Whole genome sequence analysis of 91 *Salmonella* Enteritidis isolates from mice caught on poultry farms in the mid 1990s

Jean Guard a, Guojie Cao b, Yan Luo b, Joseph D. Baugher b, Sherrill Davison c, Kuan Yao b, Maria Hoffmann b, Guodong Zhang b, Rebecca Bell b, Jie Zheng b, Eric Brown b, Marc Allard b

US National Poultry Research Center, US Department of Agriculture, Athens, GA, 30605 a, Center for Food Safety and Applied Nutrition, Food and Drug Administration, College Park, MD, 20740 b, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA, 19104 c

Author for Correspondence: Jean Guard, US National Poultry Research Center, US Department of Agriculture, 950 College Station Road, Athens, GA, 30605; Tel: 706-546-3446; Email: Jean.Guard@ars.usda.gov
Abstract

*Salmonella enterica* serovar Enteritidis (SE), the most commonly reported serovar of human salmonellosis, has been frequently associated with poultry farms, eggs and egg products. Mice are known vectors of SE contamination in these facilities. The objective of this study was to use whole-genome sequencing (WGS) to analyze SE from mice obtained at poultry farms in Pennsylvania. Documenting pathogen diversity can identify reliable biomarkers for rapid detection and speed up outbreak investigations. We sequenced 91 SE isolates from 83 mice (62 spleen isolates, 29 intestinal isolates) caught at 15 poultry farms between 1995-1998 using an Illumina NextSeq 500. We identified 742 single nucleotide polymorphisms (SNPs) capable of distinguishing each isolate from one another. Isolates were divided into two major clades: there were more SNPs differences within Clade B than counterparts in Clade A. All isolates containing antimicrobial resistance genes belong to Subgroup B2. Clade-defining SNPs provided biomarkers distinguishing isolates from 12 individual subgroups, which were separated by farm location or year of collection. Nonsynonymous changes from the clade-defining SNPs proffered a better understanding of possible genetic variations among these isolates. For a broader view of SE diversity, we included data from NCBI Pathogen Detection Isolates Browser, in which subgroups in Clade B formed new SNP Clusters.

Importance

WGS and SNPs analyses are excellent and powerful tools for investigating SE phylogenies. Identifying the evolutionary relationships among SE isolates from mouse, poultry, environmental, and clinical isolates, along with patterns of genetic diversity,
advances understanding of SE and the role mice may play in SE contamination and spread among poultry population. Our data was able to identify SE isolates from different farms or years of collection. Moreover, the annotations of clade-defining SNPs provided information about possible protein functions among these SE isolates from each subgroup. Clade-defining or farm-unique biomarkers were useful for rapid detection and outbreak investigations.

Keywords: *Salmonella*, WGS, phylogenetics, mouse, poultry, egg
Introduction

Salmonella enterica serovar Enteritidis (SE) is a long-standing public health concern in the US (1); salmonellosis can result in hospitalization or death of infants, the elderly, and those with compromised immune systems (2, 3). This pathogen has been strongly associated with poultry farms, eggs, and egg products (4, 5). In 2010, SE linked to shell eggs resulted in an outbreak requiring the recall of a half billion eggs (https://www.cdc.gov/salmonella/2010/shell-eggs-12-2-10.html) (6).

One of the challenges in resolving foodborne outbreaks associated with SE is the extreme genomic homogeneity within a specific geographic location or ecology system and its broad host range (6, 7). Mice are important biological vehicles contributing to SE dissemination and amplification in chicken houses, especially among laying hens (8, 9). In fact, SE has been strongly correlated with rodent activity; chickens in caged housing where mice are present are more likely to carry SE (10). Understanding the evolutionary relationships among SE isolates from mice, poultry, environmental surfaces, and clinical cases is important both for outbreak investigations and for identifying strains with genetic markers for virulence or capacity for rapid host adaptation, such as mutations in the mismatch repair gene mutS that can contribute to rapid evolution in immunocompromised hosts (11).

Whole genome sequencing (WGS) methods have identified variations across otherwise indistinguishable isolates from eggs and egg products (6, 12), SE associated with reptile feeder mice (13), S. Montevideo from red and black pepper (14). Genome-wide single nucleotide polymorphisms (SNPs) detected by WGS are considered as the most valuable genetic markers for investigating the evolutionary relationships among SE
homogeneous isolates (1, 7, 15). Application of WGS have also been useful in other microorganisms, including *E. coli* (16), *Vibrio cholera* (17), and *Staphylococcus aureus* (18).

Importantly, WGS can be also applied to historic isolates, some of which have been stored for decades. Data from those historic isolates should allow us to understand the origin and persistence of important traits. In this current project, we sequenced 91 SE isolated from 82 mice at poultry farms during the 1990s, which lets us to compare both site and host-adaptions with those of isolates from more recent sampling. Documenting these genomes and fitting them into large-scale phylogeny projects such as GenomeTrakr (https://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS/ucm363134.htm) and NCBI Pathogen Detection Isolates Browser (https://www.ncbi.nlm.nih.gov/pathogens/) will refine our understanding of SE contamination and spread in poultry facilities (19). Further, identifying and characterizing biomarkers can facilitate the development of rapid and reliable tests that could guide appropriate interventions during future outbreaks.
Materials and Methods

Bacterial isolates

Ninety-one SE isolates from mouse spleens (n=62) and intestines (n=29), collected from 15 poultry farms in Pennsylvania during 1995-1998, are listed in Table 1. Among these isolates, eight pairs were isolated from the spleen and intestine of the same mouse; these were designated as m1 through m8. These isolates are archived under Bioproject Number PRJNA186035 (https://www.ncbi.nlm.nih.gov/bioproject/186035).

Whole genome sequencing and assembly

Genomic DNA was extracted after incubation of culture for 16 hours at 37 ºC in Trypticase Soy Broth (TSB) using the DNeasy Blood and Tissue Kit (Qiagen Inc, Valencia, CA). Concentrations of DNA were measured using a Qubit 3.0 fluorometer (Life Technologies, MD). Libraries were prepared according to Nextera XT protocols and sequenced on the Illumina NextSeq 500 (Illumina, San Diego, CA) using NextSeq 500/550 High Output Kit v2 (300 cycles). Raw reads were assembled de novo using SPAdes software v3.8.2 with default settings (20). We obtained chromosome draft genomes between 4.69M bps and 4.80M bps. These genomes were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (21).

We selected SE CFSAN051873 (spleen, 1996, Farm VIII) to serve as the reference genome, using the PacBio platform we obtained a fully closed genome for CFSAN051873 as follows (22). Genomic DNA was sheared into approximately 20-kb fragments using g-TUBE (Covaris, Inc., Woburn, MA). The library was prepared based on the 20-kb PacBio sample preparation protocol and sequenced using P6/C4 chemistry on four single-molecule real-time (SMRT) cells with a 240-min collection time.
The continuous long-read data were *de novo* assembled using the PacBio hierarchical genome assembly process (HGAP version 3.0) with default parameters (23). The assembled sequence was annotated using PGAP (21).

**Genomic and phylogenetic analysis**

The Fastq data from NextSeq runs were put into the Center for Food Safety and Applied Nutrition (CFSAN) SNP pipeline v0.8 to create a SNP matrix (24) with SE CFSAN051873 (CP_022003.1) as the reference genome. GARLI (Genetic Algorithm for Rapid Likelihood Inference: https://code.google.com/archive/p/garli/) v2.01 (25) was used to construct maximum-likelihood (ML) phylogenetic trees (ratematrix = 6rate; ratehetmodel = gamma). Multiple runs were performed (n=100) to ensure that results were consistent. To estimate support for each node, phylogenies were created for 1,000 bootstrap replicates of the data set from GARLI. Python program SumTrees was used to generate one consensus tree with bootstrap values at a 70% threshold (https://pythonhosted.org/DendroPy/programs/sumtrees.html) and FigTree v 1.4.3 was used to export the figures (http://tree.bio.ed.ac.uk/software/figtree/). NCBI Pathogen Detection Isolates Browser (https://www.ncbi.nlm.nih.gov/pathogens) was used to show phylogenetic relationship among SE isolates from broader ranges of geographical locations and sources. Custom script was used to identify clade-defining SNPs and Tool for Rapid Annotation of Microbial SNPs (TRAMS) tool to perform annotations on clade-defining SNPs (26). The pairwise distance matrix, shown as number of SNP differences among isolates, was calculated using MEGA7 with 1,000 bootstrap iterations (27).
**Results**

**Phylogenetic analysis**

**Overview**

We identified 742 SNPs and generated the maximum-likelihood phylogenetic tree arising from these SNPs, as depicted in Fig 1. Tree tips were marked using isolate name, source, year, farm, and NCBI Pathogen Detection Isolates Browser SNP Cluster. For example, CFSAN051866 was labeled as CFSAN051866_spleen_1996_FarmVII_SCA, which provides the following details: this bacterium was isolated from a mouse spleen in 1996, that mouse came from Farm VII, and the isolate fits within SNP Cluster A (28), which was designated according to the NCBI Pathogen Detection Isolates Browser (Table 1). Subgroup names and the number of clade-defining SNPs were labeled on the internal branches. For example, Subgroup B1 had the most clade-defining SNPs (179 SNPs), while Subgroup A5 had only 6 clade-defining SNPs.

**Phylogenetic Tree Construction**

We recognized two major clades: Clades A and B, which further subdivided into 12 subgroups: A1 to A8 and B1 to B4. It was notable that all isolates carrying antimicrobial resistance genes belonged exclusively to Subgroup B2. Moreover, isolates in each subgroup had varied ranges of SNP differences. The maximum SNP differences within Subgroups A1 and A5 were 33 (CFSAN063779 and CFSAN063803) and 27 (CFSAN063788 and CFSAN063792) SNPs, respectively, while the maximum number in Subgroup A3 was only 6 SNPs (CFSAN051856 and CFSAN051861). Subgroups A1 and A2 were the two largest subgroups,
containing 18 and 17 isolates, respectively. Subgroup B3 only contained two isolates, as the smallest group in the tree.

**Impact of Farm**

Not all subgroups in each clade showed the same pattern of geographic distribution, although some subgroups were exclusively comprised of isolates from a single farm. For example, all Subgroup A2 isolates were from Farm III and all Subgroup B1 isolates from Farm X. In contrast, Subgroups B3 only contained two from Farm I; A3 contained isolates from Farms V and VII.

Our phylogeny revealed that some isolates from different farms can be grouped together and were closely related: isolates in Subgroup A3 obtained from Farms V and VII with few to no SNP differences among them. For example, CFSAN051854 (CFSAN051854_spleen(m7)_1996_FarmV_SCA) and CFSAN051864 (CFSAN051864_spleen_1996_FarmVII_SCA) where zero SNP differences were observed (Table S1).

Isolates from some farms were only distantly related and, unsurprisingly, our phylogeny showed these belonging to different subgroups. For example, Subgroups B1 and B2 both contained isolates from Farm X, indicating that these were distantly related to the rest of the isolates from our sequencing.

There were several cases in which isolates from different farms were found to belong to the same subgroup: isolates from Subgroup B2, which contained 11 clade-defining SNPs, came from Farms VI, X, and XI. Isolates in Subgroup A3 were found at Farms V and VII, and there were only very small differences among their SNPs (Table S1).
Although isolates in Subgroups A1 and A5 were found at different farms, isolates from the same farm shared common ancestors. Specifically, all Subgroup A1 isolates were from Farm XII and XV, isolates from Farm XII formed a cluster and shared a common ancestor, and another common ancestor was shared by all isolates collected from Farm XV.

**Impact of Isolation Year**

Isolates in each subgroup were collected during the same year, with only two exceptions: A1 contained isolates from 1995 and 1996, and A5 contained isolates from 1997 and 1998. In Subgroup A1, isolates from 1995 were grouped together sharing common ancestor, which also applied to those from 1996 in Subgroup A1. In another case, all Subgroup A5 isolates were collected from 1998 except CFSAN063788, which was from 1997.

**Impact of Isolation Organ**

As expected, isolates from the same mouse appeared very closely related: SNP differences ranged from zero (m8) to two SNPs (m2). Most subgroups contained isolates from both organs. Although Subgroups A4, A8, B3, and B4 only contained isolates originating from spleens, our phylogenetic analyses did not reveal any organ-defining SNPs that could be reliably used to distinguish between SE isolates taken from spleens and those obtained from intestines.

**Pathogen Detection SNP Cluster analysis**

At the time of this research (Dec 7\textsuperscript{th}, 2017), the NCBI Pathogen Detection Isolates Browser (https://www.ncbi.nlm.nih.gov/pathogens) contained more than 94,000 *Salmonella enterica* genomes. At the time of our analysis, 86 of our
isolates fit into five existing Pathogen Detection SNP Clusters, as follows. All 68 isolates, but CFSAN063803, within our eight Clade A subgroups belonged to one single Pathogen Detection SNP Cluster, which was designated as SNP Cluster A (SCA, at the time designated as SCA PDS000002757.323) (28). CFSAN063803 did not fit within any of the established SNP Cluster at that time. The four subgroups we recognized as Clade B belonged to four different Pathogen Detection SNP Clusters, which were designated as SCB, SCC, SCD, and SCE, respectively.

The data from Pathogen Detection Isolates Browser matched our phylogenetic analysis. Among our sequenced isolates, some farms contained isolates that were distantly related according to Pathogen Detection Isolates Browser data. For example, isolates collected from mice at Farm I, which we identified as Subgroups A8, B3, and B4, were members of three existing Pathogen Detection SNP Clusters: SCA, SCB, and SCC, respectively.

Our isolates in SCA had been collected from mice at 12 different farms between 1995 and 1998. However, SCA also encompassed 5,468 genomes already in the Pathogen Detection Isolates Browser. This provides the opportunity to explore additional levels of relatedness across SE isolates, as well as identify patterns across multiple years. For example, in the Pathogen Detection phylogenetic tree, Subgroup A1 isolates from 1995 shared a common ancestor with SE NYVetLIRN-37 (Sequence Read Archive (SRA) number: SRR6107632), which was isolated from dust taken from a poultry coop at Massachusetts in April 2017 (https://www.ncbi.nlm.nih.gov/Structure/tree#!/tree/Salmonella/PDG00000002.1
Another example, SE WAPHL_SAL-A00192, which was isolated from an avian source from Washington in 2003, shared a common ancestor with Subgroup A5 isolates (https://www.ncbi.nlm.nih.gov/Structure/tree/#!/tree/Salmonella/PDG000000002.1). It was notable that isolates from egg yolk and chicken drag swab appeared closely related to isolates in Subgroups A4 and A6. For example, SE CRJJGF_00137 (egg yolk, 2002, US, SRR1686612) and SE OH-10-18938-5 (chicken drag swab, 2010, Ohio, SRR5278942) were closely related to CFSAN051834 and CFSAN051835 in Subgroup A4 (https://www.ncbi.nlm.nih.gov/Structure/tree/#!/tree/Salmonella/PDG000000002.1). The SCB (designated at that time as PDS000004690.16) encompassed a total of 24 isolates including those five isolates of our Subgroup B4. These 24 isolates in SCB were obtained from human, animal, food, and environmental sources in US and Canada (Figure 2). Within SCB, our Subgroup B4 isolates were clustered together and shared a most recent common ancestor with five NCBI isolates collected from human stool (SE PNUSAS011122, US, 2016), turkey (SE SA19943269, Canada), and chicken drag swab (SE OH-15-14655, OH, US, 2015, SE OH-12-29345, OH, US, 2012 & SE OH-13-28244, OH, US, 2013). The remaining 14 isolates in SCB formed a separate cluster, these were 13 clinical isolates and one environmental isolate that all shared a different common
ancestor from the rest of SCB. The minimum distance between isolates in SCB was one SNP while the maximum number was 104.

SCC (designated at that time as PDS000011158.1) consisted of two isolates from Subgroup B3. No other genomes from Pathogen Detection Isolates Browser fit within SCC. Similarly, no other NCBI genome fit within SCD (designated at that time as PDS000011157.1), which contained only Subgroup B1 isolates.

SCE (designated at that time as PDS000004693.11) comprised 20 isolates from chicken, mouse, and human. These isolates had been collected from the states of Tennessee, Georgia, and Pennsylvania, in the US. Eight of the Subgroup B2 isolates that fit within SCE shared a common ancestor (Figure 3). Intriguingly, all SCE isolates, with the exception isolate PNUSAS014592, carried at least one of the following antimicrobial resistance genes: tetA, aadA, blaTEM-1.

### Clade-defining SNPs

We identified clade-defining SNPs and annotations identifying synonymous/nonsynonymous changes in amino acids, positions in reference genes, strands, and gene functions are presented in Table 3.

#### Clade A polymorphisms

We identified 11 SNPs that defined Subgroup A1, including seven nonsynonymous changes, three synonymous changes, and one nonsense mutation. Type VI secretion protein IcmF (reference locus tag BCA92_14555) contained one C to A mutation, which resulted in amino acid changing A to D.
Another unique genetic signature change within Subgroup A1 occurred in the colanic acid synthesis gene `wcaF` (BCA92_08715), which changed C to T change. The nonsense mutation resulted in a stop codon which interrupted `hpaE` (BCA92_14790), encoding for enzymes involved in catabolism in the aromatic pathway.

In Subgroup A2, which contained isolates exclusively from Farm III, we discovered 19 clade-defining SNPs, including 16 in coding region. The LysR family transcriptional regulator (dBCA92_19265) contained one G to A mutation resulting in a stop codon.

Other notable findings in other subgroups included nonsynonymous mutations in `zwf` (Subgroup A3, BCA92_10040, oxidoreductase in glucose metabolism), `asnB` (Subgroup A3, BCA92_16545, asparagine synthase B), `ushA` (Subgroups A4&A5 and A6, BCA92_17470, 5′-nucleotidase), and `frsA` (Subgroup A6, BCA92_18395, esterase).

**Clade B polymorphisms**

Among the 179 clade-defining SNPs in Subgroup B1, 146 SNPs were in coding regions, including 85 nonsynonymous mutations and four nonsense mutations. Subgroup B2, which contained isolates carrying resistance genes, contained 11 SNPs with nine in coding regions. Among the isolates in Subgroups B3 and B4, we identified multiple nonsynonymous mutations, including `deoD` (BCA92_20135, purine-nucleoside phosphorylase), `cysQ` (BCA92_20970, 3′(2′),5′-bisphosphate nucleotidase activity and magnesium ion binding), `hisD` (BCA92_08915, histidinol dehydrogenase and zinc ion binding), `tolA`
(BCA92_16260, cell envelope integrity protein in transporter activity), and \textit{fimH}
(BCA92_19875, fimbrial adhesion).
Discussion

The dissemination of SE via mice, particularly on poultry farms, is considered to be one of the most serious threats to poultry industry today (2). Here, we characterized a set of 91 SE that (i) represented two organs in mice that have been associated with dissemination of SE among poultry and hence to humans, (ii) were isolated at 15 farms in Pennsylvania during the mid-1990s, which was a time during which few SE isolates from mice have previously been sequenced, and (iii) analyzed in combination with the open access NCBI Pathogen Detection Isolates Browser. These steps allow us to construct a more nuanced picture of SE dissemination during the 1990s, and also identify connections between historic isolates and current SE phenotypes.

Our study demonstrated that WGS not only reliably distinguishes among closely related SE isolates from mice and trace a genome back to its farm of origin and year of isolation, but also allows sufficient resolution to distinguish between SE isolates, even those collected from different organs (spleens and intestines) of individual mouse. In addition, our analyses showed that (i) isolates carrying antimicrobial resistance genes formed a separate subgroup, which could indicate a shared mechanism which enables that feature, (ii) open access WGS database contributes comprehensive perspectives to our understanding of selected isolates, and (iii) new clade-defining markers and NCBI Pathogen Detection Isolates Browser SNP Clusters were identified, offering tool with high resolution in outbreak investigations and rapid detections to identify specific clade related to certain years or locations.

Our results strongly suggested it was possible for unique ecologies of SE to develop on individual farms, although local adaptation is not inevitable. Farms I and X exemplify this
range of possibilities: Farm I exhibited heterogeneous isolates while isolates from Farm X were shown to be highly similar. Isolates can spread from one location to another in multiple ways: insects (29), wild birds (30), wild animals (31, 32), and even wind (33) can move contamination from one place to another. However, among these possible transmission routes, mice are ubiquitous pests (8-10), and their behaviors may help shape those unique local ecologies: mice migrate periodically and also defend their territories. Understanding the genetic relatedness among the SE carried by mice and the SE found in veterinary, food, and human sampling will help improve safety and security in poultry industry.

**WGS data identified a subgroup consisting exclusively of isolates carrying antimicrobial resistance genes.**

Previously, WGS has been used to differentiate drug-resistant *S. enterica* isolates from different locations, which can exhibit notable differences in resistant-relevant genotypic and phenotypic characteristics (34). Other research has shown WGS can be valuable in predicting phenotypic resistance among both *S. enterica* (34, 35) and *E. coli* (36). In the current study, WGS analyses revealed that all our Subgroup B2 isolates carried *bla*TEM-1 and *tetA*. It is possible that Subgroup B2 isolates share specific genetic features that permit them to obtain and carry antimicrobial resistance genes via horizontal gene transfer, or make it more likely for those genes to be maintained. For example, bacteria that carry non-functional Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) /cas system could acquire plasmids carrying antimicrobial resistance genes. Possession of a fully-functioning CRISPR/cas system is reversely correlated with antimicrobial resistance in bacteria (37-39).
Open access genome databases allow greatly expanded genomic and phylogenetic investigations

In the NCBI Pathogen Detection Isolates Browser, comprehensive data was available for each genome, including up to 40 columns of detail such as WGS run qualities, outbreak relatedness, and antimicrobial resistance genotypes. The Browser also assigns specific cluster ID numbers computed based on SNP distances. Although these cluster numbers can change as new information is added to the Browser, this feature allows researchers to quickly identify isolates most closely related to target isolates, which can assist in recognizing possible connections among clinical illness cases. The phylogenetic analyses from the Browser were consistent with our phylogenetic tree.

Multiple subgroups in the current study formed distinct SNP Clusters containing isolates exclusively from our collection, like B3 isolates in SCC. We identified clinical isolates and poultry related isolates closely related to our isolates, such as SE PNUSAS011122 (human stool, US, 2016) and SE OH-15-14655 (chicken drag swab, OH, US, 2015) with B4 isolates in SCB. Our data has the potential to bridge surveillance data with long-term and large-scale genomics and phylogenetics studies (19).

Genetic variations in clade-defining SNPs showed possible unique genotypic and phenotypic features.

Distinctive genetic features are extremely useful for epidemiologic investigations. Finding such genetic identifiers can help rapidly determine outbreak lineages and accurately distinguish highly clonal clades (6). The nonsynonymous changes we identified in this study suggested that a combination of several genetic factors has facilitated the survival and growth of SE, resulting in different contamination risks for
each subgroup. For example, the *icmF* we identified in Subgroup A1 was part of Type VI Secretion System, which is known to be required for full virulence in mice (40, 41).

Similarly, *fimH* alleles have been associated with the abilities of *Salmonella* to bind onto avian or mammalian cells (42). Despite the clonal structure of SE, isolates vary greatly in the ability to contaminate eggs, which is biologically independent of phage types those isolates belong to (15, 43, 44). The heterogeneity of metabolic profiles in SE isolates might provide an explanation for the variation in contamination capability (15).

The accumulation of mutations that affect gene function is a significant part of the process by which *S. enterica* becomes host adapted (45). Such host adaptations may well be occurring at some of the farms where we collected SE from local mice, with important consequences for the safety and security of the poultry supply chain. Notably, serovars Enteritidis, Gallinarum, and Pullorum can circulate within the same farm, and sometimes within the same bird, as evidenced by field analyses conducted in South America (46). Therefore, WGS also has potential for detecting evolutionary trends within SE that could threaten the poultry industry supply chain. Our data also pave the way for research on poultry pathogenic serovars *S. Gallinarum* and *S. Pullorum*, which diverged independently from an Enteritidis-like ancestor (3, 47, 48).
Acknowledgements

The authors thank Dr. Lili Fox Vélez for scientific writing support and Oak Ridge Institute for Science and Education (ORISE) Research Participation Program at the U.S. Food and Drug Administration support on this project.
**Figure Legends**

Figure 1. Maximum likelihood phylogenetic tree of 91 *S. Enteritidis* isolates from mice spleens and intestines. We constructed the phylogenetic tree using 742 single nucleotide polymorphisms (SNPs). All sequenced isolates were divided into two major clades, Clade A and B, which were further grouped into 12 subgroups.

Figure 2. Phylogenetic tree of SNP Cluster B (SCB, designated at that time as PDS000004690.16) from NCBI Pathogen Detection Isolates Browser. The phylogenetic tree encompassed 24 isolates including our five sequenced isolates belonging to Subgroup B4.

Figure 3. Phylogenetic tree of SNP Cluster E (SCE, designated at that time as PDS000004693.11) from NCBI Pathogen Detection Isolates Browser. The phylogenetic tree encompassed 20 isolates including our eight sequenced isolates belonging to Subgroup B2.
References

1. Zheng J, Pettengill J, Strain E, Allard MW, Ahmed R, Zhao S, Brown EW. 2014. Genetic diversity and evolution of *Salmonella enterica* serovar Enteritidis strains with different phage types. J Clin Microbiol 52:1490-500.

2. Darrell Trampel TH, Richard Gast. 2014. Integrated farm management to prevent *Salmonella* Enteritidis contamination of eggs. Journal of Applied Poultry Research 2014:353-365.

3. Thomson NR, Clayton DJ, Windhorst D, Vernikos G, Davidson S, Churcher C, Quail MA, Stevens M, Jones MA, Watson M, Barron A, Layton A, Pickard D, Kingsley RA, Bignell A, Clark L, Harris B, Ormond D, Abdellah Z, Brooks K, Cherevach I, Chillingworth T, Woodward J, Norberczak H, Lord A, Arrowsmith C, Jagels K, Moule S, Mungall K, Sanders M, Whitehead S, Chabalgoity JA, Maskell D, Humphrey T, Roberts M, Barrow PA, Dougan G, Parkhill J. 2008. Comparative genome analysis of *Salmonella* Enteritidis PT4 and *Salmonella* Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. Genome Res 18:1624-37.

4. Feasey NA, Hadfield J, Keddy KH, Dallman TJ, Jacobs J, Deng X, Wigley P, Barquist L, Langridge GC, Feltwell T, Harris SR, Mather AE, Fookes M, Aslett M, Msefula C, Kariuki S, Maclennan CA, Onsare RS, Weil FX, Le Hello S, Smith AM, McClelland M, Desai P, Parry CM, Cheesbrough J, French N, Campos J, Chabalgoity JA, Betancor L, Hopkins KL, Nair S, Humphrey TJ, Lunguya O, Cogan TA, Tapia MD, Sow SO, Tennant SM, Bornstein K, Levine MM, Lacharme-Lora L, Everett DB, Kingsley RA, Parkhill J, Heyderman RS, Dougan G, Gordon MA, Thomson NR. 2016. Distinct *Salmonella* Enteritidis lineages associated with enterocolitis in high-income settings and invasive disease in low-income settings. Nat Genet 48:1211-1217.

5. Inns T, Ashton PM, Herrera-Leon S, Lighthill J, Foulkes S, Jombart T, Rehman Y, Fox A, Dallman T, E DEP, Browning L, Coia JE, Edeghere O, Vivancos R. 2017. Prospective use of whole genome sequencing (WGS) detected a multi-country outbreak of *Salmonella* Enteritidis. Epidemiol Infect 145:289-298.

6. Allard MW, Luo Y, Strain E, Pettengill J, Timme R, Wang C, Li C, Keys CE, Zheng J, Stones R, Wilson MR, Musser SM, Brown EW. 2013. On the evolutionary history, population genetics and diversity among isolates of *Salmonella* Enteritidis PFGE pattern JEGX01.0004. PLoS One 8:e55254.

7. Henzler DJ, Opitz HM. 1992. The role of mice in the epizootiology of *Salmonella* enteritidis infection on chicken layer farms. Avian Dis 36:625-31.
9. Umali DV, Lapuz RR, Suzuki T, Shirota K, Katoh H. 2012. Transmission and shedding patterns of *Salmonella* in naturally infected captive wild roof rats (Rattus rattus) from a *Salmonella*-contaminated layer farm. Avian Dis 56:288-94.

10. Whiley H, Ross K. 2015. *Salmonella* and eggs: from production to plate. Int J Environ Res Public Health 12:2543-56.

11. Klemm EJ, Gkrania-Klotsas E, Hadfield J, Forbester JL, Harris SR, Hale C, Heath JN, Wileman T, Clare S, Kane L, Gouplding D, Otto TD, Kay S, Doffinger R, Cooke FJ, Carmichael A, Lever AM, Parkhill J, MacLennan CA, Kumararatne D, Dougan G, Kingsley RA. 2016. Emergence of host-adapted *Salmonella* Enteritidis through rapid evolution in an immunocompromised host. Nat Microbiol 1:15023.

12. Inns T, Lane C, Peters T, Dallman T, Chatt C, McFarland N, Crook P, Bishop T, Edge J, Hawker J, Elson R, Neal K, Adak GK, Cleary P, Outbreak Control T. 2015. A multi-country *Salmonella* Enteritidis phage type 14b outbreak associated with eggs from a German producer: ‘near real-time’ application of whole genome sequencing and food chain investigations, United Kingdom, May to September 2014. Euro Surveill 20.

13. Sanch Kanagarajah AW, Gayle Dolan, Claire Jenkins, Philip Ashton, Antonio Isidro Carrion Martin, Robert Davies, Andrew Frost, Timothy Dallman, Elizabeth De Pinna, Jeremy Hawker, Kathie Grant, Richard Elson. 2017. Whole genome sequencing reveals an outbreak of *Salmonella* Enteritidis associated with reptile feeder mice in the United Kingdom, 2012-2015. Food Microbiology http://dx.doi.org/10.1016/j.fm.2017.04.005.

14. Lienau EK, Strain E, Wang C, Zheng J, Ottesen AR, Keys CE, Hammack TS, Musser SM, Brown EW, Allard MW, Cao G, Meng J, Stones R. 2011. Identification of a salmonellosis outbreak by means of molecular sequencing. N Engl J Med 364:981-2.

15. Guard J, Morales CA, Fedorka-Cray P, Gast RK. 2011. Single nucleotide polymorphisms that differentiate two subpopulations of *Salmonella enteritidis* within phage type. BMC Res Notes 4:369.

16. Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, Paxinos EE, Sebra R, Chin CS, Iliopoulos D, Klammer A, Peluso P, Lee L, Kisyulyk AO, Bullard J, Kasarskis A, Wang S, Eid J, Rank D, Redman JC, Steyert SR, Frimodt-Moller J, Struve C, Petersen AM, Krogfelt KA, Nataro JP, Schadt EE, Waldor MK. 2011. Origins of the E. coli strain causing an outbreak of hemolytic-uremic syndrome in Germany. N Engl J Med 365:709-17.

17. Hendriksen RS, Price LB, Schupp JM, Gillece JD, Kaas RS, Engelthaler DM, Bortolaia V, Pearson T, Waters AE, Upadhyay BP, Shrestha SD, Adhikari S, Shakya G, Kelm PS, Aarestrup FM. 2011. Population genetics of Vibrio cholerae from Nepal in 2010: evidence on the origin of the Haitian outbreak. MBio 2:e00157-11.

18. Harris SR, Cartwright EJ, Torok ME, Holden MT, Brown NM, Ogilvy-Stuart AL, Ellington MJ, Quail MA, Bentley SD, Parkhill J, Peacock SJ. 2013. Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: a descriptive study. Lancet Infect Dis 13:130-6.

19. Allard MW, Strain E, Melka D, Bunning K, Musser SM, Brown EW, Timme R. 2016. Practical Value of Food Pathogen Traceability through Building a Whole-Genome Sequencing Network and Database. J Clin Microbiol 54:1975-83.

20. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G,
Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455-77.

Klimke W, Agarwala R, Badretdin A, Chetvernin S, Ciufo S, Fedorov B, Kiryutin B, O'Neill K, Resch W, Resenchuk S, Schafer S, Tolstoy I, Tatusova T. 2009. The National Center for Biotechnology Information's Protein Clusters Database. Nucleic Acids Res 37:D216-23.

Yao K, Muruvanda T, Roberts RJ, Payne J, Allard MW, Hoffmann M. 2016. Complete Genome and Methylome Sequences of Salmonella enterica subsp. enterica Serovar Panama (ATCC 7378) and Salmonella enterica subsp. enterica Serovar Sloterdijk (ATCC 15791). Genome Announc 4.

Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563-9.

Davis S PJ, Luo Y, Payne J, Shpuntoff A, Rand H, Strain E. 2015. CFSAN SNP Pipeline: an automated method for constructing SNP matrices from next-generation sequence data. PeerJ Computer Science 1:e20 https://doi.org/10.7717/peerj-cs.20.

Zwickl DJ. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD dissertation, The University of Texas at Austin.

Reumerman RA, Tucker NP, Herron PR, Hoskisson PA, Sangal V. 2013. Tool for rapid annotation of microbial SNPs (TRAMS): a simple program for rapid annotation of genomic variation in prokaryotes. Antonie Van Leeuwenhoek 104:431-4.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33:1870-4.

Toro M, Retamal P, Ayers S, Barreto M, Allard M, Brown EW, Gonzalez-Escalona N. 2016. Whole-Genome Sequencing Analysis of Salmonella enterica Serovar Enteritidis Isolates in Chile Provides Insights into Possible Transmission between Gulls, Poultry, and Humans. Appl Environ Microbiol 82:6223-6232.

Bernard Greenberg JAK, Marc J. Klowden. 1970. Factors affecting the transmission of Salmonella by flies: natural resistance to colonization and bacterial interference. Infection and Immunity 2:800-809.

Tizard I. 2004. Salmonellosis in wild birds. Journal of Exotic Pet Medicine 13:50-66.

M.N. Skow JJM, C. Rahbek, J. Lodal, J.B. Jespersen, J.C.Jorgensen, H.H. Dietz, M. Chriel, D.L. Baggesen. 2008. Transmission of Salmonella between wildlife and meat-production animals in Denmark. Journal of Applied Microbiology 105:1558-1568.

F. Hillbert FJMS, R. Chopra-Dewasthaly, P. Paulsen. 2012. Salmonella in the wildlife-human interface. Food Research International 45:603-608.

B. W. Mitchell RJB, M. E. Berrang, J. S. Bailey, N. A. Cox. 2002. Reducing airborne pathogens, dust and Salmonella transmission in experimental hatching cabinets using an electrostatic space charge system. Poultry Science 81:49-55.

Carroll LM, Wiedmann M, den Bakker H, Siler J, Warchocki S, Kent D, Lyalina S, Davis M, Sischo W, Besser T, Warnick LD, Pereira RV. 2017. Whole-Genome Sequencing of Drug-Resistant Salmonella enterica Isolates from Dairy Cattle and Humans in New York and Washington States Reveals Source and Geographic Associations. Appl Environ Microbiol 83.
35. McDermott PF, Tyson GH, Kabera C, Chen Y, Li C, Folster JP, Ayers SL, Lam C, Tate HP, Zhao S. 2016. Whole-Genome Sequencing for Detecting Antimicrobial Resistance in Nontyphoidal Salmonella. Antimicrob Agents Chemother 60:5515-20.

36. Tyson GH, McDermott PF, Li C, Chen Y, Tadesse DA, Mukherjee S, Bodeis-Jones S, Kabera C, Gaines SA, Loneragan GH, Edrington TS, Torrence M, Harhay DM, Zhao S. 2015. WGS accurately predicts antimicrobial resistance in Escherichia coli. J Antimicrob Chemother 70:2763-9.

37. Gawryszewska I, Malinowska K, Kuch A, Chrobak-Chmiel D, Laniewska-Trokenheim L, Hryniewicz W, Sadowy E. 2017. Distribution of antimicrobial resistance determinants, virulence-associated factors and CRISPR loci in isolates of Enterococcus faecalis from various settings and genetic lineages. Pathog Dis doi:10.1093/femspd/ftx021.

38. Palmer KL, Gilmore MS. 2010. Multidrug-resistant enterococci lack CRISPR-cas. MBio 1.

39. Price VJ, Huo W, Sharifi A, Palmer KL. 2016. CRISPR-cas and restriction-modification act additively against conjugative antibiotic resistance plasmid transfer in Enterococcus faecalis. mSphere 1.

40. Liu J, Guo JT, Li YG, Johnston RN, Liu GR, Liu SL. 2013. The type VI secretion system gene cluster of Salmonella typhimurium: required for full virulence in mice. J Basic Microbiol 53:600-7.

41. Mulder DT, Cooper CA, Coombes BK. 2012. Type VI secretion system-associated gene clusters contribute to pathogenesis of Salmonella enterica serovar Typhimurium. Infect Immun 80:1996-2007.

42. Guo A, Cao S, Tu L, Chen P, Zhang C, Jia A, Yang W, Liu Z, Chen H, Schifferli DM. 2009. FimH alleles direct preferential binding of Salmonella to distinct mammalian cells or to avian cells. Microbiology 155:1623-33.

43. Jean Guard, Devendra Shah, Cesar Morales, Doug Call. 2011. Evolutionary trends associated with niche specialization as modeled by whole genome analysis of egg-contaminating Salmonella enterica serovar Enteritidis. Salmonella: From Genome to Function, Chapter 5; Editors: Steffen Porwollik:91-106.

44. Pan Z, Carter B, Nunez-Garcia J, Abuoun M, Fookes M, Ivens A, Woodward MJ, Anjum MF. 2009. Identification of genetic and phenotypic differences associated with prevalent and non-prevalent Salmonella Enteritidis phage types: analysis of variation in amino acid transport. Microbiology 155:3200-13.

45. Langridge GC, Fookes M, Connor TR, Feltwell T, Feasey N, Parsons BN, Seth-Smith HM, Barquist L, Stedman A, Humphrey T, Wigley P, Peters SE, Maskell DJ, Corander J, Chabalgoity JA, Barrow P, Parkhill J, Dougan G, Thomson NR. 2015. Patterns of genome evolution that have accompanied host adaptation in Salmonella. Proc Natl Acad Sci U S A 112:863-8.

46. Pulido-Landinez M, Sanchez-Ingunza R, Guard J, do Nascimento VP. 2014. Presence of Salmonella enteritidis and Salmonella gallinarum in commercial laying hens diagnosed with fowl typhoid disease in Colombia. Avian Dis 58:165-70.

47. Porwollik S, Santiviago CA, Cheng P, Florea L, Jackson S, McClelland M. 2005. Differences in gene content between Salmonella enterica serovar Enteritidis isolates and comparison to closely related serovars Gallinarum and Dublin. J Bacteriol 187:6545-55.
Porwollik S, Boyd EF, Choy C, Cheng P, Florea L, Proctor E, McClelland M. 2004. Characterization of *Salmonella enterica* subspecies I genovars by use of microarrays. *J Bacteriol* 186:5883-98.
Figure 1. Maximum likelihood phylogenetic tree of 91 S. Enteritidis isolates from mice spleens and intestines. We constructed the phylogenetic tree using 742 single nucleotide polymorphisms (SNPs). All sequenced isolates were divided into two major clades, Clade A and B, which were further grouped into 12 subgroups.
Table 1. The metadata and general genomic information of 91 sequenced S. Enteritidis in current study.

| Strain            | Year | Source  | Farm ID | Clade | SNP Cluster | Contig# (>500bp) | AMR Gene | SRA Accession |
|-------------------|------|---------|---------|-------|-------------|------------------|----------|---------------|
| CFSAN051823       | 1995 | Spleen* | I (ST11)| A8    | A           | 37               |          | SRR5063209    |
| CFSAN051824       | 1995 | Spleen* | I (ST11)| A8    | A           | 36               |          | SRR5063211    |
| CFSAN051825       | 1995 | Spleen* | I (ST11)| A8    | A           | 40               |          | SRR5063208    |
| CFSAN051826       | 1995 | Spleen* | I (ST3632)| B4  | N/A         | 39               |          | SRR5063210    |
| CFSAN051827       | 1995 | Spleen* | I (ST3632)| B4  | B           | 37               |          | SRR5063216    |
| CFSAN051828       | 1995 | Spleen* | I (ST3632)| B4  | B           | 38               |          | SRR5064765    |
| CFSAN051829       | 1995 | Spleen* | I (ST3632)| B4  | B           | 40               |          | SRR5064766    |
| CFSAN051830       | 1995 | Spleen* | I (ST3632)| B4  | B           | 37               |          | SRR5064767    |
| CFSAN051831       | 1995 | Spleen* | I (ST3632)| B4  | B           | 35               |          | SRR5064768    |
| CFSAN051832       | 1995 | Spleen* | I (ST11)| B3    | C           | 38               |          | SRR5064769    |
| CFSAN051833       | 1995 | Spleen* | I (ST11)| B3    | C           | 39               |          | SRR5064773    |
| CFSAN051834       | 1995 | Spleen* | II (ST11)| A4   | A           | 39               |          | SRR5064774    |
| CFSAN051835       | 1995 | Spleen* | II (ST11)| A4   | A           | 36               |          | SRR5064771    |
| CFSAN051836(m1)   | 1995 | Spleen  | III (ST11)| A2   | A           | 39               |          | SRR5064772    |
| CFSAN051837       | 1995 | Spleen* | III (ST11)| A2   | A           | 39               |          | SRR5064777    |
| CFSAN051838(m2)   | 1995 | Spleen  | III (ST11)| A2   | A           | 38               |          | SRR5064775    |
| CFSAN051839(m3)   | 1995 | Spleen  | III (ST11)| A2   | A           | 37               |          | SRR5064776    |
| CFSAN051840(m4)   | 1995 | Spleen  | III (ST11)| A2   | A           | 40               |          | SRR5064779    |
| CFSAN051841(m4)   | 1995 | Intestine| III (ST11)| A2   | A           | 39               |          | SRR5064781    |
| CFSAN051842       | 1995 | Intestine| III (ST11)| A2   | A           | 36               |          | SRR5064783    |
| CFSAN051843       | 1995 | Intestine| III (ST11)| A2   | A           | 41               |          | SRR5064782    |
| CFSAN051844(m2)   | 1995 | Intestine| III (ST11)| A2   | A           | 42               |          | SRR5064786    |
| CFSAN051845(m3)   | 1995 | Intestine| III (ST11)| A2   | A           | 40               |          | SRR5064785    |
| CFSAN051846(m1)   | 1995 | Intestine| III (ST11)| A2   | A           | 44               |          | SRR5064787    |
| CFSAN051847       | 1995 | Spleen  | IV (ST11)| A6    | A           | 37               |          | SRR5064784    |
| CFSAN051848       | 1995 | Spleen  | IV (ST11)| A6    | A           | 58               |          | SRR5064788    |
| CFSAN051849       | 1995 | Intestine| IV (ST11)| A6    | A           | 41               |          | SRR5064792    |
| CFSAN051851       | 1995 | Intestine| IV (ST11)| A6    | A           | 38               |          | SRR5064795    |
| Accession | Year  | Location | Sublocation | Assignments | SeqLen | BLAST IDs |
|-----------|-------|----------|-------------|-------------|--------|-----------|
| CFSAN051852(m5) | 1996 | Spleen | V (ST11) | A3 A 40 | SRR5064794 |
| CFSAN051853(m6) | 1996 | Spleen | V (ST11) | A3 A 43 | SRR5064790 |
| CFSAN051854(m7) | 1996 | Spleen | V (ST11) | A3 A 70 | SRR5064797 |
| CFSAN051855 | 1996 | Intestine | V (ST11) | A3 A 39 | SRR5064802 |
| CFSAN051856 | 1996 | Intestine | V (ST11) | A3 A 38 | SRR5064799 |
| CFSAN051858(m5) | 1996 | Intestine | V (ST11) | A3 A 44 | SRR5064800 |
| CFSAN051859 | 1996 | Intestine | V (ST11) | A3 A 45 | SRR5064803 |
| CFSAN051860(m6) | 1996 | Intestine | V (ST11) | A3 A 45 | SRR5064804 |
| CFSAN051861 | 1996 | Intestine | V (ST11) | A3 A 39 | SRR5064801 |
| CFSAN051862(m7) | 1996 | Intestine | V (ST11) | A3 A 39 | SRR5064806 |
| CFSAN051864 | 1996 | Spleen | VII (ST11) | A3 A 39 | SRR5064810 |
| CFSAN051865 | 1996 | Spleen | VII (ST11) | A3 A 40 | SRR5064809 |
| CFSAN051866 | 1996 | Spleen | VII (ST11) | A3 A 41 | SRR5064812 |
| CFSAN051867 | 1996 | Spleen | VII (ST11) | A3 A 40 | SRR5064811 |
| CFSAN051868 | 1996 | Spleen | VII (ST11) | A3 A 37 | SRR5064813 |
| CFSAN051870 | 1996 | Spleen | VII (ST11) | A3 A 39 | SRR5064814 |
| CFSAN051871 | 1996 | Spleen | VII (ST11) | A3 A 49 | SRR5064817 |
| CFSAN051872 | 1996 | Spleen | VIII (ST11) | A7 A 99 | SRR5064815 |
| CFSAN051873 | 1996 | Spleen | VIII (ST11) | A7 A 35 | SRR5064816 |
| CFSAN051874 | 1996 | Spleen | VIII (ST11) | A7 A 39 | SRR5064818 |
| CFSAN051875 | 1996 | Spleen | VIII (ST11) | A7 A 39 | SRR5064821 |
| CFSAN051876 | 1996 | Spleen | VIII (ST11) | A7 A 39 | SRR5064820 |
| CFSAN051877 | 1996 | Spleen | VIII (ST11) | A7 A 40 | SRR5064854 |
| CFSAN051878 | 1996 | Spleen | VIII (ST11) | A7 A 115 | SRR5064855 |
| CFSAN051880 | 1996 | Spleen | X (ST11) | A7 A 40 | SRR5064857 |
| CFSAN051881(m8) | 1997 | Intestine | IX (ST11) | B1 D 43 | SRR5065189 |
| CFSAN051882 | 1997 | Intestine | X (ST11) | B1 D 39 | SRR5065192 |
| CFSAN051883 | 1997 | Spleen | X (ST11) | B2 N/A 46 | SRR5065190 |
| CFSAN051884 | 1997 | Spleen | X (ST11) | B1 D 40 | SRR5065191 |
| CFSAN051885 | 1997 | Spleen | X (ST11) | B1 D 40 | SRR5065194 |
| CFSAN051886(m8) | 1997 | Spleen | XI (ST11) | B1 N/A 38 | SRR5065196 |
| CFSAN051887 | 1997 | Spleen | XI (ST11) | B2 N/A 40 | SRR5065195 |
| CFSAN051888 | 1997 | Spleen | VI (ST11) | B2 E 42 | SRR5065193 |
| CFSAN  | Year | Tissue | ST (Group) | MLST | WGS | Genotypic Profile | Isolate ID |
|--------|------|--------|------------|------|-----|------------------|------------|
| 051889 | 1997 | Spleen | VI (ST11)  | B2   | E   | 53 *bla*TEM-1, tetA* | SRR5065198|
| 051890 | 1997 | Spleen | XI (ST11)  | B2   | E   | 42 *bla*TEM-1, tetA* | SRR5065197|
| 051891 | 1997 | Spleen | XI (ST11)  | B2   | E   | 41 *bla*TEM-1, tetA* | SRR5065199|
| 063779 | 1996 | Spleen | XII (ST11) | A1   | A   | 15 *bla*TEM-1, tetA* | SRR5884037|
| 063780 | 1996 | Spleen | XII (ST11) | A1   | A   | 15 *bla*TEM-1, tetA* | SRR5884036|
| 063781 | 1996 | Spleen | XII (ST11) | A1   | A   | 15 *bla*TEM-1, tetA* | SRR5970532|
| 063782 | 1996 | Spleen | XII (ST11) | A1   | A   | 14 *bla*TEM-1, tetA* | SRR5884033|
| 063783 | 1996 | Intestine | XII (ST11) | A1   | A   | 14 *bla*TEM-1, tetA* | SRR5884041|
| 063784 | 1997 | Intestine | XI (ST11)  | B2   | E   | 32 *bla*TEM-1, tetA* | SRR5884042|
| 063785 | 1997 | Intestine | XI (ST11)  | B2   | E   | 28 *bla*TEM-1, tetA* | SRR5884050|
| 063786 | 1997 | Intestine | XI (ST11)  | B2   | E   | 28 *bla*TEM-1, tetA* | SRR5884043|
| 063787 | 1997 | Intestine | XI (ST11)  | B2   | E   | 29 *bla*TEM-1, tetA* | SRR5884057|
| 063788 | 1997 | Spleen | XIII (ST11) | A5   | A   | 22 *bla*TEM-1, tetA* | SRR5819771|
| 063789 | 1998 | Spleen | XIII (ST11) | A5   | A   | 18 *bla*TEM-1, tetA* | SRR5819768|
| 063790 | 1998 | Intestine | XIII (ST11) | A5   | A   | 78 *bla*TEM-1, tetA* | SRR5819773|
| 063791 | 1998 | Spleen | XIV (ST11)  | A5   | A   | 18 *bla*TEM-1, tetA* | SRR5819769|
| 063792 | 1998 | Spleen | XIV (ST11)  | A5   | A   | 20 *bla*TEM-1, tetA* | SRR5819774|
| 063793 | 1995 | Spleen | XV (ST11)  | A1   | A   | 17 *bla*TEM-1, tetA* | SRR5819770|
| 063794 | 1995 | Spleen | XV (ST11)  | A1   | A   | 14 *bla*TEM-1, tetA* | SRR5819775|
| 063795 | 1995 | Spleen | XV (ST11)  | A1   | A   | 14 *bla*TEM-1, tetA* | SRR5819777|
| 063796 | 1995 | Spleen | XV (ST11)  | A1   | A   | 15 *bla*TEM-1, tetA* | SRR5819776|
| 063797 | 1995 | Spleen | XV (ST11)  | A1   | A   | 15 *bla*TEM-1, tetA* | SRR5819784|
| 063798 | 1995 | Spleen | XV (ST11)  | A1   | A   | 18 *bla*TEM-1, tetA* | SRR5819788|
| 063799 | 1995 | Spleen | XV (ST11)  | A1   | A   | 14 *bla*TEM-1, tetA* | SRR5819790|
| 063800 | 1995 | Spleen | XV (ST11)  | A1   | A   | 15 *bla*TEM-1, tetA* | SRR5819779|
| 063801 | 1995 | Spleen | XV (ST11)  | A1   | A   | 19 *bla*TEM-1, tetA* | SRR5819785|
| 063802 | 1995 | Intestine | XV (ST11) | A1   | A   | 25 *bla*TEM-1, tetA* | SRR5819789|
| 063803 | 1995 | Intestine | XV (ST11) | A1   | N/A | 30 *bla*TEM-1, tetA* | SRR5819781|
| 063804 | 1995 | Intestine | XV (ST11) | A1   | A   | 47 *bla*TEM-1, tetA* | SRR5819782|
| 063805 | 1995 | Intestine | XV (ST11) | A1   | A   | 25 *bla*TEM-1, tetA* | SRR5819778|

*Spleen* Spleen cultured but intestine not cultured, unknown about intestine.
SNP Cluster Number
A PDS000002757.323
B PDS000004690.16
C PDS000011158.1
D PDS000011157.1
E PDS000004693.11
Table 2. The number of SNP differences (standard deviation) between 12 subgroups.

|     | A1 | A2   | A3   | A4   | A5   | A6   | A7   | A8   | B1   | B2   | B3   |
|-----|----|------|------|------|------|------|------|------|------|------|------|
| A1  |    | A2   | 44(5)|      |      |      |      |      |      |      |      |
| A3  | 50(6)| 45(6)|      |      |      |      |      |      |      |      |      |
| A4  | 40(5)| 36(5)| 39(6)|      |      |      |      |      |      |      |      |
| A5  | 44(5)| 40(5)| 43(6)| 32(5)|      |      |      |      |      |      |      |
| A6  | 54(6)| 50(6)| 55(7)| 46(6)| 50(6)|      |      |      |      |      |      |
| A7  | 51(6)| 47(6)| 52(6)| 43(6)| 47(6)| 57(7)|      |      |      |      |      |
| A8  | 50(6)| 45(6)| 51(6)| 41(6)| 45(6)| 55(7)| 24(5)|      |      |      |      |
| B1  | 92(8)| 90(8)| 95(8)| 86(8)| 90(8)| 99(8)| 97(8)| 95(8)|      |      |      |
| B2  | 84(8)| 82(8)| 87(8)| 78(8)| 82(8)| 91(8)| 89(8)| 87(8)| 33(5)|      |      |
| B3  | 137(9)| 135(9)| 140(10)| 131(9)| 135(9)| 144(10)| 142(10)| 140(10)| 86(8)| 78(8)|      |
| B4  | 133(9)| 131(9)| 137(9)| 127(9)| 132(9)| 141(10)| 138(10)| 137(9)| 83(8)| 75(7)| 46(6)|
| Location | Accession | Annotation | Locus_tag | Gene | Positions in coding | Nucleotide change | Amino acid change | Synonymous / Nonsynonymous | Strand | Product name |
|----------|-----------|------------|-----------|------|---------------------|------------------|-------------------|--------------------------|--------|--------------|
| A1 (18 samples / 14 SNPs) | | | | | | | | | | |
| 66674 | NZ_CP022003.1 | coding | BCA92_00310 | 2562 | TCG -> TCA | S -> S | S | + | transcriptional regulator |
| 292426 | NZ_CP022003.1 | coding | BCA92_01355 | 84 | CGT -> TGT | R -> C | N | - | methyl-accepting chemotaxis protein II |
| 480015 | NZ_CP022003.1 | coding | BCA92_02350 | 689 | GAC -> GCC | D -> G | N | - | 505 ribosomal protein L11 methyltransferase |
| 108137 | NZ_CP022003.1 | coding | BCA92_05455 | 705 | GCC -> GCT | A -> V | N | - | ribonucleoside-diphosphate reductase subunit alpha |
| 123727 | NZ_CP022003.1 | coding | BCA92_06615 | 181 | CTG -> TGT | L -> L | S | + | IMP dehydrogenase |
| 1408128 | NZ_CP022003.1 | coding | BCA92_07015 | 323 | CCT -> CTT | P -> L | N | - | glucose-specific phosphotransferase enzyme IIIA component |
| 168417 | NZ_CP022003.1 | coding | BCA92_08220 | 490 | CTG -> CTA | L -> L | S | - | hypothetical protein |
| 1755377 | NZ_CP022003.1 | coding | BCA92_08715 | wcaF | 251 | GCT -> GTT | A -> V | N | + | colanic acid biosynthesis acetyltransferase WcaF |
| 2549901 | NZ_CP022003.1 | intergenic | | | | | | | | |
| 2732526 | NZ_CP022003.1 | intergenic | | | | | | | | |
| 2803507 | NZ_CP022003.1 | coding | BCA92_14555 | 512 | GCC -> GAC | A -> D | N | + | type VI secretion protein LcmF |
| 2845163 | NZ_CP022003.1 | coding | BCA92_14790 | hpoE | 1001 | TGG -> TAG | W -> * | nonsense | - | S-carboxymethyl-2-hydroxymuconate semialdehyde dehydrogenase |
| 3031053 | NZ_CP022003.1 | coding | BCA92_15710 | 399 | GCG -> AGC | A -> T | N | - | glutathione ABC transporter permease |
| 3837497 | NZ_CP022003.1 | intergenic | | | | | | | | |
| A2 (11 samples / 19 SNPs) | | | | | | | | | |
| 92535 | NZ_CP022003.1 | coding | BCA92_00440 | 33 | CAG -> CAC | Q -> H | N | + | alpha-xyllosidase |
| 443373 | NZ_CP022003.1 | coding | BCA92_02080 | 96 | CGC -> GGT | R -> R | S | + | 305 ribosomal protein S19 |
| 594188 | NZ_CP022003.1 | intergenic | | | | | | | | |
| 586670 | NZ_CP022003.1 | coding | BCA92_02880 | pnp | 410 | GCG -> GTG | A -> V | N | + | polyribonucleotide nucleotidyltransferase |
| 635474 | NZ_CP022003.1 | intergenic | | | | | | | | |
| 840701 | NZ_CP022003.1 | coding | BCA92_04270 | 138 | CGC -> GGT | R -> R | S | + | transcriptional regulator |
| 901519 | NZ_CP022003.1 | coding | BCA92_04520 | fucI | 525 | GAA -> AAA | E -> K | N | - | L-fucose isomerase |
| 1033097 | NZ_CP022003.1 | coding | BCA92_05185 | 376 | CAG -> CAA | Q -> Q | S | - | hydrogenase formation protein HypD |
| 1156296 | NZ_CP022003.1 | coding | BCA92_05925 | 193 | CGA -> AGA | R -> R | S | + | late control protein D |
| 1381137 | NZ_CP022003.1 | coding | BCA92_06875 | 1098 | GCC -> GCT | A -> A | S | + | alcohol dehydrogenase EuG |
| 1724432 | NZ_CP022003.1 | coding | BCA92_08600 | 263 | TGG -> TAG | W -> * | nonsense | - | DNA-binding response regulator |
| 2294640 | NZ_CP022003.1 | coding | BCA92_11725 | 124 | GCC -> GCA | A -> A | S | - | hydrogenase formation protein |
| 2852680 | NZ_CP022003.1 | coding | BCA92_14830 | 1189 | GAT -> AAT | D -> N | N | + | two-component sensor histidine kinase |
| 3172034 | NZ_CP022003.1 | coding | BCA92_16425 | 826 | CGG -> TGG | R -> W | N | + | two-component sensor histidine kinase |
| 3749568 | NZ_CP022003.1 | coding | BCA92_19265 | 307 | TGG -> TGA | W -> * | nonsense | - | LysR family transcriptional regulator |
| 3782361 | NZ_CP022003.1 | coding | BCA92_19405 | leuC | 181 | ATG -> GTG | M -> V | N | + | 3-isopropylmalate dehydrogenase large subunit |
| 3976348 | NZ_CP022003.1 | coding | BCA92_20320 | 1523 | CGT -> CCT | R -> p | N | - | methyl-accepting chemotaxis protein II |
| 4629508 | NZ_CP022003.1 | coding | BCA92_23640 | 998 | AAC -> AGC | N -> S | N | + | tRNA uridine-5-carboxymethylaminomethyl(34) synthesis enzyme MnmG |
| 4645676 | NZ_CP022003.1 | coding | BCA92_23640 | 998 | AAC -> AGC | N -> S | N | + | tRNA uridine-5-carboxymethylaminomethyl(34) synthesis enzyme MnmG |
| A3 (17 samples / 23 SNPs) | | | | | | | | | |
| 499284 | NZ_CP022003.1 | coding | BCA92_02430 | 1084 | CTG -> TTG | L -> L | S | + | DUF3971 domain-containing protein |
| 537836 | NZ_CP022003.1 | intergenic | BCA92_02620 | 323 | CGC -> CAC | R -> H | N | - | glutamate synthase small subunit |
| 943216 | NZ_CP022003.1 | intergenic | BCA92_04915 | mutS | 892 | AAC -> AAT | N -> N | S | - | DNA mismatch repair protein MutS |
| 983144 | NZ_CP022003.1 | coding | BCA92_05960 | 312 | CTC -> TTC | L -> F | N | - | SsrA-binding protein |
| 1175265 | NZ_CP022003.1 | coding | BCA92_06430 | 87 | CGT -> CTA | L -> L | S | + | Icsc subfamily cysteine desulfurase |
| 1270601 | NZ_CP022003.1 | coding | BCA92_07010 | 148 | GAT -> TAT | D -> Y | N | + | cytoplasmic protein |
| 1532601 | NZ_CP022003.1 | intergenic | BCA92_10040 | 729 | TTT -> CTT | F -> L | N | - | 505 ribosomal protein L16 arginine hydroxylase |
A3&4&5 (24 samples / 2 SNPs)
337856 NZ_CP022003.1 pseudogene coding BCA92_17470 ushA 491 GTG -> GGG V -> G N - bifunctional UDP-sugar hydrolase/S'-nucleotidase
432515 NZ_CP022003.1 intergenic coding BCA92_17470 ushA 491 GTG -> GGG V -> G N - bifunctional UDP-sugar hydrolase/S'-nucleotidase

A4 (2 samples / 14 SNPs)
272586 NZ_CP022003.1 coding BCA92_01265 1197 GGC -> GGA G -> G S + phosphoesterase PA-phosphatase
366931 NZ_CP022003.1 coding BCA92_01675 1925 GCC -> GTC A -> V N + 4-alpha-glucanotransferase
889829 NZ_CP022003.1 intergenic coding BCA92_06810 tkt 210 TCT -> CCT S -> P N - transketolase
1371258 NZ_CP022003.1 coding BCA92_10055 1191 GCC -> TCC A -> S N - adenylyloucinate lyase
1447810 NZ_CP022003.1 intergenic coding BCA92_07710 302 CCA -> CTA P -> L N + 2-succinyl-6-hydroxy-2, 4-cyclohexadiene-1-carboxylate synthase
1544666 NZ_CP022003.1 coding BCA92_12510 234 CGG -> TGG R -> W N - hypothetical protein
2348301 NZ_CP022003.1 intergenic coding BCA92_10055 1191 GCC -> TCC A -> S N - adenylyloucinate lyase
2745105 NZ_CP022003.1 intergenic coding BCA92_14515 234 CGG -> TGG R -> W N - hypothetical protein
2980813 NZ_CP022003.1 coding BCA92_15430 1526 TTA -> TGA L -> * nonsense - ATP-dependent endonuclease
3172203 NZ_CP022003.1 coding BCA92_16425 995 GCC -> GTC A -> V N + two-component sensor histidine kinase
3760541 NZ_CP022003.1 coding BCA92_19310 ddl 819 ATT -> GTT I -> V N - D-alanine--D-alanine ligase
3791617 NZ_CP022003.1 coding BCA92_19450 530 GAG -> GTG E -> V N - arabinose operon regulatory protein

A5 (5 samples / 6 SNPs)
1143752 NZ_CP022003.1 coding BCA92_05850 94 AGG -> AGT R -> S N - hypothetical protein
1344925 NZ_CP022003.1 coding BCA92_06700 1216 CTG -> CTA L -> L S - beta-barrel assembly-enhancing protease
1629966 NZ_CP022003.1 coding BCA92_08125 1203 GCA -> ACA A -> T N - microcin C ABC transporter ATP-binding protein YejF
3172160 NZ_CP022003.1 coding BCA92_16425 952 ACC -> CCC T -> P N + two-component sensor histidine kinase
3471679 NZ_CP022003.1 coding BCA92_17940 67 TTA -> TTG L -> L S - RNA guanosine(34) transglycosylase Tgt
3775456 NZ_CP022003.1 coding BCA92_19375 256 GGA -> GGT G -> G S - acetalactate synthase small subunit

A6 (4 samples / 31 SNPs)
84482 NZ_CP022003.1 coding BCA92_00390 2709 GCT -> ACT A -> T N - intestinal colonization autotransporter adhesin MisL
534862 NZ_CP022003.1 coding BCA92_02610 28 TCA -> TCG S -> S S - cytosine permease
624925 NZ_CP022003.1 coding BCA92_03070 pfIB 1937 CGC -> CAC R -> H N + formate acetyltransferase
642429 NZ_CP022003.1 coding BCA92_03180 198 TCG -> TCT S -> S S + ribosomal RNA large subunit methyltransferase G
681388 NZ_CP022003.1 coding BCA92_03360 595 CTG -> CTA L -> L S - hypothetical protein
764440 NZ_CP022003.1 coding BCA92_03790 159 GAC -> GAT D -> D S + hypothetical protein
774876 NZ_CP022003.1 coding BCA92_03855 79 GGA -> GGG G -> G S - 16S rRNA (uracil(1498)-N(3))-methyltransferase
812123 NZ_CP022003.1 coding BCA92_04070 76 ACG -> TCG T -> S N + 2-octaprenyl-6-methoxophenyl hydroxylase
1172189 NZ_CP022003.1 coding BCA92_05955 9325 GGT -> GGA G -> G S - Ig-like domain repeat protein
1210085  NZ_CP022003.1  coding  BCA92_06140 870  TTG -> CTG  L -> L  S - protein acetyltransferase
1721423  NZ_CP022003.1  coding  BCA92_08580 1563  CTC -> ATC  L -> I  N - hypothetical protein
1740362  NZ_CP022003.1  coding  BCA92_08645 2453  CAC -> AAC  H -> N  S - diguanylate cyclase/phosphodiesterase
2100643  NZ_CP022003.1  intergenic
2356200  NZ_CP022003.1  coding  BCA92_12000 492  GTC -> ATC  V -> I  N - colanic acid/biofilm transcriptional regulator
2464926  NZ_CP022003.1  coding  BCA92_12610 165  GCC -> GCA  A -> A  S + oxidoreductase
2521018  NZ_CP022003.1  coding  BCA92_12900 1178  GCG -> GAG  A -> E  N + tryptophan synthase subunit beta
2676025  NZ_CP022003.1  intergenic
3109364  NZ_CP022003.1  coding  BCA92_16080 766  GTG -> TTG  V -> L  N + molybdenum-dependent transcriptional regulator
3171171  NZ_CP022003.1  coding  BCA92_16425 503  GAT -> GCT  D -> A  N - two-component sensor histidine kinase
3211370  NZ_CP022003.1  coding  BCA92_16625 110  AGC -> AAC  S -> N  N + glutamate/aspartate ABC transporter substrate-binding protein
3289717  NZ_CP022003.1  coding  BCA92_17010 656  GCC -> GGC  G -> D  N
3353022  NZ_CP022003.1  coding  BCA92_17345 587  ATC -> ACC  I -> T  N - 2-hydroxy-3-oxopropionate reductase
3371497  NZ_CP022003.1  coding  BCA92_17435 385  CCT -> TTC  L -> F  N + paraspin
3378738  NZ_CP022003.1  pseudogene  BCA92_17470  ushA  573  GAG -> TAG  E -> * nonsense - bifunctional UDP-sugar hydrolase/5'-nucleotidase
3556348  NZ_CP022003.1  coding  BCA92_18320 340  CAG -> CAT  Q -> H  N - LysR family transcriptional regulator
3569685  NZ_CP022003.1  coding  BCA92_18395  frsA  701  ACC -> ATC  T -> I  N - esterase
3629990  NZ_CP022003.1  coding  BCA92_18995 109  GTC -> ATC  V -> I  N + RNA 2',3'-cyclic phosphodiesterase
3819853  NZ_CP022003.1  coding  BCA92_19590 5  CAT -> GCT  H -> R  N - MFS transporter
4252918  NZ_CP022003.1  coding  BCA92_21725 10483  GGT -> GGC  G -> G  S - Ig-like domain repeat protein
4299506  NZ_CP022003.1  coding  BCA92_22755 278  GTG -> GAT  G -> D  N + rhamnulokinase

A7 (8 samples / 13 SNPs)
216261  NZ_CP022003.1  coding  BCA92_01020 575  AAC -> AAC  N -> N  Same as reference + fimbrial assembly protein
738389  NZ_CP022003.1  coding  BCA92_03650 695  GTG -> GCT  G -> G  Same as reference + amidohydrolase
900194  NZ_CP022003.1  coding  BCA92_04515 719  GGC -> GGC  G -> G  Same as reference + L-fuculokinase
1392314  NZ_CP022003.1  coding  BCA92_06920 89  CAT -> CAT  H -> H  N Same as reference - hypothetical protein
1662042  NZ_CP022003.1  coding  BCA92_08285 671  GAC -> GAC  D -> D  N Same as reference + DNA-binding transcriptional regulator GalS
1710502  NZ_CP022003.1  coding  BCA92_08850 405  CCT -> CCT  P -> P  N Same as reference + GntR family transcriptional regulator
2972241  NZ_CP022003.1  coding  BCA92_15395 410  GCC -> GCC  G -> G  Same as reference - ATP-dependent Clp protease ATP-binding subunit ClpA
3059259  NZ_CP022003.1  coding  BCA92_15835 1  GTG -> GCT  V -> V  Same as reference - mechanosensitive channel protein
3292066  NZ_CP022003.1  coding  BCA92_17025  entF  329  GCC -> GGC  G -> G  Same as reference - non-ribosomal peptide synthetas
3320204  NZ_CP022003.1  coding  BCA92_17060 326  GCT -> GCT  A -> A  N Same as reference - DNA-binding transcriptional regulator RamA
3400882  NZ_CP022003.1  coding  BCA92_17570 926  CTA -> CTA  L -> L  N Same as reference + efflux transporter periplasmic adaptor subunit
3622599  NZ_CP022003.1  coding  BCA92_18690 75  GAC -> GAC  D -> D  N Same as reference + Rcs stress response system protein RcsF
3917407  NZ_CP022003.1  coding  BCA92_20020 196  CCC -> CCC  P -> P  N Same as reference + fimbrial protein SthA

A7B8 (11 samples / 15SNPs)
47216  NZ_CP022003.1  coding  BCA92_00220 182  TCC -> TCC  S -> S  Same as reference - hypothetical protein
472646  NZ_CP022003.1  coding  BCA92_00225 4  GAC -> GAG  D -> D  N Same as reference + ltb operon leader peptide ltbL
742751  NZ_CP022003.1  intergenic
1061287  NZ_CP022003.1  coding  BCA92_05350 85  TTG -> TTG  L -> L  N Same as reference - alanine--tRNA ligase
1593242  NZ_CP022003.1  coding  BCA92_13275  cydB  580  AGC -> AGC  S -> S  Same as reference + cytochrome d ubiquinol oxidase subunit II
2597837  NZ_CP022003.1  coding  BCA92_13315 828  TGT -> TGT  C -> C  Same as reference + K+/H+ antiporter
2832926  NZ_CP022003.1  coding  BCA92_14720 244  TAT -> TAT  Y -> Y  N Same as reference - protein-disulfide reductase
3064961  NZ_CP022003.1  coding  BCA92_15860 388  AGA -> AGA  R -> R  N Same as reference - ATP-dependent DNA helicase DinG
3124704  NZ_CP022003.1  coding  BCA92_16160 699  GTA -> GTA  V -> V  N Same as reference - cation transporter
3325008  NZ_CP022003.1  coding  BCA92_17205 198  GCC -> GCC  A -> A  N Same as reference - outer membrane usher protein
3378960  NZ_CP022003.1  pseudogene  BCA92_17470  ushA  795  TAG -> TAG  * Same as reference - bifunctional UDP-sugar hydrolase/5'-nucleotidase
3496206  NZ_CP022003.1  coding  BCA92_18055 14  CCC -> CCC  P -> P  N Same as reference - anti-RssB factor
3917922  NZ_CP022003.1  coding  BCA92_20025 95  GGC -> GGC  G -> G  N Same as reference + fimbrial assembly protein
A8 (3 samples / 10 SNPs)

10925  NZ_CP022003.1  coding  BCA92_20260  311  ATC -> ATC  I -> I  Same as reference +  hypothetical protein

149531  NZ_CP022003.1  intergenic

625620  NZ_CP022003.1  coding  BCA92_00725  illdD  679  GGG -> GGA  G -> G  S  -  alpha-hydroxy-acid oxidizing enzyme

730018  NZ_CP022003.1  intergenic

1397217  NZ_CP022003.1  intergenic

2593361  NZ_CP022003.1  coding  BCA92_13290  treA  1091  GAT -> GCT  D -> A  N  +  trehalase

2984205  NZ_CP022003.1  coding  BCA92_15450  351  CTC -> CTT  L -> L  S  +  hybrid-cluster NAD(P)-dependent oxidoeductase

3168124  NZ_CP022003.1  coding  BCA92_16410  1254  GCG -> GCT  A -> A  S  +  potassium-transporting ATPase A chain

4362892  NZ_CP022003.1  coding  BCA92_22115  45  AAC -> GGC  N -> S  N  -  50S ribosomal protein L1

4569415  NZ_CP022003.1  coding  BCA92_23240  rarD  395  GCG -> GTG  A -> V  N  +  protein RarD

B1 (5 samples / 179 SNPs)

46817  NZ_CP022003.1  intergenic

123811  NZ_CP022003.1  coding  BCA92_00595  849  CCG -> CCT  P -> P  S  +  glycosyl transferase

126654  NZ_CP022003.1  coding  BCA92_00610  540  GGG -> GGA  G -> G  S  +  heptose kinase

127012  NZ_CP022003.1  coding  BCA92_00615  50  GCG -> GAG  A -> E  N  +  3-deoxy-D-manno-oct-2ulosonate III transferase WaaZ

175283  NZ_CP022003.1  coding  BCA92_00825  sgbH  61  GGC -> GCT  A -> A  S  +  3-keto-L-gulonate-6-phosphate decarboxylase

175453  NZ_CP022003.1  coding  BCA92_00825  sgbH  231  GGG -> TGG  G -> W  N  -  3-keto-L-gulonate-6-phosphate decarboxylase

185431  NZ_CP022003.1  coding  BCA92_00875  1073  AAC -> AGC  N -> S  N  -  valine--pyruvate transaminase

236543  NZ_CP022003.1  coding  BCA92_01140  bcsA  1185  CCG -> CCT  P -> P  S  -  cellulose synthase catalytic subunit

246329  NZ_CP022003.1  coding  BCA92_01165  1225  CGG -> TGC  R -> C  N  +  phosphodiesterase

261105  NZ_CP022003.1  coding  BCA92_01220  1106  GGG -> GAG  G -> E  N  +  hypothetical protein

325363  NZ_CP022003.1  coding  BCA92_01525  504  CTC -> CTT  L -> L  S  +  pirin family protein

335659  NZ_CP022003.1  coding  BCA92_01560  glgX  490  CCT -> CTT  P -> P  S  +  glycogen debranching enzyme

335522  NZ_CP022003.1  coding  BCA92_01565  glgC  522  GAG -> GAT  E -> D  N  +  glucose-1-phosphate adenylyltransferase

343262  NZ_CP022003.1  coding  BCA92_01590  1141  ACG -> ACA  T -> T  S  -  dihydroxy-acid dehydratase

347430  NZ_CP022003.1  coding  BCA92_01605  715  GGC -> AGC  G -> S  N  +  phosphate ABC transporter substrate-binding protein

390667  NZ_CP022003.1  coding  BCA92_01780  949  GCC -> GCT  A -> A  S  -  carboxypeptidase/penicillin-binding protein 1A

463914  NZ_CP022003.1  coding  BCA92_02250  512  GCG -> GAG  A -> G  N  -  L-threo[6carboxymu]alenolate synthase type 1 TsaC

477827  NZ_CP022003.1  coding  BCA92_02320  581  GTC -> GCC  V -> A  N  +  acrEF/envCd operon transcriptional regulator

511803  NZ_CP022003.1  coding  BCA92_02485  418  CTT -> GTG  L -> V  N  +  GntR family transcriptional regulator

542173  NZ_CP022003.1  coding  BCA92_02625  3232  GCG -> GCA  A -> S  -  glutamate synthase large subunit

555580  NZ_CP022003.1  coding  BCA92_02705  190  GAT -> GAD  D -> E  N  -  3-deoxy-D-manno-octulosonate 8-phosphate phosphatase

556668  NZ_CP022003.1  coding  BCA92_02710  691  GTA -> GTG  V -> V  S  -  arabinoose 5-phosphate isomerase KdsD

694167  NZ_CP022003.1  coding  BCA92_03430  parC  1653  CGT -> GGC  R -> S  S  -  DNA topoisomerase IV subunit A

694941  NZ_CP022003.1  intergenic

700972  NZ_CP022003.1  coding  BCA92_03460  362  CTG -> CGG  L -> R  N  +  YgiQ family radical SAM protein

707698  NZ_CP022003.1  intergenic

712812  NZ_CP022003.1  coding  BCA92_03520  556  AGC -> AGA  S -> R  N  -  NAD[P]-dependent oxidoreductase

724491  NZ_CP022003.1  coding  BCA92_03595  221  CGG -> CTC  R -> I  N  +  hydrogenase 2 accessory protein HypG

765899  NZ_CP022003.1  coding  BCA92_03795  710  GAT -> GGT  D -> G  N  +  L-asparaginase 2

771202  NZ_CP022003.1  coding  BCA92_03830  412  ACG -> GCG  T -> A  N  +  twitching motility protein PilT

837891  NZ_CP022003.1  coding  BCA92_04225  514  CTT -> TTT  L -> F  N  +  porin family protein

856839  NZ_CP022003.1  coding  BCA92_04320  860  CAG -> GGG  Q -> R  N  -  LysF family transcriptional regulator

872799  NZ_CP022003.1  coding  BCA92_04390  621  CAC -> CAT  H -> H  S  +  thymidylate synthase

887476  NZ_CP022003.1  coding  BCA92_04440  660  CTC -> GTC  L -> V  N  -  N-acetylglutamate synthase

962056  NZ_CP022003.1  coding  BCA92_04790  1339  GAA -> TAA  E -> *  nonsense  +  sulfate adenylyltransferase

1011450  NZ_CP022003.1  coding  BCA92_05070  335  TCA -> TTA  S -> L  N  +  chaperone protein SicP

1050530  NZ_CP022003.1  coding  BCA92_05280  214  GGG -> AGG  G -> R  N  +  NorR family transcriptional regulator

1131930  NZ_CP022003.1  intergenic
113290 NZ_CP022003.1 coding BCA92_05765 193 TGC -> CGC C -> G N + Dns family protein
115381 NZ_CP022003.1 coding BCA92_05915 997 TGC -> GGC C -> G N + phage tail tape measure protein
116524 NZ_CP022003.1 coding BCA92_05955 2370 ATC -> CTC I -> L N - lg-like domain repeat protein
116738 NZ_CP022003.1 coding BCA92_05955 4519 TGG -> TGT W -> C N - lg-like domain repeat protein
1186125 NZ_CP022003.1 coding BCA92_06035 trmD 294 GGC -> GGA G -> G S + tRNA (guanosine(37)-N3)-methyltransferase TrmD
1189004 NZ_CP022003.1 intergenic
1189098 NZ_CP022003.1 intergenic
1221880 NZ_CP022003.1 coding BCA92_06205 rpoE 128 TCG -> TTG S -> L N + ECF RNA polymerase sigma-E factor
1255004 NZ_CP022003.1 intergenic
1263328 NZ_CP022003.1 coding BCA92_06390 384 TTT -> TTG F -> L N + nickel transporter
1289858 NZ_CP022003.1 coding BCA92_06505 1912 TCG -> CGG S -> A N + dimethyl sulfoxide reductase subunit A
1365643 NZ_CP022003.1 coding BCA92_06785 671 GCC -> GTG A -> V N + oxidoreductase FeS-binding subunit
1380936 NZ_CP022003.1 coding BCA92_06860 211 CAT -> TAT H -> Y N + ethanolamine utilization protein EutN
1387797 NZ_CP022003.1 coding BCA92_06890 573 GGC -> GGA G -> G S + ethanolamine ammonia-lyase heavy chain
1397007 NZ_CP022003.1 coding BCA92_06950 855 GGA -> GGC G -> G S + iron-dependent peroxidase
1477635 NZ_CP022003.1 intergenic
1485223 NZ_CP022003.1 coding BCA92_07425 1107 TCC -> TCT S -> S S + amidophosphorosyltransferase
1524616 NZ_CP022003.1 coding BCA92_07630 621 TCC -> TCT S -> S S + NADH-quione oxidoreductase subunit F
1564854 NZ_CP022003.1 coding BCA92_07815 632 ACA -> ATA T -> I N - type III secretion system effector deubiquitinase SseL
1612441 NZ_CP022003.1 coding BCA92_08025 936 CGT -> GCC R -> R S + cytochrome c biogenesis protein CcmH
1646199 NZ_CP022003.1 coding BCA92_08210 961 ACC -> GCC T -> A N + PTS fructose transporter subunit EIIBC
1660947 NZ_CP022003.1 coding BCA92_08280 868 CCG -> CGP P -> S N + DUF418 family protein
1697466 NZ_CP022003.1 coding BCA92_08465 130 TTT -> TTC F -> F S - lipoprotein
1722080 NZ_CP022003.1 coding BCA92_08585 98 GGA -> GAA G -> E N - hypothetical protein
1735620 NZ_CP022003.1 coding BCA92_08705 wcaD 483 AAC -> AAA N -> N S + putative colanic acid polymerase WcaD
1771599 NZ_CP022003.1 coding BCA92_08780 rfbB 781 AAC -> ACG N -> H N + dTDP-glucose 4,6-dehydratase
1771638 NZ_CP022003.1 coding BCA92_08780 rfbB 820 TGT -> CTT G -> S S + dTDP-glucose 4,6-dehydratase
1781330 NZ_CP022003.1 coding BCA92_08830 898 AGA -> GGA R -> G N + transporter
1795919 NZ_CP022003.1 coding BCA92_08895 hisA 703 ACC -> ACT T -> T S - carboxamide isomerase
1809044 NZ_CP022003.1 coding BCA92_08950 2335 TCA -> TCG S -> S S - E3 ubiquitin–protein ligase
1809044 NZ_CP022003.1 coding BCA92_08955 57 CCT -> CCC P -> P S + hypothetical protein
1811339 NZ_CP022003.1 coding BCA92_08965 1662 GAA -> GAG E -> E S + thiosulfate reductase
1813310 NZ_CP022003.1 intergenic
1822810 NZ_CP022003.1 coding BCA92_09030 923 GAC -> GCC D -> A N - propanediol utilization protein
1854608 NZ_CP022003.1 coding BCA92_09220 53 N/A N/A N/A N/A trNA-Asn
1891422 NZ_CP022003.1 coding BCA92_09450 25 CCG -> CCA P -> P S - recombinase
1905319 NZ_CP022003.1 coding
1923249 NZ_CP022003.1 coding
1925726 NZ_CP022003.1 coding
1979211 NZ_CP022003.1 coding BCA92_09700 graB 636 GTT -> ATT V -> I S - glutaredoxin 2
1997211 NZ_CP022003.1 coding BCA92_09980 612 CAT -> CAC H -> H N + lipoprotein-releasing system ATP-binding protein LoolD
2005269 NZ_CP022003.1 intergenic
2010154 NZ_CP022003.1 intergenic
2011688 NZ_CP022003.1 coding BCA92_10165 636 GGC -> GGT G -> G S + hypothetical protein
2084889 NZ_CP022003.1 coding BCA92_10605 718 TGG -> TGA W -> * nonsense - L-cystine transporter
2111961 NZ_CP022003.1 coding BCA92_10755 108 ACT -> ACC T -> T S + phosphoenolpyruvate synthase
2127652 NZ_CP022003.1 coding BCA92_10815 469 ATC -> ATA I -> I S - MFS transporter
2137572 NZ_CP022003.1 coding BCA92_11065 267 CCA -> CGC P -> P S + Esc/iYsc/Hrc family type III secretion inner membrane ring protein
2190356 NZ_CP022003.1 coding BCA92_11160 486 AGC -> GCC S -> G N - Bcr/Cfa family drug resistance efflux transporter
2198111 NZ_CP022003.1 coding BCA92_11210 987 CCT -> TCT P -> S N - alkene reductase
2216104 NZ_CP022003.1 coding BCA92_11315 337 GCC -> GCT A -> A S - electron transport complex subunit RsxC
2216131 NZ_CP022003.1 coding BCA92_11315 364 GAT -> GAC D -> D S - electron transport complex subunit RsxC
2216191 NZ_CP022003.1 coding BCA92_11315 424 GCT -> GCC A -> A S - electron transport complex subunit RsxC
232095 NZ_CP022003.1 coding BCA92_11390 182 TGG -> TAG W -> * nonsense - amidohydrolase
2254370 NZ_CP022003.1 coding BCA92_11505 584 CGC -> CTC R -> L N - choline ABC transporter permease
2256276 NZ_CP022003.1 coding BCA92_11520 337 GAT -> GAG D -> E N - DMSO reductase maturation protein DsmD
2274959 NZ_CP022003.1 coding BCA92_11605 1157 GGC -> GAC G -> D N + dipeptidyl carboxypeptidase II
2333624 NZ_CP022003.1 coding BCA92_11905 233 TAC -> TGC Y -> C N + EamA family transporter
2337952 NZ_CP022003.1 intergenic
2371331 NZ_CP022003.1 intergenic
2380433 NZ_CP022003.1 coding BCA92_12115 749 GTG -> GGC V -> A N + hypothetical protein
2438493 NZ_CP022003.1 coding BCA92_12425 327 ACA -> ACG T -> T S + hypothetical protein
2467879 NZ_CP022003.1 coding BCA92_12625 145 GAT -> AAT D -> N N + aromatic alcohol reductase
2496756 NZ_CP022003.1 coding BCA92_12725 72 TGC -> TGT C -> S + osmotically-inducible lipoprotein B
2529046 NZ_CP022003.1 intergenic
2549474 NZ_CP022003.1 ncRNA BCA92_13060 134 N/A N/A N/A N/A
2581346 NZ_CP022003.1 intergenic
2593804 NZ_CP022003.1 coding BCA92_13290 treA 1534 ACC -> GCC T -> A N + trehalase
2595691 NZ_CP022003.1 coding BCA92_13305 emtA 475 GCC -> GCA A -> A S - murine transglycosylase
2625180 NZ_CP022003.1 coding BCA92_13455 1476 GTC -> ATC V -> I N - Terc family protein
2631795 NZ_CP022003.1 coding BCA92_13500 1009 ATC -> ATT I -> I S - cell division protein FtsI
263083 BCA92_13510 110 AAA -> AGA K -> R N - DUF2627 domain-containing protein
2636289 NZ_CP022003.1 coding BCA92_13535 447 GTC -> GTA V -> S + MFS transporter
2663553 NZ_CP022003.1 coding BCA92_13695 290 GGC -> GAC G -> D N - DNA breaking-rejoining protein
2664866 NZ_CP022003.1 coding BCA92_13710 380 GTC -> GAC V -> D N - DUF2514 domain-containing protein
2696527 NZ_CP022003.1 coding BCA92_13930 edd 305 AAA -> AGA K -> R N - phosphogluconate dehydratase
2737304 NZ_CP022003.1 coding BCA92_14130 motB 187 CTG -> CTA L -> L S - flagellar motor protein MotB
2748251 NZ_CP022003.1 intergenic
2749728 NZ_CP022003.1 coding BCA92_14210 123 TGG -> GGG W -> G N - YecA family protein
2753476 NZ_CP022003.1 intergenic
2786139 NZ_CP022003.1 intergenic
2817536 NZ_CP022003.1 coding BCA92_14635 360 TGC -> TGA C -> * nonsense + acetylneuraminic ABC transporter
2877805 NZ_CP022003.1 coding BCA92_14980 59 GCA -> GTA A -> V N - hypothetical protein
2893251 NZ_CP022003.1 coding BCA92_15040 326 GAA -> GCA E -> A N + DUF159 family protein
2975968 NZ_CP022003.1 coding BCA92_15415 835 GGC -> GGT G -> S - macrolide ABC transporter permease/ATP-binding protein MacB
2992160 NZ_CP022003.1 intergenic
3003806 NZ_CP022003.1 coding BCA92_15555 433 CCG -> CCT P -> P S - polyamine ABC transporter ATP-binding protein
3081209 NZ_CP022003.1 coding BCA92_15945 potG 226 TTC -> TTA F -> L N - molybdopterin synthase sulfur carrier subunit
3112990 NZ_CP022003.1 coding BCA92_16100 moaD 759 AAA -> AGA K -> N N + UDP-glucose 4-epimerase
3116229 NZ_CP022003.1 coding BCA92_16145 galE 733 CAT -> TAT H -> Y N + LysR family transcriptional regulator
3175555 NZ_CP022003.1 coding
3191555 NZ_CP022003.1 intergenic
3196084 NZ_CP022003.1 intergenic
3228078 NZ_CP022003.1 coding BCA92_16700 457 GGA -> GCC G -> G S - galactonate dehydratase
3248933 NZ_CP022003.1 intergenic
3251648 NZ_CP022003.1 coding BCA92_16825 dpiB 253 GTG -> GTC V -> V S - histidine kinase
3256983 NZ_CP022003.1 coding BCA92_16845 citF 1258 ATG -> CTG M -> L N + citrate lyase subunit alpha
3347362 NZ_CP022003.1 coding BCA92_17325 587 CCG -> CTG P -> L N - uracil/xanthine transporter
3379696 NZ_CP022003.1 pseudogene BCA92_17470 ushA 1531 GAT -> GAG D -> E N - bifunctional UDP-sugar hydrolase/S'-nucleotidase
3399407 NZ_CP022003.1 coding BCA92_17565 246 AGT -> GGT S -> G N - DNA-binding transcriptional repressor AcrR
3407165 NZ_CP022003.1 coding BCA92_17605 208 AGA -> AGC R -> S N - 50S ribosomal protein L31
3431395 NZ_CP022003.1 intergenic
3439004 NZ_CP022003.1 coding BCA92_17760 57 ATA -> ATG I -> M N + cytochrome c ubiquinol oxidase subunit IV
3511771 NZ_CP022003.1 coding BCA92_18120 prpE 1710 CGC -> TGC R -> E N - propionate--CoA ligase
3513417 NZ_CP022003.1 coding BCA92_18125 1429 ATC -> ATT I -> I S - 2-methylcitrate dehydratase
3561008 NZ_CP022003.1 coding BCA92_18345 13 CAG -> CAA Q -> Q S - MFS transporter
3608574 NZ_CP022003.1 coding BCA92_18615 471 ACC -> GCC T -> A N - class I SAM-dependent methyltransferase
| Accession | Description | Gene | Start (bp) | End (bp) | ORF Direction | NT Change | AA Change | Function |
|-----------|-------------|------|------------|----------|----------------|-----------|-----------|----------|
| B2 (10 samples / 11 SNPs) |
| 1090131 | NCP22003.1 | coding | BCA92_05535 | 193 | GGC -> GCA | A -> A | S | transcriptional regulator |
| 1750882 | NCP22003.1 | coding | BCA92_08690 | 267 | TAC -> TAT | Y -> F | S | - |
| 2070137 | NCP22003.1 | coding | BCA92_10520 | 115 | CGT -> TGT | R -> C | N | - |
| 2371470 | NCP22003.1 | intergenic | BCA92_13245 | 1446 | CTG -> CTA | L -> L | S | hydrogenase 2 large subunit |
| 2748916 | NCP22003.1 | coding | BCA92_14205 | 597 | AGC -> AGT | S -> S | S | - |
| 2818725 | NCP22003.1 | intergenic | BCA92_20815 | 704 | CCA -> CCA | L -> P | N | - |
| 4066198 | NCP22003.1 | coding | BCA92_21680 | 1368 | CTG -> TGT | L -> L | S | - |
| 4233664 | NCP22003.1 | coding | BCA92_21765 | 1758 | TCG -> TCA | S -> S | S | - |
| 4302445 | NCP22003.1 | coding | BCA92_21935 | 630 | GAA -> GAE | E -> E | S | - |
| B3 (2 samples / 23 SNPs) |
| 7774 | NCP22003.1 | coding | BCA92_00030 | 575 | CGC -> CAC | R -> H | N | - |

**Function**: B2 binds DNA and is involved in various cellular processes.
| Gene ID  | Description                              | Start | End   | Amino Acid Change | Function                                      |
|----------|------------------------------------------|-------|-------|-------------------|-----------------------------------------------|
| NZ_CP022003.1 | coding BCA92_00665 | 391   | GGT -> GGC | G -> G | S | - murein hydrolase activator EnvC |
| NZ_CP022003.1 | coding BCA92_00745 | 15    | AAG -> GAG | K -> E | N | - hypothetical protein |
| NZ_CP022003.1 | coding BCA92_03940 | 201   | GGC -> GGA | G -> G | S | + alpha/beta hydrolase |
| NZ_CP022003.1 | intergenic BCA92_04990 | 199   | CAA -> AAA | Q -> K | N | + surface presentation of antigens protein SpaK |
| NZ_CP022003.1 | coding BCA92_08610 | 1348  | CAG -> CAA | Q -> Q | S | - MFS transporter |
| NZ_CP022003.1 | intergenic BCA92_122710 | 124   | TCG -> GCG | S -> A | N | + ABC transporter ATP-binding protein |
| NZ_CP022003.1 | coding BCA92_13245 | 1366  | GCG -> ACG | A -> T | N | + hydrogenase 2 large subunit |
| BCA92_13405 | coding BCA92_02545 | 25    | GCC -> GGC | R -> S | N | + hypothetical protein |
| BCA92_15070 | coding BCA92_18425 | 480   | GCC -> GCT | A -> A | S | + transpeptidase |
| BCA92_19540 | coding BCA92_00480 | 379   | TAC -> TAT | Y -> Y | S | - guanylate kinase |
| BCA92_01315 | coding BCA92_01111 | 890   | CCG -> CTG | P -> L | N | + hypothetical protein |
| BCA92_01895 | coding BCA92_02545 | 55    | TTA -> TAT | D -> A | S | + cell filamentation protein Fic |
| BCA92_03986 | coding BCA92_04045 | 131   | C -> A | A | N/A |
| BCA92_04565 | coding BCA92_05445 | 441   | AAT -> GAT | N -> D | N | - NADPH-dependent 7-cyano-7-deazaguanine reductase QueF |
| BCA92_05445 | coding BCA92_06805 | 212   | GCC -> GTG | A -> V | N | + hypothetical protein |
| BCA92_06895 | coding BCA92_07290 | 1141  | TCC -> TCT | F -> L | N | + acetyl-CoA C-acetyltransferase FadL |
| BCA92_07495 | coding BCA92_08395 | 493   | TCC -> TCT | F -> L | N | + transporter |
| BCA92_08425 | coding BCA92_08453 | 1056  | GTA -> ATA | V -> I | N | - nucleoside permease |
| BCA92_08545 | coding BCA92_09795 | 1569  | GGC -> GTG | G -> G | S | - flagellar hook-associated protein FlgK |
| BCA92_09745 | coding BCA92_10605 | 543   | GTG -> ATG | V -> M | N | - L-cystine transporter |
| BCA92_11245 | coding BCA92_11605 | 112   | GGG -> GCG | G -> G | S | + NAD(P)H transhydrogenase (Re/Si-specific) subunit alpha |
| BCA92_11645 | coding BCA92_13105 | 28    | AAA -> AAG | K -> K | S | - respiratory nitrate reductase subunit gamma |
| BCA92_13190 | coding BCA92_13645 | 301   | GTG -> GTA | V -> V | S | - glutamyl-trNA reductase |
| Gene Accession | Gene Name | Start | End | Strand | Description |
|----------------|-----------|-------|-----|--------|-------------|
| NZ_CP022003.1 | tolA      | 631   | V   | -      | cell envelope integrity protein TolA |
| NZ_CP022003.1 | tolA      | 632   | A   | -      | cell envelope integrity protein TolA |
| NZ_CP022003.1 | BC92_16260| 56    | CCG | -      | two-component sensor histidine kinase |
| NZ_CP022003.1 | leuS      | 1228  | AAA | K      | leucine--tRNA ligase |
| NZ_CP022003.1 | BC92_17425| 284   | CCG | -      | iron export ABC transporter permease subunit FetB |
| NZ_CP022003.1 | BC92_17470| 519   | GCC | -      | bifunctional UDP-sugar hydrolase/5'-nucleotidase |
| NZ_CP022003.1 | BC92_17635| 1022  | GCC | -      | ammonium transporter |
| NZ_CP022003.1 | BC92_18190| 700   | CGC | R      | RIP metalloprotease RseP |
| NZ_CP022003.1 | BC92_18795| 231   | GAG | *      | L-ribulose-5-phosphate 4-epimerase |
| NZ_CP022003.1 | BC92_19680| 422   | ACC | M      | phosphoribosyl-dephospho-CoA transferase |
| NZ_CP022003.1 | BC92_19690| 705   | GCC | -      | citrate lyase subunit beta |
| NZ_CP022003.1 | BC92_20320| 342   | ACC | -      | anaerobic ribonucleoside triphosphate reductase |
| NZ_CP022003.1 | BC92_20385| 1685  | GCC | -      | type I restriction endonuclease |
| NZ_CP022003.1 | BC92_20425| 552   | TTA | -      | MFS transporter |
| NZ_CP022003.1 | BC92_20825| 1609  | TAC | -      | anaerobic sulfite reductase subunit A |
| NZ_CP022003.1 | BC92_21080| 612   | GAT | Y      | bifunctional isocitrate dehydrogenase kinase/phosphatase |
| NZ_CP022003.1 | BC92_22020| 602   | GGT | A      | ribosome maturation factor |
| NZ_CP022003.1 | BC92_22190| 46    | CTG | -      | DNA-directed RNA polymerase subunit beta |
| NZ_CP022003.1 | BC92_23470| 1243  | TCC | -      | ketol-acid reductoisomerase |