6	Gene: Gene, CDS: ATP-dependent protease LA-related protein	
NC_002505	1587149	SNP	Y	T	100	10	10	Gene: Gene, CDS: ATP-dependent protease LA-related protein	Xxx196Ser
NC_002505	1587765	SNP	W	T	99.5	205	206	Gene: Gene, CDS: ATP-dependent protease LA-related protein	Xxx401Val
NC_002505	1591816	SNP	A	G	100	194	194	Gene: Gene, Restriction site: BsaBI, CDS: ABC transporter ATPase component	

Supplementary Table S1 (continued). Single Nucleotide Polymorphisms (SNPs) in the E7946 Sequence.

Reference Genome used for mapping	Reference Position	Variation Type	Reference nucleotide*	SNP Variation*	Frequency (%)**	Counts	Coverage	Annotations for site	Amino Acid Change
NC_002505	1670764	SNP	W	T	100	234	234		
NC_002505	1696639	SNP	Y	C	99.2	130	131	Gene: Gene, CDS: catalase	Asx522Asp
NC_002505	1696645	SNP	Y	C	99.2	119	120	Gene: Gene, CDS: catalase	Xxx520Ala
NC_002505	1735241	SNP	M	C	99.5	201	202	Gene: Gene	
NC_002505	1792474	SNP	C	A	100	246	246	Gene: Gene	
NC_002505	2010846	SNP	G	A	100	112	112	Gene: Gene, CDS: methyl-accepting chemotaxis protein	Arg483Cys
NC_002505	2076300	SNP	A	G	100	187	187	Gene: Gene, CDS: C4-dicarboxylate transport sensor protein	Asn187Ser
NC_002505	2122955	SNP	R	G	100	145	145	Gene: Gene, CDS: O-succinylbenzoic acid--CoA ligase	
NC_002505	2122956	SNP	M	C	100	144	144	Gene: Gene, CDS: O-succinylbenzoic acid--CoA ligase	Xxx110Trp
NC_002505	2148579	SNP	A	G	100	201	201	Gene: Gene, CDS: protease IV	
NC_002505	2153724	SNP	T	A	98.9	174	176	Gene: Gene, CDS: hypothetical protein	Ile106Asn
NC_002505	2179341	SNP	C	G	100	177	177	Gene: maf, CDS: *maf*	Leu47Val
NC_002505	2318560	SNP	G	A	100	9	9	Gene: Gene, CDS: hypothetical protein	Leu19Phe
NC_002505	2378839	SNP	N	G	100	214	214	Gene: Gene, CDS: GGDEF family protein	Xxx227Ser
NC_002505	2379972	SNP	A	G	100	128	128	Gene: *upp*, CDS: *upp*	Ile135Thr
NC_002505	2488985	SNP	G	T	99.6	252	253	Gene: Beta Galactosidase (LacZ)	
NC_002505	2496376	SNP	T	C	100	196	196	Gene: Gene, CDS: elongation factor G	Ser370Gly
NC_002505	2590300	SNP	R	A	100	194	194	Gene: Gene, CDS: 2`,3`-cyclic-nucleotide 2`-phosphodiesterase, putative	Xxx481Glu
NC_002505	2682957	SNP	R	G	100	178	178	Gene: Gene, CDS: hypothetical protein	Xxx73Pro
NC_002505	2682978	SNP	S	G	100	142	142	Gene: Gene, CDS: hypothetical protein	Xxx66Pro
NC_002505	2881400	SNP	C	A	98.8	164	166	Gene: Gene, CDS: bifunctional (p)ppGpp synthetase II/ guanosine-3',5'-bis pyrophosphate 3'-pyrophosphohydrolase	Pro142His
NC_002506	44129	SNP	A	G	100	57	57	Gene: Gene, CDS: serine/threonine transporter SstT	Val390Ala
NC_002506	85485	SNP	T	G	100	124	124	Gene: Gene, CDS: hypothetical protein	Ile247Leu
NC_002506	152561	SNP	C	A	99.6	247	248	Gene: *glpT*, CDS: *glpT*	Gly44Val
NC_002506	161478	SNP	C	G	99.3	133	134	Gene: Gene, CDS: hypothetical protein	Arg377Gly
NC_002506	186754	SNP	C	T	100	5	5	Gene: Gene, CDS: hypothetical protein	
NC_002506	219397	SNP	T	C	99.2	131	132	Gene: Gene, CDS: hypothetical protein	
NC_002506	277301	SNP	A	G	100	157	157	Gene: Gene, CDS: sensor histidine kinase	
NC_002506	310944	SNP	G	T	100	17	17		
NC_002506	310945	SNP	C	A	100	16	16		
NC_002506	310946	SNP	T	A	100	9	9		
NC_002506	312308	SNP	G	A	100	4	4		
NC_002506	312309	SNP	C	A	100	10	10		
NC_002506	312310	SNP	G	A	100	16	16		
NC_002506	356433	SNP	Y	T	100	24	24		
NC_002506	366180	SNP	Y	C	100	186	186	Gene: Gene, CDS: toxin resistance protein	
NC_002506	566512	SNP	C	G	97.7	209	214	Gene: Gene, CDS: hypothetical protein	Arg256Pro
NC_002506	566513	SNP	G	C	100	204	204	Gene: Gene, CDS: hypothetical protein	Arg256Gly

Supplementary Table S1 (continued). Single Nucleotide Polymorphisms (SNPs) in the E7946 Sequence.

Reference Genome used for mapping	Reference Position	Variation Type	Reference nucleotide*	SNP Variation*	Frequency (%)**	Counts	Coverage	Annotations for site	Amino Acid Change
NC_002506	588148	SNP	C	G	100	149	149	Gene: Gene, CDS: PTS system,_sucrose-specific IIBC component	Arg115Pro
NC_002506	646810	SNP	Y	T	100	220	220	Gene: Gene, CDS: phosphoglycerate transport regulatory protein PgtB,_putative	
NC_002506	788542	SNP	G	T	99.6	228	229		
NC_002506	811408	SNP	T	C	100	216	216	Gene: Gene, CDS: hypothetical protein	
NC_002506	886005	SNP	C	T	100	91	91		
NC_002506	894743	SNP	G	T	99.5	205	206	Gene: *malF*, CDS: *malF*	Phe397Leu
NC_002506	901045	SNP	C	T	100	206	206	Gene: Gene, CDS: hypothetical protein	Ser36Phe
NC_002506	980511	SNP	C	G	96.6	199	206	Gene: Gene	
NC_002506	980512	SNP	G	C	100	201	201	Gene: Gene	
NC_002506	980632	SNP	C	G	98.3	229	233	Gene: Gene	
NC_002506	980633	SNP	G	C	97	227	234	Gene: Gene	
NC_002506	986611	SNP	A	G	100	6	6	Gene: Gene	
NC_002506	993684	SNP	C	A	100	101	101	Gene: Gene, CDS: tagE protein	
NC_002506	993701	SNP	T	C	100	7	7	Gene: Gene, CDS: tagE protein	Arg200Gly
NC_002506	1004685	SNP	T	G	100	126	126	Gene: Gene, CDS: LysR family transcriptional regulator	Val73Gly
AB374228	4195	SNP	A	T	99.6	239	240	Gene: *cII*, CDS: *cII*	Gln103Leu
AB374228	4296	SNP	A	T	100	87	87	Gene: *cII*, CDS: *cII*	Ser137Cys
AB374228	4302	SNP	A	G	100	99	99	Gene: *cII*, CDS: *cII*	Thr139Ala

*IUPAC notation
**Frequency is equal to number of SNP reads (count) divided by the total reads (coverage)

Supplementary Table S2. Deletion/Insertion Polymorphisms (DIPs) in the E7946 Sequence.

Reference Genome used for mapping	Reference Position	Length	Reference nucleotide*	DIP Variation*	Frequency (%)**	Counts	Coverage	Annotations for site	Amino Acid Change
NC_002505	359986	1	-	T	85.3	185	217	Gene: Gene, CDS: ribosome-associated GTPase	Change, frameshift
NC_002505	424645	1	-	G	87	120	138	Gene: Gene, CDS: regulatory protein CsrD	Change, frameshift
NC_002505	424670	1	G	-	88.2	67	76	Gene: Gene, CDS: regulatory protein CsrD	Change, frameshift
NC_002505	479266	1	G	-	94.2	196	208		
NC_002505	540850	1	C	-	90.3	176	195		
NC_002505	553569	1	-	C	97.4	190	195	Gene: *dnaG*, CDS: *dnaG*	Change, frameshift
NC_002505	555404	1	-	G	79.2	80	101	Gene: *gcp*, CDS: *gcp*	Change, frameshift
NC_002505	612959	1	-	A	86.8	165	190	Gene: Gene	
NC_002505	623228	1	-	G	94	205	218	Gene: Gene, CDS: glutamyl-Q tRNA(Asp) synthetase	Change, frameshift
NC_002505	638168	1	-	T	90.5	218	241	Gene: Gene	
NC_002505	807832	1	-	G	95.5	171	179		
NC_002505	849965	1	-	G	89.2	91	102	Gene: Gene, Gene: *oadA-2*, CDS: *oadA-2*	Change, frameshift
NC_002505	872713	1	-	A	81.4	114	140		
NC_002505	881244	1	-	T	87.8	158	180	Gene: Gene, CDS: hypothetical protein	Change, frameshift
NC_002505	998435	1	-	C	95.8	205	214	Gene: Gene, Gene: Gene, CDS: hypothetical protein, CDS: hypothetical protein	Change, frameshift
NC_002505	1087962	1	-	G	95.4	103	108	Gene: *uvrB*, CDS: *uvrB*	Change, frameshift
NC_002505	1145407	1	-	G	90.6	116	128	Gene: Gene, CDS: AraC/XylS family transcriptional regulator	Change, frameshift
NC_002505	1228191	1	-	C	91.9	113	123	Gene: Gene, Gene: Gene, CDS: sensor histidine kinase, CDS: response regulator	Change, frameshift
NC_002505	1337746	1	T	-	85.5	106	124		
NC_002505	1389484	1	-	C	90.8	138	152	Gene: Gene	
NC_002505	1470952	1	-	G	96	167	174	Gene: Gene, CDS: GGDEF family protein	Change, frameshift
NC_002505	1486534	1	-	C	88.5	54	61	Gene: Gene	
NC_002505	1489715	1	-	T	81.7	161	197	Gene: Gene, CDS: chemotaxis protein methyltransferase CheR	Change, frameshift
NC_002505	1530608	1	-	C	79.7	55	69		
NC_002505	1618601	1	-	C	91.8	157	171	Gene: Gene, CDS: hypothetical protein	Change, frameshift
NC_002505	1625933	1	-	C	91.6	98	107	Gene: Gene	
NC_002505	1649049	1	C	-	90.4	103	114	Gene: Gene, CDS: hypothetical protein	Change, frameshift
NC_002505	1676600	1	-	C	87.6	113	129	Gene: Gene, CDS: hypothetical protein	Change, frameshift
NC_002505	1676702	1	-	C	95.2	216	227	Gene: Gene, CDS: hypothetical protein	Change, frameshift
NC_002505	1676704	1	-	A	86	203	236	Gene: Gene, CDS: hypothetical protein	Change, frameshift
NC_002505	1748765	1	-	C	92.4	133	144	Gene: Gene	
NC_002505	1771830	1	-	A	85.8	199	232	Gene: Gene, CDS: hypothetical protein	Change, frameshift
NC_002505	1875074	1	-	T	84.1	191	227	Gene: Gene, CDS: hypothetical protein	Change, frameshift
NC_002505	1900854	1	T	-	86.9	185	213	Gene: Gene, CDS: hypothetical protein	Change, frameshift

Supplementary Table S2 (continued). Deletion/Insertion Polymorphisms (DIPs) in the E7946 Sequence.

Reference Genome used for mapping	Reference Position	Length	Reference nucleotide*	DIP Variation*	Frequency (%)**	Counts	Coverage	Annotations for site	Amino Acid Change
NC_002505	1990030	1	-	T	93.8	121	129	Gene: Gene, CDS: UDP-2,3-diacylglucosamine hydrolase	Change, frameshift
NC_002505	2051793	1	-	C	80.9	161	199		
NC_002505	2081077	1	-	C	90.1	183	203		
NC_002505	2194678	1	-	G	88.1	155	176	Gene: Gene, CDS: hypothetical protein	Change, frameshift
NC_002505	2262543	1	T	-	83.3	169	203	Gene: Gene, Gene: Gene, CDS: hypothetical protein, CDS: hypothetical protein	Change, frameshift
NC_002505	2266655	1	-	C	94.7	89	94	Gene: Gene	
NC_002505	2318560	1	G	-	91	91	100	Gene: Gene, CDS: hypothetical protein	Change, frameshift
NC_002505	2358194	1	-	C	92.9	221	238	Gene: *vibF*, CDS: *vibF*	Change, frameshift
NC_002505	2365248	1	-	G	85.8	187	218		
NC_002505	2489046	1	-	G	91.9	148	161	Gene: Beta Galactosidase (LacZ)	
NC_002505	2564032	1	-	C	90.1	182	202	Gene: *lpxC*, CDS: *lpxC*	Change, frameshift
NC_002505	2612864	1	-	C	87.1	128	147	Gene: Gene	
NC_002505	2657595	1	-	C	95.5	213	223	Gene: Gene, CDS: hypothetical protein	Change, frameshift
NC_002505	2738434	1	-	C	92.7	203	219	Gene: Gene, CDS: hypothetical protein	Change, frameshift
NC_002505	2839613	1	-	C	96.8	181	187	Gene: Gene, CDS: ribonuclease activity regulator protein RraA	Change, frameshift
NC_002505	2879855	1	T	-	91	181	199	Gene: Gene, CDS: hypothetical protein	Change, frameshift
NC_002505	2907037	1	-	G	90.3	187	207	Gene: Gene	
NC_002506	17393	1	C	-	96.5	165	171	Gene: Gene, Gene: Gene, CDS: hypothetical protein, CDS: glycogen branching enzyme	Change, frameshift
NC_002506	53608	1	T	-	90.8	177	195	Gene: Gene, Gene: Gene, CDS: hypothetical protein, CDS: protease-related protein	Change, frameshift
NC_002506	57465	1	-	G	89.6	103	115	Gene: Gene, Gene: Gene, CDS: GGDEF family protein, CDS: hypothetical protein	Change, frameshift
NC_002506	97575	1	G	-	92.2	202	219	Gene: Gene, CDS: proton/glutamate symporter	Change, frameshift
NC_002506	144626	1	A	-	82.4	122	148	Gene: *rbsC*, CDS: *rbsC*	Change, frameshift
NC_002506	184512	1	-	C	87	47	54		
NC_002506	186168	1	-	G	98.9	86	87	Gene: Gene, Gene: Gene, CDS: hypothetical protein, CDS: hypothetical protein	Change, frameshift
NC_002506	255374	1	-	G	83	78	94	Gene: Gene	
NC_002506	279450	1	-	C	87.6	113	129	Gene: Gene, CDS: hypothetical protein	Change, frameshift
NC_002506	294983	1	-	G	86.6	97	112	Gene: Gene	
NC_002506	299055	1	G	-	94.8	184	194	Gene: Gene	
NC_002506	299080	1	C	-	87.1	121	139	Gene: Gene	
NC_002506	403877	1	-	G	95.1	117	123	Gene: Gene, CDS: hypothetical protein	Change, frameshift
NC_002506	472297	1	A	-	97.4	149	153	Gene: Gene	
NC_002506	474527	1	-	C	90	208	231	Gene: Gene, CDS: hypothetical protein	Change, frameshift
NC_002506	530511	1	A	-	96.1	149	155		

Supplementary Table S2 (continued). Deletion/Insertion Polymorphisms (DIPs) in the E7946 Sequence.

Reference Genome used for mapping	Reference Position	Length	Reference nucleotide*	DIP Variation*	Frequency (%)**	Counts	Coverage	Annotations for site	Amino Acid Change
NC_002506	535093	1	G	-	93.4	211	226	Gene: Gene, Gene: Gene, CDS: hypothetical protein	Change, frameshift
NC_002506	545436	1	-	C	92.3	179	194		
NC_002506	609601	1	-	G	94.1	144	153	Gene: Gene	
NC_002506	631352	1	-	G	95.3	204	214	Gene: Gene	
NC_002506	631405	1	-	G	86.7	104	120	Gene: Gene	
NC_002506	686744	1	-	C	96	166	173		
NC_002506	690275	1	A	-	97.2	141	145	Gene: Gene	
NC_002506	743256	1	-	G	89	121	136		
NC_002506	765560	1	G	-	92.9	65	70	Gene: Gene, CDS: magnesium transporter MgtE,_putative	Change, frameshift
NC_002506	769970	1	-	T	94.8	147	155	Gene: Gene, CDS: diaminobutyrate--2-oxoglutarate aminotransferase	Change, frameshift
NC_002506	771746	1	-	G	97.6	160	164		
NC_002506	775705	1	-	G	89.3	216	242	Gene: Gene, CDS: acetoacetyl-CoA synthetase	Change, frameshift
NC_002506	869466	1	-	C	93.7	178	190		
NC_002506	889857	1	-	G	90.1	146	162	Gene: Gene	
NC_002506	893324	1	A	-	94.2	210	223		
NC_002506	969852	1	-	T	92.5	209	226	Gene: Gene, Gene: Gene, CDS: hypothetical protein, CDS: hypothetical protein	Change, frameshift
NC_002506	980877	1	G	-	95	191	201	Gene: Gene	
NC_002506	980886	1	-	G	92.8	192	207	Gene: Gene	
NC_002506	981512	1	A	-	92.4	121	131	Gene: Gene	
NC_002506	986611	2	AA	--	97.5	118	121	Gene: Gene	
NC_002506	993674	1	C	-	88.3	91	103	Gene: Gene, CDS: *tagE* protein	Change, frameshift
NC_002506	993700	2	CT	--	94	78	83	Gene: Gene, CDS: *tagE* protein	Change, frameshift
NC_002506	1036512	1	-	G	95.7	267	279	Gene: Gene, CDS: hypothetical protein	Change, frameshift
NC_002506	1045002	1	-	G	94.3	215	228	Gene: Gene	
NC_002506	1060870	1	-	C	94.9	168	177	Gene: Gene	
NC_002506	1060888	1	-	C	91.7	133	145	Gene: Gene	
NC_002506	1062459	1	-	G	96.2	176	183	Gene: Gene, CDS: hypothetical protein	Change, frameshift
AB374228	4914	1	A	-	74.2	196	264		
AB374228	5689	1	-	G	92.8	206	222	Gene: *kp9*, CDS: *kp9*	Change, frameshift
AB374228	7865	1	-	C	80.5	99	123	Gene: *rep*, CDS: *rep*	Change, frameshift
AB374228	17096	1	-	A	94.5	224	237	Gene: *kp24*, CDS: *kp24*	Change, frameshift
AB374228	19381	1	-	G	98.7	74	75	Gene: *kp28*, CDS: *kp28*	Change, frameshift
AB374228	22704	2	--	GC	82.2	176	214	Gene: *kp36*, CDS: *kp36*	Change, frameshift

*IUPAC notation
**Frequency is equal to number of SNP reads (count) divided by the total reads (coverage)