Supplementary Information for Turbelidze et al.

Methods to Sequence and assignment of backbone ORF SNP sets.

Genomic DNA was isolated from the study isolates and indexed fragment libraries were constructed for sequencing on the Illumina MiSeq platform. Pair-end reads of 150bp were generated to a depth of ca. 100x sequence coverage and each genome was assembled with Velvet (v1.1.04-64). All assemblies passed our in-house contamination screen and satisfied the metrics for a high quality draft sequence defined by the NIH Human Microbiome Project (1). While variant calling was read-based (below), the evaluation of assemblies provided assurance that there was no significant contamination or other quality issues in the data.

Pair-end MiSeq reads of isolate was mapped to the reference genome Sakai (gi:15829254, NC002695.1) using BWA (version 0.5.9 (2) and SNPs were detected using Samtools (version 0.1.12) (3). A minimum read depth of 10 was required for each SNP detected. Nucmer (an algorithm from MUMmer v3.19) was applied to align the assembled contigs of each strain to the backbone ORFs of the reference Sakai genome and to generate primary SNPs calls. Only the SNPs detected by both Samtools and Nucmer and that were in the backbone ORFS of Sakai were further analyzed. SNPs were also identified for strains 86-24, 87-14, 493-89, TB182A and TW14359 by comparing their backbone ORFs to the backbone ORFs of Sakai utilizing Nucmer. Further comparison of the SNPs between the strains was performed in R (4). ANNOVAR was used appropriately to annotate the final SNPs (5).

SNPs were validated by orthogonal sequencing using Life Technologies Ion Personal Genome Machine Sequencer. Life Technologies kits were used for library construction (Ion Fragment Library Kit, size selected for 390 base pairs +/- 5%), template preparation (Ion OneTouch 200 Template Kit v2 DL), and sequencing (Ion PGM 300 Sequencing Kit, 640 flows on a 318 chip). Sixteen samples were prepared and sequenced as replicates over two runs. Library construction methodologies were identical for both runs. Subsequent alignment and variant calling, and detection of SNPs and insertions and deletions, were performed on an individual sample basis, in a parallel approach. One path utilized Life Technologies Torrent Suite (6) software and another, independent, evaluation was performed with the Burrows-Wheeler Aligner (2, 7) package and SAMtools (www.samtools.sourceforge.net) (3). The Torrent Suite analysis pipeline, which includes the Torrent Variant Caller, produces a variant report using settings adjusted to match the error model (i.e., homopolymer errors) in Ion Torrent data. In both approaches, all samples were aligned to *E. coli* O157:H7 strain Sakai, complete genome to call variants.

For the purposes of this paper, radial SNPs were sought among each of the 3,442,673 nucleotides in the 3,586 Subgroup C Backbone ORFs (radial SNPs and ORFs defined in Tables S5 and S7 in reference (8)). Backbone ORF SNP sets were defined as sets of SNPs that are not present in strain TW14359, (8, 9) but present in one or more study isolates, except the isolate from case M23. In the first round of sequencing, we used a pilot set of isolates to determine if isolates from two exposed cases were isogenic and differed from isolates from cases in Webster County, Missouri, and Washington State. Next, we sequenced additional isolates, including isolates from cases who reported definite, possible, or no exposure to the incriminated vehicle or outlet. SNP analysis was performed by authors (H.G. and P.I.T.) who did not know the epidemiologic history or the origin of the isolates, except for the two Washington State *E. coli* O157:H7. The code for the isolates and their epidemiologic data was held by a third author (G.T.), and not broken until after the phylogenetic relationship of the isolates was constructed by P.I.T. (Figure 1).

We characterized SNPs in each of the newly sequenced *E. coli* O157:H7 as belonging to terminal or proximal backbone ORF SNP sets. Terminal backbone ORF SNP sets are defined as sets of one or more SNPs that by themselves differentiate that group from other groups of sequenced isolates in the TW14359 lineage. Proximal backbone ORF SNP sets are defined as sets of one or more SNPs that are shared by isolates that each possess additional and different terminal backbone ORF SNP sets. One SNP set in this study was a proximal (for one set of isolates) and a terminal (for three additional isolates) backbone ORF SNP set.

All SNPs (except those from the isolate from case M23) that were not present in strain TW14359 were amplified by PCR for confirmatory (Sanger) sequencing using primers in Table S1 and published conditions (8).

1. **Human Microbiome Jumpstart Reference Strains C, Nelson KE, Weinstock GM, Highlander SK, Worley KC, Creasy HH, Wortman JR, Rusch DB, Mitreva M, Sodergren E, Chinwalla AT, Feldgarden M, Gevers D, Haas BJ, Madupu R, Ward DV, Birren BW, Gibbs RA, Methe B, Petrosino JF, Strausberg RL, Sutton GG, White OR, Wilson RK, Durkin S, Giglio MG, Gujja S, Howarth C, Kodira CD, Kyrpides N, Mehta T, Muzny DM, Pearson M, Pepin K, Pati A, Qin X, Yandava C, Zeng Q, Zhang L, Berlin AM, Chen L, Hepburn TA, Johnson J, McCorrison J, Miller J, Minx P, Nusbaum C, Russ C, Sykes SM, Tomlinson CM, Young S, Warren WC, Badger J, Crabtree J, Markowitz VM, Orvis J, Cree A, Ferriera S, Fulton LL, Fulton RS, Gillis M, Hemphill LD, Joshi V, Kovar C, Torralba M, Wetterstrand KA, Abouellleil A, Wollam AM, Buhay CJ, Ding Y, Dugan S, FitzGerald MG, Holder M, Hostetler J, Clifton SW, Allen-Vercoe E, Earl AM, Farmer CN, Liolios K, Surette MG, Xu Q, Pohl C, Wilczek-Boney K, and Zhu D.** 2010. A catalog of reference genomes from the human microbiome. Science. **328:**994-999.
2. **Li H, and Durbin R.** 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. **25:**1754-1760.
3. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, and Genome Project Data Processing S.** 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. **25:**2078-2079.
4. **Team DC.** 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
5. **Wang K, Li M, and Hakonarson H.** 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. **38:**e164.
6. **Feng P, Lampel KA, Karch H, and Whittam TS.** 1998. Genotypic and phenotypic changes in the emergence of Escherichia coli O157:H7. The Journal of infectious diseases. **177:**1750-1753.
7. **Eppinger M, Daugherty S, Agrawal S, Galens K, Sengamalay N, Sadzewicz L, Tallon L, Cebula TA, Mammel MK, Feng P, Soderlund R, Tarr PI, Debroy C, Dudley EG, Fraser CM, and Ravel J.** 2013. Whole-Genome Draft Sequences of 26 Enterohemorrhagic Escherichia coli O157:H7 Strains. Genome Announc. **1:**e0013412.
8. **Leopold SR, Magrini V, Holt NJ, Shaikh N, Mardis ER, Cagno J, Ogura Y, Iguchi A, Hayashi T, Mellmann A, Karch H, Besser TE, Sawyer SA, Whittam TS, and Tarr PI.** 2009. A precise reconstruction of the emergence and constrained radiations of Escherichia coli O157 portrayed by backbone concatenomic analysis. Proc. Natl. Acad. Sci. U. S. A. **106:**8713-8718.
9. **Wendel AM, Johnson DH, Sharapov U, Grant J, Archer JR, Monson T, Koschmann C, and Davis JP.** 2009. Multistate outbreak of Escherichia coli O157:H7 infection associated with consumption of packaged spinach, August-September 2006: the Wisconsin investigation. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. **48:**1079-1086.

Table S1. SNP Details.
Section A. SNPs previously identified in strain TW14359 in Leopold, et al, Proc Natl Acad Sci U S A. 2009;106:8713-8, and deposited in GenBank under accession number GENBANK: EU892409.1.

Backbone ORF SNP set number	Locus	Amino Acid change, if not synonymous	Study strains in which SNP found	start and stop coordinates (based on *E.coli* O157 Sakai)	ORF	Full name of ORF and comments
1	6674GA	H/Y	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	6546/7976	ECs7	putative inner membrane transport protein
1	14317GA	A/T	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	14185/15315	ECs15	DnaJ protein
1	430790CT		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	430715/431524	ECs405	2-keto-4-pentenoate hydratase
1	479657CT		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	479123/479812	ECs449	Positive response regulator for pho regulon PhoB
1	556909CT	R/C	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	556777/557154	ECs521	putative gene 58
1	656910AG		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	656307/656939	ECs597	fimbrial Z protein
1	661692CT		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	659139/662111	ECs600	bacteriophage N4 adsorption protein NfrA
1	723747CT		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	723442/724680	ECs647	putative oxidoreductase
1	787375AT		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	786116/787780	ECs710	glutamine tRNA synthetase
1	866365CT		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	865949/866992	ECs778	quinolinate synthetase A protein
1	941648CG	T/S	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	941413/941658	ECs862	molybdopterin biosynthesis protein D chain
1	946192CT	A/T	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	946103/946864	ECs868	hypothetical protein
1	988913GA	A/V	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	988714/989463	ECs904	molybdopterin biosynthesis MoeB protein
1	1049196GA	R/H	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	1047548/1049206	ECs962	hypothetical protein
1	1151284AG		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	1151201/1153363	ECs1044	hypothetical protein
1	1226849GA		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	1224980/1227076	ECs1140	hypothetical protein
1	1725872CT		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	1725087/1728830	ECs1729	nitrate reductase 1 alpha subunit
1	1837132GT	G/C	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	1836538/1839213	ECs1849	aconitate hydrase 1
1	1902934AG	Y/C	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	1902564/1904198	ECs1911	putative transport periplasmic protein
1	2002772GA		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	1998948/2002793	ECs2015	ATP-dependent helicase HrpA
1	2055091TC	K/E	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	2054201/2055094	ECs2067	hypothetical protein
1	2118004GA		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	2117977/2122008	ECs2117	hypothetical protein
1	2346918CT	P/L	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	2346842/2348446	ECs2375	hypothetical protein
1	2410770GA		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	2410648/2412039	ECs2435	part of a kinase
1	2416801TC	D/G	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	2416659/2418011	ECs2440	phospho-beta-glucosidase
1	2423779TA	L/Q	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	2423028/2423915	ECs2447	putative excinuclease subunit
1	2452510GT		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	2452172/2452831	ECs2475	hypothetical protein
1	2452556TC	F/L	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	2452172/2452831	ECs2475	hypothetical protein
1	2473176GA	R/H	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	2472122/2473465	ECs2494	hypothetical protein
1	2508129GT	F/L	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	2508081/2508890	ECs2532	putative enzyme
1	2512684CT		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	2511995/2513479	ECs2538	putative transport protein
1	2544711GA		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	2544493/2545248	ECs2568	putative ATP-binding component of a transport system
1	2769998GA	P/S	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	2769387/2770400	ECs2828	regulator of length of O-antigen component of lipopolysaccharide chains

1	2792291GT	R/S	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	2791287/2792507	ECs2849	putative colanic acid biosynthesis glycosyl transferase
1	2799943GA		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	2799895/2801118	ECs2855	putative colanic biosynthesis glycosyl transferase
1	2804346GA	A/V	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	2804279/2805025	ECs2860	putative colanic acid biosynthesis glycosyl transferase
1	2807112TC	S/G	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	2806228/2807445	ECs2862	putative glycosyl transferase
1	2887014CT	A/V	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	2886917/2887648	ECs2929	hypothetical protein
1	2887845AC	T/P	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	2887641/2888777	ECs2930	hypothetical protein
1	2973134TG	S/R	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	2972763/2974283	ECs3041	ATP-binding component of methyl-galactoside transport system
1	3053617CT	G/S	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3053138/3053914	ECs3108	hypothetical protein
1	3069642GA	T/I	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3066200/3069952	ECs3116	putative ATP-binding component of a transport system
1	3069841CT	D/N	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3066200/3069952	ECs3116	putative ATP-binding component of a transport system
1	3095943CT	R/H	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3095531/3096133	ECs3140	aluminum-inducible protein
1	3098294TC		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3097584/3098552	ECs3142	putative sugar transferase
1	3164771CA	R/S	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3164303/3165298	ECs3205	cell division protein
1	3165502GA		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3165295/3166473	ECs3206	putative transport protein
1	3222307CT	E/K	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3221900/3222535	ECs3255	hypothetical protein
1	3281701GA	R/C	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3280421/3281782	ECs3312	ethanolamine ammonia-lyase heavy chain
1	3286717TC	I/V	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3286544/3287947	ECs3317	ethanolamine utilization protein EutE
1	3293290GA		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3292222/3294501	ECs3325	putative multimodular enzyme
1	3296574CT	A/V	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3295760/3297763	ECs3327	transketolase 2 isozyme
1	3322711CT	P/S	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3322357/3323007	ECs3347	hydrogenase 4 membrane subunit
1	3323566GA		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3323012/3324592	ECs3348	hydrogenase 4 membrane subunit
1	3384416GC		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3383943/3384746	ECs3399	extragenic suppressor protein SuhB
1	3439915GA		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3439397/3440119	ECs3449	hypothetical protein
1	3443184GC	V/L	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3442923/3444281	ECs3452	phosphatidylserine synthase
1	3529080TC		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3528967/3530169	ECs3540	ATP-binding component of glycine betaine / proline transport system
1	3535377GT	Q/H	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3534928/3535458	ECs3546	regulator of plasmid mcrB operon
1	3656053GT	D/E	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3655441/3656088	ECs3660	L-fuculose-1-phosphate aldolase
1	3672410GT	A/S	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3671642/3672973	ECs3675	N-acetylglutamate synthase
1	3697431CT		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3696255/3698414	ECs3693	2-acyl-glycerophospho-ethanolamine acyltransferase
1	3763970AG	T/A	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3763916/3765235	ECs3756	hypothetical protein
1	3838445CT	R/C	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3838187/3838780	ECs3830	putative ribosomal protein
1	3947080CT		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	3946901/3947506	ECs3945	L-tartrate dehydratase subunit B
1	4007053GA		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	4006883/4007227	ECs3999	threonine dehydratase operon activator protein
1	4055003GA	A/V	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	4054330/4055274	ECs4047	tRNA pseudouridine 5S synthase
1	4058766GA	A/V	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	4058537/4060024	ECs4050	transcription termination-antitermination factor NusA
1	4062632GA	P/S	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	4062486/4064129	ECs4053	putative alkaline phosphatase I
1	4151813GA	V/I	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	4149992/4151878	ECs4138	putative integral transmembrane protein
1	4198837GA		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	4198735/4200513	ECs4201	NEM-activable K+/H+ antiporter
1	4223005GC		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	4222256/4223083	ECs4223	hypothetical protein
1	4253565GA		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	4252582/4254903	ECs4251	ferrous iron transport protein B
1	4273017GA		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	4272795/4273553	ECs4266	repressor of the glp operon

1	4305379CA	A/S	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	4305029/4305772	ECs4295	glycerophosphodiester phosphodiesterase
1	4362446GA	P/S	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	4361280/4363964	ECs4359	putative ATP-binding component of a transport system
1	4367684GA	P/L	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	4367329/4368939	ECs4362	hypothetical protein
1	4413603GA		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	4413429/4414031	ECs4400	putative regulator
1	4487234GT		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	4487192/4487665	ECs4456	hypothetical protein
1	4705476GA		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	4705284/4707113	ECs4671	N-acetyl glucosamine-1-phosphate uridyltransferase
1	4773029GA		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	4772007/4773359	ECs4727	probable UDP-N-acetyl-D-mannosaminuronic acid transferase
1	4778904GA		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	4777992/4779647	ECs4731	HemY protein
1	4786550CT		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	4784811/4787357	ECs4736	hypothetical protein
1	4971458CT	G/S	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	4971435/4972535	ECs4896	outer membrane vitamin B12 receptor protein BtuB
1	4988108TG		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	4987896/4988600	ECs4907	50S ribosomal subunit protein L10
1	4992513CT		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	4990141/4994169	ECs4910	RNA polymerase beta prime subunit
1	5137547AC		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	5136401/5138050	ECs5049	putative transport protein
1	5141169GA	R/H	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	5140913/5142349	ECs5052	periplasmic cytochrome c(552)
1	5259986GT	R/L	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	5259916/5260341	ECs5154	hypothetical protein
1	5270403GA	P/L	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	5270168/5270596	ECs5165	hypothetical protein
1	5337294GA		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	5336379/5339234	ECs5235	valine tRNA synthetase
1	5343889CT		WC1, WC2, Wa1, Wa2, M1-22, K1, K2	5342906/5343991	ECs5239	hypothetical protein
1	5460247AC	V/G	WC1, WC2, Wa1, Wa2, M1-22, K1, K2	5459982/5460308	ECs5323	hypothetical protein
5	343210GA	H/Y	WA1, WA2, M1-22, K1, K2	343154/343744	ECs324	putative regulator
5	1226896GA	R/W	WA1, WA2, M1-22, K1, K2	1224980/1227076	ECs1140	hypothetical protein
5	3285750GA		WA1, WA2, M1-22, K1, K2	3285697/3286533	ECs3316	ethanolamine utilization protein EutJ
5	3319230GA	G/S	WA1, WA2, M1-22, K1, K2	3317913/3319931	ECs3344	hydrogenase 4 membrane subunit
5	4352057GA		WA1, WA2, M1-22, K1, K2	4351620/4352384	ECs4346	ATP-binding protein of nickel transport system
5	5396102CT	P/S	WA1, WA2, M1-22, K1, K2	5395865/5396467	ECs5271	type 1 fimbriae regulatory recombinase protein FimB

Section B. Details regarding SNPs identified in this study.

Backbone ORF SNP set number	Locus	Amino Acid change, if not synonymous	Study strains in which SNP found	Forward / Reverse Primer (5'-3') used for confirmation of SNP by Sanger sequencing	start and stop coordinates (based on E.coli O157 Sakai)	ORF	Full name of ORF and comments
2	801508CT	R/H	WC1, WC2	AGT AAG GTT CCC GAT CAC CAG/ TAC AGA TTC AGG GAT GCG TGG	800157/ 802841	ECs0723	sensor for high-affinity potassium transport system
2	1245136CT	P/L	WC1, WC2	AAC CGT ATC AGT TGC ATG ACC/ AGG TAG CGA TGA AGT GTT AGT C	1243938/ 1245179	ECs1158	periplasmic glucose-1-phosphatase
2	1908590GT	S/Y	WC1, WC2	AAT TGC TCA TCG CAA ATC ACC/ TGA CCG ATA TCC TGG AAA GTT C	1908406/ 1909098	ECs1917	hypothetical protein
2	2140505GT	S/I	WC1, WC2	ACG TCT TTA AAC ATC AAC TCA GCC/	2140174/	ECs2134	hypothetical protein

				AAC GTC AGG GTT GAT CAT TGG	2140962		
2	2616879GA		WC1, WC2	AGA GTT GAT GTA GTT GTG TTC CAG/ AGA ACA GCT CAA TCT CAG TCC	2615145/ 2616911	ECs2651	excinuclease ABC subunit C
2	4349372GC		WC1, WC2	TGA TAA CGC GTT GTA TGG CAC/ AAT CAT GCC AAA ACG ACC GTC	4348272/ 4349846	ECs4343	periplasmic binding protein for nickel
3	484705TG	D/E	WC1	ATG TTA AAT GCA TGG CAC CTG/ AGA CAT GAT CCT GTT CAG CTT C	484496/ 486313	ECs0453	maltodextrin glucosidase
3	611124CT	H/Y	WC1	AGC GAT ATT GAT TCG CGA CAG/ ATC TGC GTA AAG ATC TCA CCG	610353/ 611159	ECs0554	putative metal resistance protein
3	747134TC	S/G	WC1	ATT CAA GCT GTG ACT GAG TGC/ TAG ATG CTT TCG GTG CCA TC	746559/ 747671	ECs0672	rod shape-determining membrane protein
3	754785TG	D/E	WC1	ACC ACC AAT GTA GAT ATC CAC C/ TGT TCA ACT CTG GCG AGT TC	753559/ 756141	ECs0680	leucine tRNA synthetase
3	1463295CA	A/E	WC1	TAG GTT GCG TTA ACA ACC AAG T/ ACT GTC TCT GCA CTA CAA TTG C	1463090/ 1463545	ECs1419	minor curlin subunit precursor CsgB
3	2332345CA	A/E	WC1	ATT GCT GCT GGA ACG TCA G/ TAT CCG TTG GCT GAT TTG TTG	2328813/ 2333429	ECs2362	probable ATP-dependent helicase Lhr
3	2416915GT	A/E	WC1	AAG CGT TCC ATA ATC AGC CTC/ AGA GAT GCT GGC TAT TGA AAT G	2416659/ 2418011	ECs2440	phospho-beta-glucosidase
3	3173905GA	A/V	WC1	AGA ATT TTG ACC TGA TCT GGC TC/ TGT CGA TGC GGA AGA TAT GTC	3173871/ 3175136	ECs3215	putative adenine-specific methylase
3	3209936CT		WC1	ACA GGG TCA GTC CTG GTA G/ TTC TAA CGT CGC AAT ACC GAC	3208775/ 3210313	ECs3246	multidrug resistance protein Y
3	3421037GA	H/Y	WC1	ATG TCA GAG ATT GCA AGT TGC/ ATT TGG AAA ACG CAC CAC CAG	3421002/ 3421382	ECs3429	CoA:apo-[acyl-carrier-protein] pantetheinephosphotransferase
3	3540151CT		WC1	AGA TCG CTC AAA CGA AGA GAG/ TGG TGA TAT TGA ACA TAT GCT GAC	3539041/ 3540597	ECs3550	gamma-glutamate-cysteine ligase
3	3779900GA	G/S	WC1	ATG TGG AGC AAT TTA CGT CTG/ ATC TCT TGT CCG GTA TAG CAG	3779462/ 3780442	ECs3770	hypothetical protein
3	4418769CA	A/S	WC1	TTA AAC GTC CCA CGC ACA TAC/ TGT GGA ACT GAT GCA GAA GAC	4418428/ 4420503	ECs4404	hypothetical protein
3	4576291GA	A/T	WC1	ACG TTG ATC CTT GAT GAT GCG/ TCA ACG GTA ATT CAG GCC ATC	4575004/ 4576713	ECs4531	hypothetical protein
3	4781387TG	I/L	WC1	TAG ATA GTT CAC CAC CGG TTG/ ACA ACG GAA GCT AAA CCT GC	4780326/ 4781522	ECs4732	HemY protein
3	5321302GT		WC1	ATG CTG AAC ATG CCA AAG TCC/ AAC GTG GGT GAT TAC TAT CAA GC	5320297/ 5321718	ECs5217	trehalose specific PTS system enzyme II
4	207231GA	A/T	WC2	TTT GGT GAA CTA AAG CCA GAT G/ ATT CCA TCC AGC AAA TGA ACC	206691/ 207839	ECs0184	lipid A-disaccharide synthase
4	653171CT		WC2	ATG GTG TTG GAT TTC GTG GTG/ TGT AAT TGG CAG ATA ATT GCG TC	652149/ 654758	ECs0594	putative outer membrane protein
4	1067276CA		WC2	ATC ACT CGC TTT GAA CTT AAC C/	1064301/	ECs0975	cell division protein

				TCA GTT TAT AAC GAC GCT CCA TC	1068329		
4	1235736AC	V/G	WC2	TCG GTG TCT GAT AAT GTC AGC/ TAC ACT CAG GAC AAC TGC TAT C	1235615/ 1236307	ECs1150	response transcriptional regulator for torA
4	1535509AT	S/C	WC2	ATC GTA AAG GCG AGC TGT TC/ ATA GCG TTT GCA CAG GTT GTC	1534552/ 1535751	ECs1494	hypothetical protein
4	1867688GA		WC2	TCA AAC TGC AAC AGC TCA TCA T/ AAC GTG CTT TCT GAC ACA GAA G	1867423/ 1868919	ECs1874	putative glutamine synthetase
4	2464012GA	A/V	WC2	AAC GAA TGG ATT CTT CAC TCA C/ ACA AAT CAA TCC TGT GCC TAA G	2463618/ 2464031	ECs2487	hypothetical protein
4	3157658GA	P/L	WC2	AGA AAG TTG CTT CAA CTC ACC/ ATG AGC CTG ATA TGA TGC CAG	3157375/ 3158037	ECs3198	putative lipoprotein
4	3319266GA	G/S	WC2	TTT GCA TAC CCA CGA CAT GG/ TGA CCT GTG TAT GAA ACG TGC	3317913/ 3319931	ECs3344	hydrogenase 4 membrane subunit
4	3342659CA	T/K	WC2	ATC ACC GTA AAG CGA TTA AAC TC/ TGA TAC GCT GAA TGA AAT CGA AC	3341257/ 3342798	ECs3364	exopolyphosphatase
4	4204517GA	G/D	WC2	ACG AAA CGG TCA ATA TCA TGC C/ AAC AGT ACA TCA GTG GGT TCT G	4204396/ 4205265	ECs4206	probable phosphoribulokinase
4	5015574CT	V/I	WC2	AAT CAC TTC AAC AAA CTG ACG AG/ AGC ATG ATT GCC AAC TAC TTC	5014747/ 5016336	ECs4929	Phosphoribosylaminoimidazolecarboxamideformyltransferase
4	5298260TC		WC2	ATT GAA TTT GAT CTC CGT CCA C/ TCT TTC GAT TCG TAC TCA TCG C	5297862/ 5299595	ECs5198	hypothetical protein
4	5422614AC	L/R	WC2	ACC ATA CAG AAT CGC CAG TAA C/ TCG CAA CCG GAT TAT TAA TCC TC	5421911/ 5423089	ECs5293	putative transport protein
6	554623GA	V/I	M1-22, K1, K2, Wa1, WA2	TGG AAA GTG CTG GAG AAG AAC/ AGA GCA AGT TCC AGA TGA TCA C	552601/ 555963	ECs0518	mechanosensitive channel protein
6	3615468GA	G/E	M1-22, K1, K2, Wa1, WA2	ACA ACA ATA CAC ACC TCA GGT G/ TTA CAT GGT GCA AGC ATC TGC	3615328/ 3615897	ECs3625	hypothetical protein
6	5196793CG	T/S	M1-22, K1, K2, Wa1, Wa2	TGA CTG ATT TCG AGG TCT GTA AG/ TAT GGT TGA TAC CTG CGC AAC	5196360/ 5197715	ECs5101	alpha-galactosidase
7	3840700GA		Wa1, Wa2	AAA CTC ATC ACG CTG TTG GTG/ TTA CGA ACT GCT CGA ACA GAC	3840152/ 3841159	ECs3832	putative alpha helix chain
7	4805937CT	P/L	Wa1, Wa2	TGA TGT CTT AAT CAT CGA CCA TCG/ AGA AGC GTA AGT TAC GTC GAT ATC	4805480/ 4806502	ECs4755	lysophospholipase L(2)
8	452998CT	H/Y	Wa2	TAA CGT TGA TCA GTT CGA TGC/ TTC AGC CTG ATA GGT GTA TGC	451255/ 454197	ECs0424	putative flagellin structural protein
8	3621144GA	V/I	Wa2	ACT TCC GCG ATG ATC ATT ACC/ AGA ATG GCG TAG ATG GTG AAG	3620709/ 3621986	ECs3631	putative transport protein
8	3668504CT		Wa2	ATC AGG CCT ACA TAA CTC CTG/ TGC CTC TGA TCG TTC CAT TAT C	3668481/ 3669578	ECs3673	membrane-bound lytic murein transglycosylase A
8	3905431CT	H/Y	Wa2	TTC TCC AGA CGA TAG TTC AGT C/ AAA TCT ACG AGA ACG GAC GTG	3904066/ 3906324	ECs3903	DNA topoisomerase IV subunit A
9	219932GA	T/I	Wa1	TTC GAC CAC ACT GCT TAA CTG/	219511/	ECs0195	hypothetical protein

				TTC CAG CAT CAA AGC TCT TTC	220335		
9	838912AC	Y/S	Wa1	ACG TTC AAC GAA GTC AAC ATG/ AGA TCT TCA ACC GTC AGC TTG	838167/ 839384	ECs0752	2-oxoglutarate dehydrogenase dihydrolipoyltranssuccinase E2 component
9	961676GT	E/*	Wa1	TTG CCT GAT TGA CGT TAA CG/ TCA TTG ATC CTA TGT TAC CCT TG	960791/ 961876	ECs0879	putative dehydrogenase
9	2041566CA		Wa1	AGT TGA TGA AGT ACA TCC AGC/ TGG TGA TAA CTA TCA TCG CCA G	2040246/ 2041796	ECs2057	L-asparagine permease
9	3017687CT	P/S	Wa1	TCA TCT ACG GTG GCA AAG AAG/ TGT GTT ATG AAT GCC AGA TGC	3017447/ 3017731	ECs3077	50S ribosomal subunit protein L25
9	4111029GA	G/D	Wa1	ATA TCA ATC ACG AAC AGG CAG/ TGC GAT ATT CGG AAT AAG ACT G	4110116/ 4111480	ECs4100	putative C4-dicarboxylate carrier protein
9	4863308CT		Wa1	TTC AAC AGT TAT TTC TCC ACC G/ ACC ACC CTG TTT GAG ATG AAG	4863038/ 4865074	ECs4800	putative glycosidase
10	6456AC		M1-22, K1, K2	AGT TTG TCG CTG ATA CGC ATC/ TGT GGT TCA TAC CAT TGC CAG	5700/ 6476	ECs0006	hypothetical protein
10	2478628GA	E/K	M1-22, K1, K2	AGT GAC GTG ACA AAT CAT GCG/ TCA CAA GCT ACG TAG CCA GTC	2478493/ 2478861	ECs2501	hypothetical protein
10	3243126CA		M1-22, K1, K2	ACT TCA ACA AGG ATT CCA GCA C/ AGT GAA CTG CTG TTA GTG CTG	3242427/ 3244655	ECs3273	hypothetical protein
10	3517964GA	R/H	M1-22, K1, K2	ATA GGG ATT ATC TGT ACC GTG C/ AGA CTC ACT AAC ATC GCT TCC	3517843/ 3518523	ECs3525	putative transcriptional regulator
10	3780989GA		M1-22, K1, K2	AGG CTC CAG ATA ACA ATC ATC AAC/ AGA GGT ATA CAT GTG CGT GTA C	3780519/ 3781178	ECs3771	putative oxidoreductase
10	4492725GA		M1-22, K1, K2	ATA CGA CGT GCT CTC CTA TCG/ TGA CGA TTA TGC CAT GCA ACA G	4492461/ 4493201	ECs4460	putative outer membrane protein
11	3894065CT	P/L	M18	ACG ATT CAG GAG TAC AAT AGG C/ TGG TGA TAT TTA GCA GGA AGC TG	3893920/ 3894579	ECs3893	hypothetical protein
12	2617146GA	T/M	K2	ACA TAG ATA ACC GTA CCA CCA G/ TGC GAT CCG TTC TGT CTA TTC	2616974/ 2617630	ECs2652	putative 2-component transcriptional regulator
12	3587723CT	D/N	K2	TGG TTC ATG GTC CAG CTT ATG/ TCA GCT TTA CCT TGG TGA GAT TG	3587082/ 3588074	ECs3595	sigma factor RpoS
13	3838874GA		M19	ACT GGT GGC TTT GGT TAT GAT C/ ACG CTG ATA ATC AAC AAA GCG	3838773/ 3839909	ECs3831	putative oxidase