http://www.egejfas.org
Ege Journal of Fisheries and Aquatic Sciences, 43(3), 216-225 (2026)

Su Ürünleri Dergisi (2026)
DOI: 10.12714/egejfas.43.3.04

RESEARCH ARTICLE ARAŞTIRMA MAKALESİ

Identification, characterization and draft genome analysis of *Vibrio anguillarum* VK1 isolated from hatchery water

Kuluçkahane suyundan izole edilen *Vibrio anguillarum* VK1'in identifikasyonu, karakterizasyonu ve taslak genom analizi

Kadriye Toklu[1] ● **İhsan Yaşa**[2*]

[1]Department of Biotechnology, Graduate School of Natural and Applied Sciences, Ege University, Bornova, İzmir, Türkiye
[2]Department of Biology, Faculty of Science, Ege University, Bornova, İzmir, Türkiye

*Corresponding author: ihsan.yasa@ege.edu.tr
Received date: 24.04.2026 Accepted date: 12.05.2026

How to cite this paper:
Toklu, K., & Yaşa, İ. (2026).Identification, characterization and draft genome analysis of *Vibrio anguillarum* VK1 isolated from hatchery water. *Ege Journal of Fisheries and Aquatic Sciences, 43*(3), 216-225. https://doi.org/10.12714/egejfas.43.3.04

Abstract: Aquaculture environments are dynamic ecosystems that may influence the evolution and adaptive potential of bacterial pathogens. In this study, a locally isolated *Vibrio anguillarum* strain VK1 (PRJNA1452808) from Türkiye was characterized using 16S rDNA phylogenetic analysis and whole-genome-based approaches. The high-quality draft genome (4.23 Mb; GC content 44.31%) comprised 194 contigs and 4,102 coding sequences, including 1,057 hypothetical proteins. Functional analyses revealed putative stress response mechanisms and iron acquisition systems, suggesting metabolic versatility and environmental adaptability. Genome-based screening identified 33 putative antimicrobial resistance-associated determinants, including multidrug efflux-related genes and genes linked to antibiotics commonly used in aquaculture. In addition, genes associated with Type III and Type VI secretion systems were detected, potentially indicating adaptive traits related to host interaction and environmental persistence. Overall, the findings suggest that strain VK1 possesses a flexible genomic structure that may support adaptation to fluctuating aquaculture conditions. However, further functional studies are needed to biologically validate the potential virulence and AMR related features.

Keywords: Aquaculture, vibriosis, next generation sequencing, virulence factors

Öz: Su ürünleri yetiştiriciliği ortamları bakteriyel patojenlerin evrimini ve adaptasyon potansiyelini etkileyebilecek dinamik ekosistemlerdir. Bu çalışmada, Türkiye'den yerel olarak izole edilen *Vibrio anguillarum* VK1 suşu (PRJNA1452808), 16S rDNA filogenetik analizi ve tüm genom tabanlı yaklaşımlar kullanılarak karakterize edilmiştir. Yüksek kaliteli taslak genom (4,23 Mb; GC içeriği %44,31), 194 kontig ve 4.102 kodlama dizisi içermekte olup, bunların arasında 1.057 hipotetik protein bulunmaktadır. Fonksiyonel analizler, olası stres yanıt mekanizmalarını ve demir edinim sistemlerini ortaya çıkararak metabolik çok yönlülük ve çevresel adaptasyon yeteneğini işaret etmektedir. Genom tabanlı tarama, çoklu ilaç atılımı ile ilgili genler ve su ürünleri yetiştiriciliğinde yaygın olarak kullanılan antibiyotiklerle bağlantılı genler de dahil olmak üzere 33 olası antimikrobiyal dirençle ilişkili belirleyiciyi tanımlamıştır. Ayrıca, konakçı etkileşimi ve çevresel kalıcılıkla ilgili adaptif özellikleri potansiyel olarak gösteren Tip III ve Tip VI salgılama sistemleriyle ilişkili genler tespit edilmiştir. Genel olarak, bulgular VK1 suşunun, değişken akuakültür ortamlarına adaptasyonu destekleyebilecek esnek bir genomik yapıya sahip olduğunu göstermektedir. Bununla birlikte, olası virülans ve AMR ile ilişkili özelliklerin biyolojik doğrulaması için daha fazla fonksiyonel çalışmaya ihtiyaç duyulmaktadır.

Anahtar kelimeler: Su ürünleri yetiştiriciliği, vibriozis, yeni nesil sekanslama, virülans faktörleri

INTRODUCTION

The rapid growth of the global population, coupled with the increasing demand for animal-derived protein, has driven the expansion of aquaculture worldwide. This intensification has led to the widespread adoption of large-scale seafood farming systems. However, alongside these developments, the prevalence of infectious diseases has risen markedly, posing a critical constraint on sustainable aquaculture production (Sampantamit et al., 2020). Among the bacterial pathogens affecting aquaculture, *Vibrio anguillarum* stands out as a significant threat, particularly due to its impact on economically important fish species (Bergman, 1909). Infections caused by this pathogen result in substantial economic losses across the industry (Yang et al., 2023). The disease, known as vibriosis, typically presents as hemorrhagic septicemia and is characterized by clinical manifestations such as skin ulcerations, hepatomegaly, and the accumulation of ascitic fluid (Guanhua et al., 2018). To date, 23 O serotypes of *V. anguillarum* have been identified based on serological classification. Nevertheless, infections in aquatic organisms are predominantly associated with the O1 and O2 serotypes, which are considered the most virulent and epidemiologically relevant (Hu et al., 2023).

A broad repertoire of virulence factors has been described in *V. anguillarum*. These include traits such as motility and chemotaxis, lipopolysaccharides, extracellular products exhibiting hemolytic and proteolytic activities, and multiple iron acquisition systems (Lages et al., 2019). Among the key determinants of virulence are the metalloprotease EmpA, hemolysins (Vah1–5 and Rtx), and iron uptake mechanisms. These systems encompass pathways involved in iron metabolism, transport of ferrous iron via the feoABC operon, and genes associated with three major siderophore systems—vanchrobactin, anguibactin, and piscibactin (Gao et al., 2018). Comparative genomic analyses of fifteen *V. anguillarum* isolates representing O1, O2, and O3 serotypes have demonstrated that putative virulence factors are highly

© Published by Ege University Faculty of Fisheries, İzmir, Türkiye

conserved across 775 reference strains, indicating limited genetic diversity. These observations suggest that pathogenicity in this species is likely governed by a multifactorial framework (Busschaert et al., 2015). However, experimental studies employing gnotobiotic European perch larvae as model hosts have not identified a clear link between genotype and virulence phenotype, underscoring the complexity of host–pathogen interactions (Frans et al., 2013). Moreover, the limited availability of reliable biomarkers, together with low interstrain genetic variability, complicates the accurate identification of highly pathogenic strains and hinders the development of functional disease control strategies in aquaculture. These challenges highlight the need for comprehensive genomic investigations supported by standardized virulence databases to better elucidate the evolutionary dynamics, physiology, and pathogenic mechanisms of *V. anguillarum* (Castillo et al., 2017).

Whole-genome sequencing (WGS) stands out as a powerful and reliable approach that enables the high-resolution study of bacterial genetic diversity. WGS allows for comprehensive analysis of multidimensional genomic features of pathogens, such as molecular epidemiology, evolutionary processes, virulence, and antibiotic resistance genes (Quainoo et al., 2017). While traditional methods of examining virulence genes with a limited number of targeted PCR amplifications cannot provide a holistic assessment due to low throughput and limited scope, the different pathogenicity levels observed among strains carrying the same virulence genes demonstrate that the presence of genes alone is insufficient to explain pathogenesis mechanisms (Zautner et al., 2017).

The genomic repertoire of a local *V. anguillarum* strain (VK1), isolated from Aydın, Türkiye, is characterized here for the first time. The strain was identified using 16S rRNA analysis and further investigated through whole-genome sequencing to elucidate its metabolic potential, stress response capacity, virulence-associated traits, and antimicrobial resistance (AMR) profile.

MATERIALS AND METHODS

Bacterial isolation

Water samples were collected in 2022 from a marine aquaculture facility located in Didim district, Aydın, Türkiye, during routine microbiological monitoring. Samples were transported to the laboratory under cold-chain conditions and processed on the same day. For bacterial isolation, seawater samples were subjected to membrane filtration, and the filters were plated onto Thiosulfate Citrate Bile Salts Sucrose (TCBS) agar followed by incubation at 27°C for 24 h (Lotz et al., 1983). Predominant colonies were subsequently subcultured on TCBS agar to ensure purity and preserved at −80°C in 30% glycerol for long-term storage. More than 20 presumptive *Vibrio anguillarum* isolates were obtained during the study. Among these isolates, strain VK1 was selected for whole-genome sequencing and detailed genomic characterization based on preliminary microbiological and phenotypic evaluations (Toklu et al., 2022).

16S rDNA amplification and phylogenetic analysis

Genomic DNA was purified from overnight active cultures using a commercial kit (Roche, High Pure PCR Template Preparation Kit) according to the manufacturer's instructions. PCR amplification was performed in a 50 µL reaction mixture using the primers 27F (5'-AGAGTTGATCCTGGCTCAG-3') and 1492R (5'-TACGGCTACCTTGTACGACTT-3'), targeting the 16S rRNA gene region (Frank et al., 2008). The thermal cycling conditions included an initial denaturation at 95°C for 5 min, followed by 30 cycles of denaturation at 94°C for 30 s, annealing at 58°C for 30 s, and extension at 72°C for 40 s, with a final extension at 72°C for 10 min. The PCR products were sequenced by Letgen Biotech Co., Ltd. The obtained sequences were compared with reference 16S rRNA sequences in the GenBank database using the NCBI-BLAST tool. Phylogenetic relationships were inferred by constructing a tree in MEGA11 using the Neighbor-Joining method (Tamura et al., 2021).

Genome sequencing and functional annotation

Genomic DNA of strain VK1 was sequenced by Refgen Biotechnology Inc. (Ankara, Turkey) using the Illumina NovaSeq 6000 platform with paired-end reads. Raw sequencing reads were subjected to quality control and trimming prior to assembly. De novo genome assembly was performed using Shovill (v1.1.0.), generating a draft genome composed of contigs. Assembly quality and genome completeness were evaluated using CheckM (v1.2.5). Gene prediction and functional annotation of coding sequences (CDSs) were performed using RAST, PATRIC, and Prokka (v1.14) to ensure comprehensive genome characterization (Eren Eroglu et al., 2025a). AMR genes were identified using ABRicate (v0.5) by screening against the ResFinder and CARD databases, applying a minimum threshold of 90% sequence identity and 90% coverage. Virulence-associated genes were detected and functionally classified based on annotations from the PATRIC and KEGG databases. The average nucleotide identity (ANI) between VK1 and the reference strain was calculated using FastANI, based on reciprocal genome fragment comparisons (Stothard et al., 2019). Taxonomic classification was further supported by the Type (Strain) Genome Server (TYGS; http://tygs.dsmz.de/), using default parameters (Meier-Kolthoff and Göker, 2019). The draft genome sequence has been deposited in GenBank under the accession number PRJNA1452808.

Pan genome based comparative analysis

Comparative pan-genome analysis of *V. anguillarum* strains from different geographical origins was performed using Panaroo v1.5.2 with default settings and a protein clustering threshold of 95% sequence identity. The analysis included strain VK1 from Türkiye together with publicly available genomes of *V. anguillarum* strain 100719-1/3A, *V. anguillarum* strain 090707-1/1A, *V. anguillarum* strain LMG 4437, and *V.*

anguillarum strain LMG 13583. Genome annotations generated by Prokka were used as input files for the analysis. Genes identified in the pan-genome were categorized as core genes (shared by all analyzed strains), accessory genes (present in two or more but not all strains), and strain-specific genes (unique to a single strain) (Tonkin-Hill et al., 2020).

RESULTS

Preliminary identification of the VK1 strain

The VK1 strain was isolated using the membrane filtration method (Figure 1A). Following incubation, distinct colonies were observed and subsequently subcultured to obtain pure isolates. When cultured on TCBS agar, the colonies appeared yellow, circular, slightly raised, and exhibited smooth margins (Figure 1B), consistent with sucrose-fermenting *Vibrio* species.

16S rRNA gene analysis showed that the VK1 isolate showed the highest similarity (98.42%) to *V. anguillarum* strain NCIMB (Figure 2), indicating that the isolate is closely related to the *V. anguillarum* group.

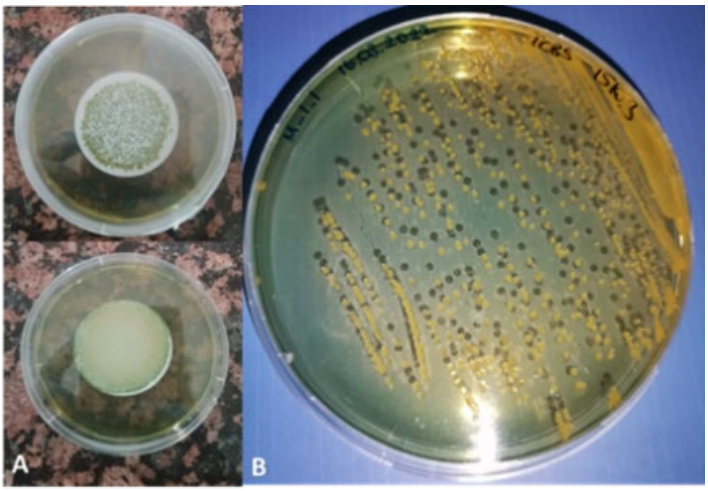

Figure 1. (A) Membrane filtration of the water sample showing bacterial retention on the filter surface. (B) Pure culture of the VK1 isolate grown on TCBS agar, displaying characteristic colony morphology.

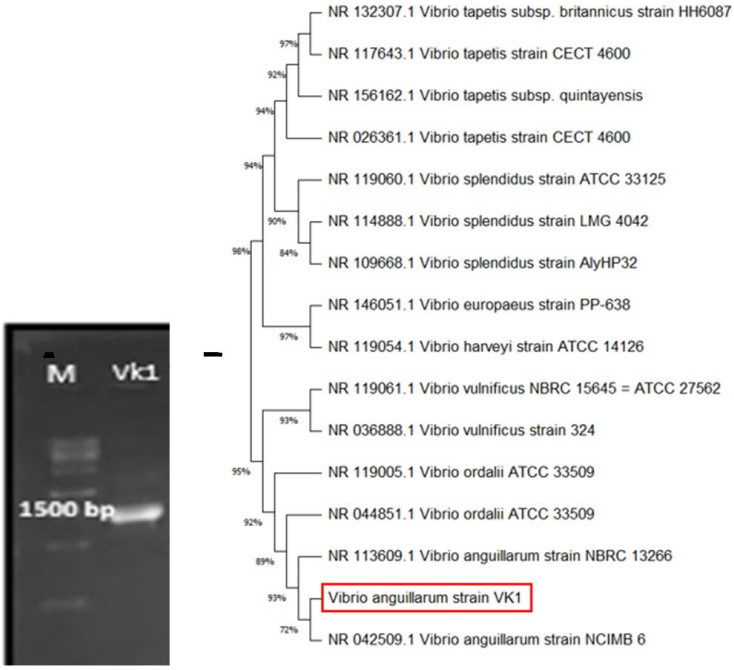

Figure 2. (A) Agarose gel electrophoresis image of the amplified 16S rRNA gene from the VK1 isolate, showing the expected ~1500 bp PCR product. M: DNA marker. (B) Phylogenetic tree based on 16S rRNA gene sequences, constructed using the Neighbor-Joining method in MEGA11, illustrating the relationship of the VK1 isolate with closely related *Vibrio* species.

Genome characteristics of VK1

Genome analysis of the VK1 strain revealed a high-quality draft genome with strong completeness. The assembled genome has a total length of 4,230,312 bp with a GC content of 44.31%, distributed across 194 contigs (Table 1). Functional annotation identified 4,102 coding sequences (CDSs), along with 100 tRNA and 8 rRNA genes. Among the CDSs, 1,057 (33.59%) were annotated as hypothetical proteins, while 3,959 (96.51%) were assigned to protein families (PLFAMs). Genome quality assessment indicated high consistency, with coarse and fine consistency values of 99.4% and 98.7%, respectively. Average nucleotide identity (ANI) analysis demonstrated that the VK1 genome shares 98.58% similarity with *V. anguillarum* PF4-E1-1 (reference strain; GCA_003390515.1). A total of 1,391 genome fragments were compared, of which 1,148 were identified as orthologous matches (Figure 3). This high ANI value strongly supports the classification of the VK1 strain within the same species as the reference genome. ANI values calculated between strain VK1 and the other reference strains ranged from 98.03% to 99.96% and are presented in Supplementary Table S1. Consistent with the ANI analysis, the whole-genome phylogenomic tree generated using TYGS also placed strain VK1 within the *V. anguillarum* cluster together with closely related reference strains (Figure 4).

Table 1. Genome characteristics of the VK1 strain

Category	Metric	Value
Assembly Statistics	Contigs	194
	Genome Length (bp)	4,230,312
	GC Content (%)	44.31
	Contig L50	11
	Contig N50 (bp)	110,506
	tRNA	100
Annotation Statistics	rRNA	8
	CDS	4,102
	CDS Ratio	0.9697
	Hypothetical CDS	1,057
	Hypothetical CDS Ratio	0.3359
	PLFAM CDS	3,959
	PLFAM CDS Ratio	0.9651
	Coarse Consistency (%)	99.4
Genome Quality	Fine Consistency (%)	98.7
	CheckM Completeness (%)	100

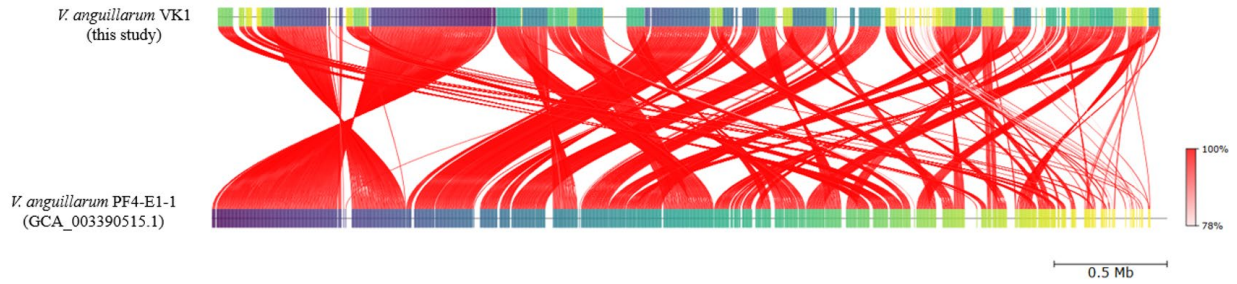

Figure 3. Reciprocal mapping between the VK1 genome (query) and the reference genome (PF4-E1-1), showing the alignment of orthologous genomic regions used for ANI calculation.

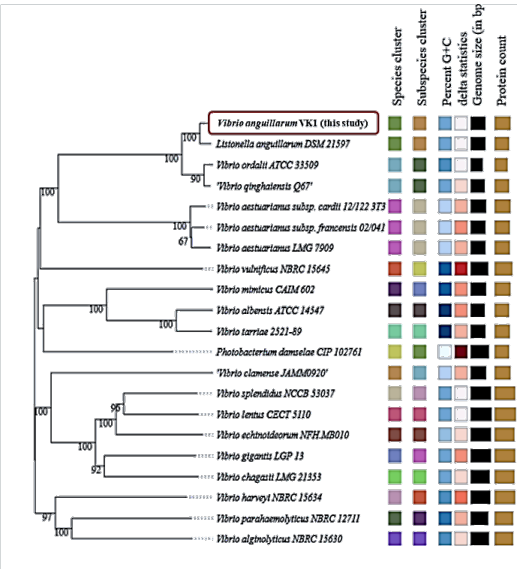

Figure 4. Phylogenomic tree based on whole-genome comparisons showing the relationship of strain VK1 with representative members of the genus *Vibrio*. Strain VK1 clustered closely with *Vibrio* reference strains, supporting its taxonomic assignment. Bootstrap values (%) are shown at branch nodes, and colored boxes indicate comparative genomic features including GC content, genome size, and protein count.

Functional annotation of the VK1 genome

According to RAST annotation, genes identified in the VK1 genome were classified into multiple subsystem categories representing diverse biological functions (Figure 5A). The most abundant categories were associated with core cellular processes, including protein metabolism, amino acids and derivatives, carbohydrate metabolism, and membrane transport. In addition, genes related to stress response, virulence, disease and defense, lipid metabolism, and nucleotide metabolism were also detected, indicating a broad functional repertoire.

Analysis based on the PATRIC database revealed that stress-related genes were distributed across several subcategories (Figure 5B). The largest proportion corresponded to genes grouped under "no subcategory," representing general stress response functions, followed by genes associated with oxidative stress, detoxification, osmotic stress, heat shock, cold shock, and periplasmic stress. A detailed list of stress-related genes is provided in the Table S2.

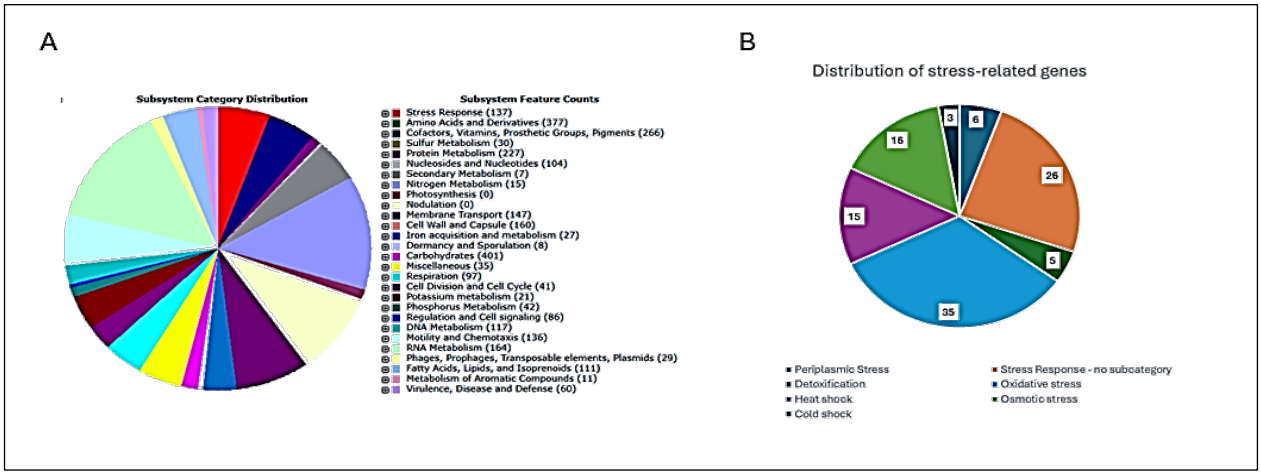

Figure 5. (A) Subsystem category distribution of the VK1 genome based on RAST annotation, showing the functional classification of genes across major metabolic and cellular processes. (B) Distribution of stress-related genes in the VK1 genome according to PATRIC annotation, categorized into different stress response groups.

In silico prediction of AMR-related genes

Genome analysis of the VK1 isolate identified a total of 33 genes/proteins associated with different classes of antibiotics (Table 2). These genes include targets involved in cell wall synthesis (e.g., *murA, ddl, alr, kasA*), DNA replication and transcription *(gyrA*, g*yrB, rpoB, rpoC, rho*), protein synthesis (*EF-Tu, EF-G, S12p, S10p, Iso-tRNA*), folate metabolism (*folA*/*dfr, folP*), lipid biosynthesis (*fabV, pgsA*), and oxidative stress response (*katG, oxyR*), as well as other metabolic processes. In addition, genes associated with antibiotic resistance mechanisms, including efflux and target protection systems (*tet(35), qnrB, tolC*/*opmH*), were also detected.

Virulence-associated genes

Genome analysis of the VK1 isolate identified a total of 71 virulence-associated genes, which were classified into different functional categories (Table S3). These genes include those related to adherence (*pilC*, *pilB*, *mshE*, *tufA*, IlpA), biofilm formation (*luxS*), effector delivery systems (components of T3SS and T6SS, as well as general secretion pathway proteins; e.g., *hcp-1*, *vscN2, vas, vip, icmF, eps* genes), immune modulation (*gmhA*/*lpcA*), motility (numerous genes associated with flagellar structure, biosynthesis, and chemotaxis; e.g., *fli, flg, flh, fla, mot, che* genes), and regulation (*csrA, rpoS*). Among these categories, motility-related genes represented the largest group, followed by genes associated with effector delivery systems and adherence.

Comparative pan genome analysis

Pan-genome analysis of *V. anguillarum* strains originating from Denmark, Norway, the USA, and the Turkish isolate analyzed in this study revealed 2,869 core genes, 1,082 accessory genes, and 2,444 strain-specific genes. Accessory gene-based phylogenetic analysis indicated that strain VK1 clustered more closely with the USA isolate (Figure 6A and 6B). Furthermore, a total of 310 genes unique to VK1 and absent in the other compared strains were identified across different COG functional categories. These strain-specific genes were mainly associated with mobile genetic elements, stress adaptation, transport systems, and regulatory functions. The identified genes included transposases, integrases, recombinases, phage-related proteins, multidrug efflux pumps, chemotaxis-associated proteins, DNA helicases, and transcriptional regulators. In addition, genes related to secretion systems, iron acquisition, toxin-associated proteins, and signal transduction mechanisms were also detected (Table S4).

Table 2. Predicted antimicrobial resistance-related genes identified in the VK1 genome and their associated antibiotic classes.

	Genes	Products	Related Antibiotics class
1	*KatG*	Catalase-peroxidase KatG (EC 1.11.1.21)	Isoniazid
2	*Iso-tRNA*	Isoleucyl-tRNA synthetase (EC 6.1.1.5)	Mupirocin
3	*EF-Tu*	Translation elongation factor Tu	Elfamycins
4	*EF-Tu*	Translation elongation factor Tu	Elfamycins
5	*Alr*	Alanine racemase (EC 5.1.1.1)	Cycloserine
6	*gyrB*	DNA gyrase subunit B (EC 5.99.1.3)	Fluoroquinolones, Quinolones, Quinolines, Aminocoumarin antibiotics
7	*dxr*	1-deoxy-D-xylulose 5-phosphate reductoisomerase	Fosmidomycin
8	*EF-Tu*	Translation elongation factor Tu	Elfamycins
9	*MurA*	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (EC 2.5.1.7)	Fosfomycin
10	*S12p*	SSU ribosomal protein S12p (S23e)	Aminoglycosides
11	*folA, Dfr*	Dihydrofolate reductase (EC 1.5.1.3)	Diaminopyrimidines
12	*gyrA*	DNA gyrase subunit A (EC 5.99.1.3)	Fluoroquinolones, Quinolones, Quinolines
13	*rpoC*	DNA-directed RNA polymerase beta' subunit (EC 2.7.7.6)	Myxopyronins, Corallopyronins, Peptide antibiotics
14	*folP*	Dihydropteroate synthase (EC 2.5.1.15)	Sulfonamides
15	*kasA*	3-oxoacyl-[acyl-carrier-protein] synthase, KASII (EC 2.3.1.179)	Isoniazid, Triclosan
16	*rpoB*	DNA-directed RNA polymerase beta subunit (EC 2.7.7.6)	Rifamycins, Peptide antibiotics
17	*rho*	Transcription termination factor Rho	Bicyclomycins
18	*EF-Tu*	Translation elongation factor Tu	Elfamycins
19	*EF-G*	Translation elongation factor G	Fusidic acid
20	*Ddl*	D-alanine--D-alanine ligase (EC 6.3.2.4)	Cycloserine
21	*S10p*	SSU ribosomal protein S10p (S20e)	Tetracyclines, Glycylcyclines
22	*EF-Tu*	Translation elongation factor Tu	Elfamycins
23	*EF-Tu*	Translation elongation factor Tu	Elfamycins
24	*QnrB family*	Pentapeptide repeat protein QnrB family	Fluoroquinolones, Quinolones, Quinolines
25	*fabV*	Enoyl-[acyl-carrier-protein] reductase [NADH] (EC 1.3.1.9), FabV => refractory to triclosan	Triclosan
26	*fabV*	Enoyl-[acyl-carrier-protein] reductase [NADH] (EC 1.3.1.9), FabV => refractory to triclosan	Triclosan
27	*TolC/OpmH*	Outer membrane channel TolC (OpmH)	Penams, Cephamycins, Cephalosporins, Monobactams, Tetracyclines, Rifamycins, Phenicol antibiotics, Aminocoumarin antibiotics, Glycylcyclines, Fluoroquinolones Quinolones, Quinolines, Triclosan, Macrolides, Glycylcyclines
28	*Tet(35)*	Tetracycline resistance, MFS efflux pump => Tet(35)	Tetracyclines
29	*gidB*	16S rRNA (guanine(527)-N(7))-methyltransferase (EC 2.1.1.170)	Aminoglycosides
30	*GdpD*	Glycerophosphoryl diester phosphodiesterase (EC 3.1.4.46)	Peptide antibiotics
31	*PgsA*	CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase (EC 2.7.8.5)	Peptide antibiotics
32	*H-NS*	DNA-binding protein H-NS	Penams, Cephamycins, Cephalosporins, Fluoroquinolones Quinolones Quinolines, Macrolides, Tetracyclines
33	*OxyR*	Hydrogen peroxide-inducible genes activator => OxyR	Isoniazid

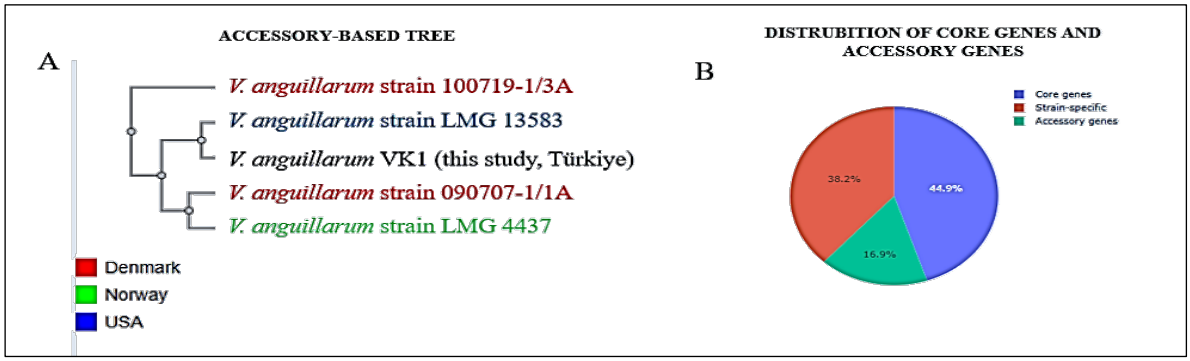

Figure 6. Pan-genome analysis of *V. anguillarum* strains from different geographical origins. (A) Accessory gene-based phylogenetic tree showing the relationship of strain VK1 with representative isolates from Denmark, Norway, and the USA. Strain VK1 clustered more closely with the USA isolate based on accessory genome composition. (B) Distribution of core, accessory, and strain-specific genes identified among the analyzed genomes.

DISCUSSION

Vibriosis remains one of the most prevalent bacterial diseases in aquaculture, with infections caused by *Vibrio* species leading to high mortality rates and significant economic losses (Sanches-Fernandes et al., 2022; Mohd et al., 2021). In this context, understanding the phenotypic traits and intraspecific genomic diversity of members of the Vibrionaceae family is critical for elucidating disease dynamics and improving control strategies (Bondad-Reantaso et al., 2023).

In the present study, genome analysis of the VK1 strain revealed a high-quality draft genome, as indicated by completeness and consistency values exceeding 99%, supporting the reliability of downstream genomic analyses (Parks et al., 2015). ANI analysis demonstrated that strain VK1 shared more than 98% similarity with reference strains, exceeding the widely accepted species delineation threshold of 95–96%, thereby confirming its taxonomic classification (Murray et al., 2021). In addition, the genome size and GC content (44.31%) are consistent with previously reported *V. anguillarum* genomes, suggesting genomic stability within the species (Naka et al., 2011; Vasquez et al., 2020). Moreover, approximately one-third of the coding sequences were annotated as hypothetical proteins, indicating the presence of a substantial proportion of uncharacterized genes. This highlights the potential for uncovering novel functions, particularly those related to environmental adaptation, virulence, or antimicrobial resistance (Prasad et al., 2025).

RAST-based subsystem profiling revealed a predominance of genes associated with core metabolic functions in the VK1 genome, particularly those involved in protein, amino acid, and carbohydrate metabolism. This enrichment suggests a high degree of metabolic versatility, likely underpinning rapid adaptation to fluctuating environmental conditions—a key determinant of survival and competitiveness in dynamic aquaculture systems (Liu et al., 2021; Phurahong et al., 2024).

Consistent with this, PATRIC-based functional insights indicate that VK1 harbors a broad repertoire of stress response mechanisms. The presence of genes linked to oxidative stress and detoxification highlights a potential capacity to withstand host-derived reactive oxygen species (Gao et al., 2018). Notably, iron limitation—a hallmark of infection sites often coupled with oxidative stress—appears to be counteracted by Fur-regulated heme uptake systems (huvA-X, tonB1-exbB1D1-huvBCD). While Fur represses these systems under iron-replete conditions, derepression during infection likely facilitates host heme acquisition, thereby sustaining respiration and enhancing virulence (Crisafi et al., 2014). In parallel, siderophore-mediated iron acquisition, particularly under variable temperature regimes, represents a well-established virulence determinant in *Vibrio* sp. (Eren Eroğlu et al., 2025b). The detection of genes associated with osmotic stress, as well as heat and cold shock responses, further supports the ecological fitness of VK1 in marine environments characterized by pronounced fluctuations in temperature and salinity (Lovell, 2017).

The VK1 genome contained 33 putative AMR associated determinants, indicating the presence of genes potentially involved in antimicrobial tolerance and cellular detoxification processes in aquatic environments. Among these, the *TolC/OpmH* efflux system was identified, which has previously been associated with the export of multiple antibiotics, including penams, cephalosporins, tetracyclines, and macrolides, as well as environmental toxic compounds (Soto and Nishiguchi, 2021). In addition, *tet(35)*- and *qnrB*-like genes were detected, which may be related to resistance against tetracyclines and fluoroquinolones commonly used in aquaculture settings (Tang et al., 2025). Several intrinsic resistance-associated genes, including *murA, ddl, alr, kasA, gyrA,* and *gyrB*, were also identified and are known to be linked to resistance mechanisms targeting different antibiotic classes (Rodkhum et al., 2008; AlJindan et al., 2023). Furthermore, the presence of mobile genetic element-associated genes suggests the potential for horizontal gene transfer in aquatic environments (Reverter et al., 2025).

Virulence genes encode factors that may contribute to bacterial colonization, persistence, and host interaction, and both canonical and non-canonical virulence-associated determinants can co-exist within pathogenic strains (Schroeder et al., 2017). In this context, genomic analysis of strain VK1 identified several putative virulence-associated genes, suggesting that its pathogenic potential may be linked not only to classical cytotoxic mechanisms but also to metabolic adaptation and regulatory flexibility. Consistent with this observation, previous studies reported that *V. anguillarum* exhibits temperature-dependent modulation of virulence, with increased virulence observed at 15°C despite an optimal growth temperature of approximately 25°C. Transcriptomic analyses further demonstrated that virulence-related factors, including siderophore systems and toxin-associated genes, may be differentially regulated depending on temperature and iron availability, highlighting the influence of environmental conditions on pathogenic behavior (Lages et al., 2019).

The VK1 genome contained several putative virulence-associated determinants, including genes related to the Type III (T3SS) and Type VI (T6SS) secretion systems. Among these, *vscN2* and *hcp-1* genes were identified, which are known to be associated with secretion system functions in pathogenic bacteria. The *vscN2* gene encodes an ATPase involved in the translocation of effector proteins through the T3SS, while *hcp-1* is associated with the tubular structure of the T6SS involved in interbacterial and host interactions (Park et al., 2004; Sha et al., 2013). The presence of both secretion system-associated genes may contribute to environmental adaptation and host interaction capabilities in strain VK1. Similar secretion system components have previously been reported in pathogenic strains associated with aquaculture environments (Guillemette et al., 2020).

Comparative pangenome analysis revealed considerable genomic diversity among the analyzed *V. anguillarum* strains originating from different geographical regions. The presence of 2,444 strain-specific genes together with 1,082 accessory genes supports the view that *V. anguillarum* possesses a highly dynamic genome structure shaped by environmental adaptation and genome plasticity. The identification of 310 VK1 specific genes, including transposases, integrases, recombinases, and phage-related proteins, further suggests the contribution of HGT and mobile genetic elements to genomic diversification within this species. Previous studies have similarly highlighted the role of insertion sequences, prophage-associated regions, and genomic islands in shaping intra-species evolution and genomic heterogeneity in *V. anguillarum* strains (Castillo et al., 2017; Machimbirike et al., 2023). In addition, the detection of genes associated with secretion systems, iron acquisition, multidrug efflux pumps, and regulatory functions suggests that strain VK1 may harbor adaptive traits potentially related to environmental persistence and host interaction. Similar virulence- and adaptation-associated accessory genes, including iron uptake systems and transport-related determinants, have also been reported in pathogenic *V. anguillarum* strains isolated from aquaculture environments (Busschaert et al., 2015; Hansen et al., 2020). Accessory genome-based phylogenetic analysis further placed strain VK1 closer to the USA isolate rather than the geographically closer European strains, suggesting that accessory genome composition may not strictly correlate with geographical origin. Similar observations have previously been reported, where phylogenetic clustering and virulence-associated characteristics varied independently of strain origin or host source (Castillo et al., 2017).

Taken together, these findings provide genomic insights into the potential adaptive and virulence-associated features of strain VK1. However, several limitations should be considered when interpreting genome-based predictions. Draft genome assemblies may contain fragmented regions associated with genomic islands, repetitive sequences, and mobile genetic elements, potentially limiting the accurate reconstruction of the genomic context of acquired genes (Ricker et al., 2012). In addition, although comparative genomic approaches such as ANI and pan-genome analyses provide valuable information for taxonomic classification and functional prediction, genomic data alone may not fully reflect phenotypic characteristics or pathogenic behavior (Sentausa and Fournier, 2013). Furthermore, predictions based on gene content and organization remain constrained by the limited availability of closely related reference genomes and the dynamic nature of bacterial genome evolution (Wolf et al., 2001). Therefore, further transcriptomic, phenotypic, and functional studies will be necessary to validate the biological significance of the identified genes and clarify their contribution to the ecology and pathogenic potential of *V. anguillarum* VK1.

CONCLUSION

This study provides genomic insights into the adaptive potential of *V. anguillarum* strain VK1 in aquaculture environments. The findings suggest that this strain may possess traits associated with environmental persistence and host interaction. Although further functional validation is required, the genomic data generated in this study may contribute to future surveillance and disease management strategies in aquaculture systems.

SUPPLEMENTARY MATERIALS

Table S1: ANI values of VK1 compared with other *V. anguillarum* strains. Table S2: Stress-related genes in VK1 genome. Table S3: Virulence-related genes in VK1 genome. Table S4: Presence–absence variation (pav) matrix of accessory genes among *V. anguillarum* strains.

ACKNOWLEDGMENTS AND FUNDING

The authors would like to express their gratitude to the staff of Kılıç Seafood Inc. at the Didim Incubation and Adaptation Facility for their support. This work is part of Kadriye Toklu's master's thesis and was substantially supported by the Ege University Scientific Research Projects Coordination Office (Grant No. 23352).

AUTHOR CONTRIBUTIONS

Kadriye Toklu: Methodology, Original draft writing, Formal analysis, Writing – review and editing. İhsan Yaşa: Conceptualization, Data organization and management, Writing – review and editing, Funding, Project management.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ETHICAL APPROVAL

This study does not involve human participants or animal experiments.

DECLARATION OF AI USE

English language editing and translation benefited in part from the use of artificial intelligence tools (ChatGPT, OpenAI, 2026).

DATA AVAILABILITY

All datasets generated in this study can be obtained from the corresponding author upon reasonable request.

REFERENCES

AlJindan, R., AlEraky, D.M., Farhat, M., Almandil, N.B., AbdulAzeez, S., & Borgio, J.F. (2023). Genomic insights into virulence factors and multi-drug resistance in *Clostridium perfringens* IRMC2505A. *Toxins*, *15*(6), 359. https://doi.org/10.3390/toxins15060359

Bergman, A.M. (1909). Die rote beulenkrankheit des aals. *Bericht aus der Koniglichen Bayerischen Versuchsstation*, 2, 10-54

Bondad-Reantaso, M.G., MacKinnon, B., Karunasagar, I., Fridman, S., Alday-Sanz, V., Brun, E., Le Groumellec, M., Li, A., Surachetpong, W.,

Karunasagar, I., Hao, B., Dall'Occo, A., Urbani, R., & Caputo, A. (2023). Review of alternatives to antibiotic use in aquaculture. *Reviews in Aquaculture*, *15*(4), 1421-1451. https://doi.org/10.1111/raq.12786

Busschaert, P., Frans, I., Crauwels, S., Zhu, B., Willems, K., Bossier, P., Michiels, C., Verstrepen, K., Lievens, B., & Rediers, H. (2015). Comparative genome sequencing to assess the genetic diversity and virulence attributes of 15 *Vibrio anguillarum* isolates. *Journal of Fish Diseases*, *38*(9), 795-807. https://doi.org/10.1111/jfd.12290

Castillo, D., Alvise, P.D., Xu, R., Zhang, F., Middelboe, M., & Gram, L. (2017). Comparative genome analyses of *Vibrio anguillarum* strains reveal a link with pathogenicity traits. *Msystems*, *2*(1), 10-1128. https://doi.org/10.1128/msystems.00174-16

Crisafi, F., Denaro, R., Genovese, M., Yakimov, M., & Genovese, L. (2014). Application of relative real-time PCR to detect differential expression of virulence genes in *Vibrio anguillarum* under standard and stressed growth conditions. *Journal of Fish Diseases*, *37*(7), 629-640. https://doi.org/10.1111/jfd.12158

Eren Eroglu, A.E., Toklu, K., & Yasa, İ. (2025a). Functional genomics of a food-related thermotolerant *Acetobacter oryzifermentans* strain AAB5: genetic determinants of stress response, CAZyme repertoire, and CRISPR-Cas system. *Functional & Integrative Genomics*, *25*(1), 253. https://doi.org/10.1007/s10142-025-01765-0

Eren Eroğlu, A.E., Toklu, K., Sarıkahya, N., & Yaşa, İ. (2025b). In Vitro Ferrophilic Responses of *Photobacterium damselae subsp. piscicida* EKL1 and characterization of the Fe (III)-Piscibactin complex. *Microorganisms*, *13*(4), 858. https://doi.org/10.3390/microorganisms13040858

Frank, J.A., Reich, C.I., Sharma, S., Weisbaum, J.S., Wilson, B.A., & Olsen, G.J. (2008). Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Applied and Environmental Microbiology*, *74*(8), 2461-2470. https://doi.org/10.1128/AEM.02272-07

Frans, I., Michiels, C.W., Bossier, P., Willems, K.A., Lievens, B., & Rediers, H. (2011). *Vibrio anguillarum* as a fish pathogen: virulence factors, diagnosis and prevention. *Journal of Fish Diseases*, *34*(9), 643-661. https://doi.org/10.1111/j.1365-2761.2011.01279

Gao, X., Pi, D., Chen, N., Li, X., Liu, X., Yang, H., Wei, W., Zhang, X. (2018). Survival, virulent characteristics, and transcriptomic analyses of the pathogenic *Vibrio anguillarum* under starvation stress. *Frontiers in Cellular and Infection Microbiology*, *8*, 389. https://doi.org/10.3389/fcimb.2018.00389

Guanhua, Y., Wang, C., Wang, X., Ma, R., Zheng, H., Liu, Q., Zhang, Y., Ma, Y., & Wang, Q. (2018). Complete genome sequence of the marine fish pathogen *Vibrio anguillarum* and genome-wide transposon mutagenesis analysis of genes essential for in vivo infection. *Microbiological Research*, *216*, 97-107. https://doi.org/10.1016/j.micres.2018.08.011

Guillemette, R., Ushijima, B., Jalan, M., Häse, C.C., & Azam, F. (2020). Insight into the resilience and susceptibility of marine bacteria to T6SS attack by *Vibrio cholerae* and *Vibrio coralliilyticus*. *PloS One*, *15*(1), e0227864. https://doi.org/10.1371/journal.pone.0227864

Hansen, M. J., Kudirkiene, E., & Dalsgaard, I. (2020). Analysis of 44 *Vibrio anguillarum* genomes reveals high genetic diversity. PeerJ, 8, e10451. https://doi.org/10.7717/peerj.10451

Hu, J. Y., Zhang, X. K., Xin, C. Q., Zhang, L., Kang, J., Gong, P. (2023). Severe infection by *Vibrio anguillarum* following a bite by a marine fish: a case report. *Emerging Microbes & Infections*, *12*(1), 2204145. https://doi.org/10.1080/22221751.2023.2204145

Lages, M.A., Balado, M., & Lemos, M.L. (2019). The expression of virulence factors in *Vibrio anguillarum* is dually regulated by iron levels and temperature. *Frontiers in Microbiology*, *10*, 2335. https://doi.org/10.3389/fmicb.2019.02335

Liu, B., Gao, Q., Zhang, X., Chen, H., Zhang, Y., Sun, Y., Yang, S., & Chen, C. (2021). CsrA regulates swarming motility and carbohydrate and amino acid metabolism in *Vibrio alginolyticus*. *Microorganisms*, *9*(11), 2383. https://doi.org/10.3390/microorganisms9112383

Lotz, M.J., Tamplin, M.L., & Rodrick, G.E. (1983). Thiosulfate-citrate-bile salts-sucrose agar and its selectivity for clinical and marine vibrio organisms. *Annals of Clinical & Laboratory Science*, *13*(1), 45-48.

Lovell, C.R. (2017). Ecological fitness and virulence features of *Vibrio parahaemolyticus* in estuarine environments. *Applied microbiology and biotechnology*, *101*(5), 1781-1794. https://doi.org/10.1007/s00253-017-8096-9

Machimbirike, V.I., Vasquez, I., Cao, T., Chukwu-Osazuwa, J., Onireti, O., Segovia, C., Khunrae, P., Rattanarojpong, T., Booman, M., Jones, S., Soto-Davila, M., Dixon, B., & Santander, J. (2023). Comparative genomic analysis of virulent *Vibrio* (*Listonella*) *anguillarum* serotypes revealed genetic diversity and genomic signatures in the O-antigen biosynthesis gene cluster. *Microorganisms*, *11*(3), 792. https://doi.org/10.3390/microorganisms11030792

Meier-Kolthoff, J. P., & Göker, M. (2019). TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nature Communications*, *10*(1), 2182.

Mohd Yazid, S.H., Mohd Daud, H., Azmai, M.N.A., Mohamad, N., & Mohd Nor, N. (2021). Estimating the economic loss due to vibriosis in net-cage cultured Asian seabass (*Lates calcarifer*): Evidence from the East coast of Peninsular Malaysia. *Frontiers in Veterinary Science*, *8*, 644009. https://doi.org/10.3389/fvets.2021.644009

Murray, C.S., Gao, Y., & Wu, M. (2021). Re-evaluating the evidence for a universal genetic boundary among microbial species. *Nature Communications*, *12*(1), 4059. https://doi.org/10.1038/s41467-021-24128-2

Naka, H., Dias, G.M., Thompson, C.C., Dubay, C., Thompson, F.L., & Crosa, J.H. (2011). Complete genome sequence of the marine fish pathogen *Vibrio anguillarum* harboring the pJM1 virulence plasmid and genomic comparison with other virulent strains of *V. anguillarum* and *V. ordalii*. *Infection and Immunity*, *79*(7), 2889-2900. https://doi.org/10.1128/iai.05138-11

Park, K.S., Ono, T., Rokuda, M., Jang, M.H., Okada, K., Iida, T., & Honda, T. (2004). Functional characterization of two type III secretion systems of *Vibrio parahaemolyticus*. *Infection and Immunity*, *72*(11), 6659-6665. https://doi.org/10.1128/iai.72.11.6659-6665.2004

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., & Tyson, G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043-1055. http://www.genome.org/cgi/doi/10.1101/gr.186072.114

Phurahong, T., Soonson, P., Thawonsuwan, J., Tanasomwang, V., Areechon, N., E-kobon, T., & Unajak, S. (2024). Comparative Genome Analysis of Piscine *Vibrio vulnificus*: Virulence-Associated Metabolic Pathways. *Microorganisms*, *12*(12), 2518. https://doi.org/10.3390/microorganisms12122518

Prasad, A., Sundararajan, V.S., JayaramanValadi, Nigam, V.K., & Suravajhala, P. (2025). Prediction of virulence factors in bacterial hypothetical proteins. *Archives of Microbiology*, *207*(10), 244. https://doi.org/10.1007/s00203-025-04447-4

Quainoo, S., Coolen, J.P., van Hijum, S.A., Huynen, M.A., Melchers, W.J., van Schaik, W., & Wertheim, H.F. (2017). Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis. *Clinical Microbiology Reviews*, *30*(4), 1015-1063. https://doi.org/10.1128/cmr.00016-17

Reverter, M., Vega-Heredia, S., & Warburton, P.J. (2025). Historical perspective and contribution of aquaculture to the AMR global pandemic. *In Antimicrobial resistance in aquaculture and aquatic environments* (pp. 17-38). Singapore: Springer Nature, Singapore. https://doi.org/10.1007/978-981-97-7320-6_2

Ricker, N., Qian, H., & Fulthorpe, R.R. (2012). The limitations of draft assemblies for understanding prokaryotic adaptation and evolution. Genomics, *100*(3), 167-175. https://doi.org/10.1016/j.ygeno.2012.06.009

Rodkhum, C., Maki, T., Hirono, I., & Aoki, T. (2008). gyrA and parC associated with quinolone resistance in *Vibrio anguillarum*. *Journal of Fish Diseases*, *31*(5). https://doi.org/10.1111/j.1365-2761.2007.00843.x

Sampantamit, T., Ho, L., Lachat, C., Sutummawong, N., Sorgeloos, P., & Goethals, P. (2020). Aquaculture production and its environmental sustainability in Thailand: Challenges and potential solutions. *Sustainability*, *12*(5), 2010. https://doi.org/10.3390/su12052010

Sanches-Fernandes, G.M., Sá-Correia, I., & Costa, R. (2022). Vibriosis outbreaks in aquaculture: addressing environmental and public health concerns and preventive therapies using gilthead seabream farming as a model system. *Frontiers in Microbiology*, *13*, 904815. https://doi.org/10.3389/fmicb.2022.904815

Schroeder, M., Brooks, B.D., & Brooks, A.E. (2017). The complex relationship between virulence and antibiotic resistance. *Genes*, *8*(1), 39. https://doi.org/10.3390/genes8010039

Sentausa, E., & Fournier, P.E. (2013). Advantages and limitations of genomics in prokaryotic taxonomy. *Clinical Microbiology and Infection*, *19*(9), 790-795. https://doi.org/10.1111/1469-0691.12181

Sha, J., Rosenzweig, J.A., Kozlova, E.V., Wang, S., Erova, T.E., Kirtley, M.L., van Lier, C.J., Chopra, A.K. (2013). Evaluation of the roles played by Hcp and VgrG type 6 secretion system effectors in *Aeromonas hydrophila* SSU pathogenesis. *Microbiology*, *159*(Pt_6), 1120-1135. https://doi.org/10.1099/mic.0.063495-0

Soto, W., & Nishiguchi, M.K. (2021). Environmental stress selects for innovations that drive *Vibrio* symbiont diversity. *Frontiers in Ecology and Evolution*, *9*, 616973. https://doi.org/10.3389/fevo.2021.616973

Stothard, P., Grant, J.R., & Van Domselaar, G. (2019). Visualizing and comparing circular genomes using the CGView family of tools. *Briefings in Bioinformatics*, *20*(4), 1576-1582. https://doi.org/10.1093/bib/bbx081

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution*, *38*(7), 3022-3027. https://doi.org/10.1093/molbev/msab120

Tang, H., Liu, Z., Hu, B., & Zhu, L. (2025). Hierarchical activation of resistance genes under tetracyclines selective pressure in complex microbial community. *Journal of Hazardous Materials*, *493*, 138399. https://doi.org/10.1016/j.jhazmat.2025.138399

Toklu, K., Yaşa, İ., & Eroğlu, A.E.E. (2022, September 9–11). Temperature Dependent Siderophore and Hemolysin Production in Vibrio anguillarum. In Proceedings of the III. International Enzyme and Bioprocess Days (EBDays 2022) (p. 46). Tokat, Türkiye. Oral presentation https://drive.google.com/file/d/1sewdj41Mr8o83DX7ftPp0KtXJpNHbGe/view

Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J. A., Gladstone, R.A., Lo, S., Beaudoin, C., Floto, R. A., Frost, S. D.W., Corander, J., Bentley, S.D., & Parkhill, J. (2020). Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biology*, *21*(1). https://doi.org/10.1186/s13059-020-02090-4

Vasquez, I., Cao, T., Chakraborty, S., Gnanagobal, H., O'Brien, N., Monk, J., Boyce, D., Westcott, J.D., Santander, J. (2020). Comparative genomics analysis of *Vibrio anguillarum* isolated from lumpfish (*Cyclopterus lumpus*) in Newfoundland reveals novel chromosomal organizations. *Microorganisms*, *8*(11), 1666. https://doi.org/10.3390/microorganisms8111666

Wolf, Y.I., Rogozin, I.B., Kondrashov, A.S., & Koonin, E.V. (2001). Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. *Genome Research*, *11*(3), 356-372.

Yang, Q., Yang, X.D., Liu, M.Q., Zeng, C., Zhao, H.K., Xiang, K.W., Hou, Z.S., Wen, H.S., Li, J.F. (2023). Transcriptome analysis of liver, gill and intestine in rainbow trout (*Oncorhynchus mykiss*) symptomatically or asymptomatically infected with *Vibrio anguillarum*. *Fish & Shellfish Immunology*, *135*, 108643. https://doi.org/10.1016/j.fsi.2023.108643

Zautner, A. E., Groß, U., Emele, M.F., Hagen, R.M., & Frickmann, H. (2017). More pathogenicity or just more pathogens?—on the interpretation problem of multiple pathogen detections with diagnostic multiplex assays. *Frontiers in Microbiology*, *8*, 1210. https://doi.org/10.3389/fmicb.2017.01210

Table S1. ANI values of VK1 compared with other *V. anguillarum* strains

	Reference Genomes	**ANI value**
V. anguillarum VK1	*V. anguillarum* strain V2022-124 (GCA_049943515_1)	99.96
V. anguillarum VK1	*V. anguillarum* strain 040915-1/1B (GCA_049943615_1)	99.96
V. anguillarum VK1	*V. anguillarum* strain V76 (GCA_049943765_1)	99.96
V. anguillarum VK1	*V. anguillarum* strain V82 (GCA_049943735_1)	99.96
V. anguillarum VK1	*V. anguillarum* strain LMG 32357 (GCA_977925695_1)	99.90
V. anguillarum VK1	*V. anguillarum* strain V100 (GCA_049943555_1)	99.90
V. anguillarum VK1	*V. anguillarum* strain Ba35-E2-R4 (GCA_003391215_1)	99.87
V. anguillarum VK1	*V. anguillarum* strain PF430-3 (GCA_001989695_1)	98.70
V. anguillarum VK1	*V. anguillarum* strain PF4 substr. R6 (GCA_002813855_1)	98.66
V. anguillarum VK1	*V. anguillarum* strain GA (GCA_037969705_1)	98.50
V. anguillarum VK1	*V. anguillarum* strain DSM 21597 (GCA_001989995_2)	98.32
V. anguillarum VK1	*V. anguillarum* strain LMG 4437 (GCA_977925735_1)	98.23
V. anguillarum VK1	*V. anguillarum* strain 040915-1/1B (GCA_015343015_1)	98.14
V. anguillarum VK1	*V. anguillarum* strain LMG 4411 (GCA_977925715_1)	98.03
V. anguillarum VK1	*V. anguillarum* strain PF4-E1-1 GCA_003390515.1	98.58

Table S2. Stress-related genes in VK1 genome

	Subcategory	Subsystem	Role
1	Periplasmic Stress	Periplasmic Stress Response	Sigma factor RpoE negative regulatory protein RseB precursor
2	Periplasmic Stress	Periplasmic Stress Response	Outer membrane protein H precursor
3	Periplasmic Stress	Periplasmic Stress Response	Outer membrane stress sensor protease DegQ, serine protease
4	Periplasmic Stress	Periplasmic Stress Response	Outer membrane stress sensor protease DegS
5	Periplasmic Stress	Periplasmic Stress Response	Survival protein SurA precursor (Peptidyl-prolyl cis-trans isomerase SurA) (EC 5.2.1.8)
6	Periplasmic Stress	Periplasmic Stress Response	Sigma factor RpoE negative regulatory protein RseA
7	Stress Response - no subcategory	Flavohaemoglobin	Flavohemoprotein (Hemoglobin-like protein) (Flavohemoglobin) (Nitric oxide dioxygenase) (EC 1.14.12.17)
8	Stress Response - no subcategory	Phage shock protein (psp) operon	Phage shock protein E
9	Stress Response - no subcategory	Phage shock protein (psp) operon	Phage shock protein A
10	Stress Response - no subcategory	Phage shock protein (psp) operon	Phage shock protein B
11	Stress Response - no subcategory	Phage shock protein (psp) operon	Phage shock protein C
12	Stress Response - no subcategory	Phage shock protein (psp) operon	Psp operon transcriptional activator
13	Stress Response - no subcategory	Bacterial hemoglobins	Flavohemoprotein (Hemoglobin-like protein) (Flavohemoglobin) (Nitric oxide dioxygenase) (EC 1.14.12.17)
14	Stress Response - no subcategory	Bacterial hemoglobins	diguanylate cyclase/phosphodiesterase (GGDEF & EAL domains) with PAS/PAC sensor(s)
15	Stress Response - no subcategory	Hfl operon	HflC protein
16	Stress Response - no subcategory	Hfl operon	Putative inner membrane protein YjeT (clustered with HflC)
17	Stress Response - no subcategory	Hfl operon	RNA-binding protein Hfq
18	Stress Response - no subcategory	Hfl operon	HflK protein
19	Stress Response - no subcategory	Hfl operon	GTP-binding protein HflX
20	Stress Response - no subcategory	SigmaB stress responce regulation	Serine phosphatase RsbU, regulator of sigma subunit
21	Stress Response - no subcategory	SigmaB stress responce regulation	Serine-protein kinase RsbW (EC 2.7.11.1)
22	Stress Response - no subcategory	SigmaB stress responce regulation	Anti-sigma B factor antagonist RsbV
23	Stress Response - no subcategory	Universal stress protein family	Universal stress protein E
24	Stress Response - no subcategory	Universal stress protein family	Universal stress protein family 3
25	Stress Response - no subcategory	Universal stress protein family	Universal stress protein A
26	Stress Response - no subcategory	Universal stress protein family	Universal stress protein B
27	Stress Response - no subcategory	Carbon Starvation	Starvation lipoprotein Slp paralog
28	Stress Response - no subcategory	Carbon Starvation	Carbon storage regulator
29	Stress Response - no subcategory	Carbon Starvation	Carbon starvation protein A
30	Stress Response - no subcategory	Carbon Starvation	Carbon starvation protein A
31	Stress Response - no subcategory	Carbon Starvation	Carbon starvation protein B
32	Stress Response - no subcategory	Sugar-phosphate stress regulation	SgrR, sugar-phosphate stress, transcriptional activator of SgrS small RNA

Table S2. (Continued)

	Subcategory	Subsystem	Role
33	Detoxification	Glutathione-dependent pathway of formaldehyde detoxification	S-(hydroxymethyl)glutathione dehydrogenase (EC 1.1.1.284)
34	Detoxification	Glutathione-dependent pathway of formaldehyde detoxification	S-formylglutathione hydrolase (EC 3.1.2.12)
35	Detoxification	Glutathione-dependent pathway of formaldehyde detoxification	Transcriptional regulator, LysR family, in formaldehyde detoxification operon
36	Detoxification	Uptake of selenate and selenite	Sulfate and thiosulfate import ATP-binding protein CysA (EC 3.6.3.25)
37	Detoxification	Uptake of selenate and selenite	DedA protein
38	Oxidative stress	CoA disulfide thiol-disulfide redox system	CoA-disulfide reductase (EC 1.8.1.14)
39	Oxidative stress	Glutathione: Redox cycle	Glutaredoxin 1
40	Oxidative stress	Glutathione: Redox cycle	Glutaredoxin
41	Oxidative stress	Glutathione: Redox cycle	Glutathione reductase (EC 1.8.1.7)
42	Oxidative stress	Glutathione: Biosynthesis and gamma-glutamyl cycle	Glutamate--cysteine ligase (EC 6.3.2.2)
43	Oxidative stress	Glutathione: Biosynthesis and gamma-glutamyl cycle	Glutathione synthetase (EC 6.3.2.3)
44	Oxidative stress	Oxidative stress	Ferric uptake regulation protein FUR
45	Oxidative stress	Oxidative stress	Iron-binding ferritin-like antioxidant protein
46	Oxidative stress	Oxidative stress	NnrS protein involved in response to NO
47	Oxidative stress	Oxidative stress	Hydrogen peroxide-inducible genes activator
48	Oxidative stress	Oxidative stress	Fumarate and nitrate reduction regulatory protein
49	Oxidative stress	Oxidative stress	Peroxidase (EC 1.11.1.7)
50	Oxidative stress	Oxidative stress	Alkyl hydroperoxide reductase subunit C-like protein
51	Oxidative stress	Oxidative stress	Paraquat-inducible protein A
52	Oxidative stress	Oxidative stress	Paraquat-inducible protein B
53	Oxidative stress	Oxidative stress	Redox-sensitive transcriptional activator SoxR
54	Oxidative stress	Oxidative stress	Non-specific DNA-binding protein Dps
55	Oxidative stress	Oxidative stress	Zinc uptake regulation protein ZUR
56	Oxidative stress	Oxidative stress	Catalase (EC 1.11.1.6)
57	Oxidative stress	Oxidative stress	Superoxide dismutase [Fe] (EC 1.15.1.1)
58	Oxidative stress	Oxidative stress	Ferroxidase (EC 1.16.3.1)
59	Oxidative stress	Glutathione: Non-redox reactions	Glutathione S-transferase, zeta (EC 2.5.1.18)
60	Oxidative stress	Glutathione: Non-redox reactions	FIG005121: SAM-dependent methyltransferase (EC 2.1.1.-)
61	Oxidative stress	Glutathione: Non-redox reactions	Glutathione S-transferase (EC 2.5.1.18)
62	Oxidative stress	Glutathione: Non-redox reactions	Lactoylglutathione lyase (EC 4.4.1.5)
63	Oxidative stress	Glutathione: Non-redox reactions	Glutathione S-transferase, omega (EC 2.5.1.18)
64	Oxidative stress	Glutathione: Non-redox reactions	Hydroxyacylglutathione hydrolase (EC 3.1.2.6)
65	Oxidative stress	Glutaredoxins	Glutaredoxin 1
66	Oxidative stress	Glutaredoxins	Cell wall endopeptidase, family M23/M37
67	Oxidative stress	Glutaredoxins	Glutaredoxin-related protein
68	Oxidative stress	Glutaredoxins	Glutaredoxin
69	Oxidative stress	Glutaredoxins	Flavohemoprotein (Hemoglobin-like protein) (Flavohemoglobin) (Nitric oxide dioxygenase) (EC 1.14.12.17)
70	Oxidative stress	Protection from Reactive Oxygen Species	Catalase (EC 1.11.1.6)
71	Oxidative stress	Protection from Reactive Oxygen Species	Peroxidase (EC 1.11.1.7)
72	Oxidative stress	Protection from Reactive Oxygen Species	Superoxide dismutase [Fe] (EC 1.15.1.1)
73	Heat shock	Heat shock dnaK gene cluster extended	rRNA small subunit methyltransferase I
74	Heat shock	Heat shock dnaK gene cluster extended	Glutathione synthetase (EC 6.3.2.3)
75	Heat shock	Heat shock dnaK gene cluster extended	Ribosomal protein L11 methyltransferase (EC 2.1.1.-)
76	Heat shock	Heat shock dnaK gene cluster extended	Ribonuclease PH (EC 2.7.7.56)
77	Heat shock	Heat shock dnaK gene cluster extended	Ribosomal RNA small subunit methyltransferase E (EC 2.1.1.-)
78	Heat shock	Heat shock dnaK gene cluster extended	Ribosome-associated heat shock protein implicated in the recycling of the 50S subunit (S4 paralog)
79	Heat shock	Heat shock dnaK gene cluster extended	RNA polymerase sigma factor RpoH
80	Heat shock	Heat shock dnaK gene cluster extended	Heat shock protein GrpE
81	Heat shock	Heat shock dnaK gene cluster extended	Chaperone protein DnaJ
82	Heat shock	Heat shock dnaK gene cluster extended	FIG001341: Probable Fe(2+)-trafficking protein YggX
83	Heat shock	Heat shock dnaK gene cluster extended	Translation elongation factor LepA

Table S2. (Continued)

	Subcategory	Subsystem	Role
84	Heat shock	Heat shock dnaK gene cluster extended	Chaperone protein DnaK
85	Heat shock	Heat shock dnaK gene cluster extended	Radical SAM family enzyme, similar to coproporphyrinogen III oxidase, oxygen-independent, clustered with nucleoside-triphosphatase RdgB
86	Heat shock	Heat shock dnaK gene cluster extended	tmRNA-binding protein SmpB
87	Heat shock	Heat shock dnaK gene cluster extended	Nucleoside 5-triphosphatase RdgB (dHAPTP, dITP, XTP-specific) (EC 3.6.1.15)
88	Osmotic stress	Osmotic stress cluster	Heat shock (predicted periplasmic) protein YciM, precursor
89	Osmotic stress	Osmotic stress cluster	Phosphatidylglycerophosphatase B (EC 3.1.3.27)
90	Osmotic stress	Osmoregulation	Glycerol uptake facilitator protein
91	Osmotic stress	Osmoregulation	Outer membrane protein A precursor
92	Osmotic stress	Ectoine biosynthesis and regulation	L-2,4-diaminobutyric acid acetyltransferase (EC 2.3.1.-)
93	Osmotic stress	Ectoine biosynthesis and regulation	Diaminobutyrate-pyruvate aminotransferase (EC 2.6.1.46)
94	Osmotic stress	Ectoine biosynthesis and regulation	L-ectoine synthase (EC 4.2.1.-)
95	Osmotic stress	Ectoine biosynthesis and regulation	Aspartokinase (EC 2.7.2.4) associated with ectoine biosynthesis
96	Osmotic stress	Synthesis of osmoregulated periplasmic glucans	Phosphoglycerol transferase I (EC 2.7.8.20)
97	Osmotic stress	Choline and Betaine Uptake and Betaine Biosynthesis	HTH-type transcriptional regulator BetI
98	Osmotic stress	Choline and Betaine Uptake and Betaine Biosynthesis	Betaine aldehyde dehydrogenase (EC 1.2.1.8)
99	Osmotic stress	Choline and Betaine Uptake and Betaine Biosynthesis	Choline dehydrogenase (EC 1.1.99.1)
100	Osmotic stress	Choline and Betaine Uptake and Betaine Biosynthesis	L-proline glycine betaine binding ABC transporter protein ProX (TC 3.A.1.12.1)
101	Osmotic stress	Choline and Betaine Uptake and Betaine Biosynthesis	L-proline glycine betaine ABC transport system permease protein ProV (TC 3.A.1.12.1)
102	Osmotic stress	Choline and Betaine Uptake and Betaine Biosynthesis	L-proline glycine betaine ABC transport system permease protein ProW (TC 3.A.1.12.1)
103	Osmotic stress	Choline and Betaine Uptake and Betaine Biosynthesis	High-affinity choline uptake protein BetT
104	Cold shock	Cold shock, CspA family of proteins	Cold shock protein CspE
105	Cold shock	Cold shock, CspA family of proteins	Cold shock protein CspD
106	Cold shock	Cold shock, CspA family of proteins	Cold shock protein CspA

Table S3. Virulence-related genes in VK1 genome

	Gene	Product	Classification
1	pilC	Type IV fimbrial assembly protein PilC	Adherence
2	mshE	MSHA biogenesis protein MshE	Adherence
3	tufA	Translation elongation factor Tu	Adherence
4	tufA	Translation elongation factor Tu	Adherence
5	tufA	Translation elongation factor Tu	Adherence
6	llpA	Methionine ABC transporter substrate-binding protein	Adherence
7	pilB	Type IV fimbrial assembly, ATPase PilB	Adherence
8	tufA	Translation elongation factor Tu	Adherence
9	luxS	S-ribosylhomocysteine lyase (EC 4.4.1.21) @ Autoinducer-2 production protein LuxS	Biofilm
10	hcp-1	T6SS component Hcp	Effector delivery system
11	epsD	General secretion pathway protein D	Effector delivery system
12	vipB/mglB	T6SS component TssC (ImpC/VipB)	Effector delivery system
13	vasE	T6SS component TssK (ImpJ/VasE)	Effector delivery system
14	vasF	T6SS outer membrane component TssL (ImpK/VasF)	Effector delivery system
15	vasD	T6SS secretion lipoprotein TssJ (VasD)	Effector delivery system
16	hcp-1	T6SS component Hcp	Effector delivery system
17	vscN2	Type III secretion cytoplasmic ATP synthase (EC 3.6.3.14, YscN,SpaL,MxiB,HrcN,EscN)	Effector delivery system
18	epsG	General secretion pathway protein G	Effector delivery system
19	VCA0109	T6SS lysozyme-like component TssE	Effector delivery system
20	epsE	General secretion pathway protein E	Effector delivery system
21	epsF	General secretion pathway protein F	Effector delivery system
22	vipA/mglA	T6SS component TssB (ImpB/VipA)	Effector delivery system
23	clpB/vasG	T6SS AAA+ chaperone ClpV (TssH)	Effector delivery system
24	icmF/vasK	T6SS component TssM (IcmF/VasK)	Effector delivery system
25	vasA	T6SS component TssF (ImpG/VasA)	Effector delivery system
26	gmhA/lpcA	D-sedoheptulose 7-phosphate isomerase (EC 5.3.1.28)	Immune modulation
27	cheR	Chemotaxis protein methyltransferase CheR (EC 2.1.1.80)	Motility
28	fliR	Flagellar biosynthesis protein FliR	Motility
29	fliN	Flagellar motor switch protein FliN	Motility
30	fliQ	Flagellar biosynthesis protein FliQ	Motility
31	fliA	RNA polymerase sigma factor for flagellar operon	Motility
32	flgD	Flagellar basal-body rod modification protein FlgD	Motility
33	fliM	Flagellar motor switch protein FliM	Motility
34	motB	Flagellar motor rotation protein MotB	Motility
35	flgO	Flagellar protein FlgO	Motility
36	flgI	Flagellar P-ring protein FlgI	Motility
37	fliE	Flagellar hook-basal body complex protein FliE	Motility
38	flhF	Flagellar biosynthesis protein FlhF	Motility
39	flaA	Flagellin protein FlaC	Motility
40	flgB	Flagellar basal-body rod protein FlgB	Motility
41	flgL	Flagellar hook-associated protein FlgL	Motility
42	flgP	Flagellar protein FlgP	Motility
43	flrA	Flagellar regulatory protein FleQ	Motility
44	flaE	Flagellin protein FlaF	Motility
45	cheB	Chemotaxis response regulator protein-glutamate methylesterase CheB (EC 3.1.1.61)	Motility
46	cheW	Positive regulator of CheA protein activity (CheW)	Motility
47	flgC	Flagellar basal-body rod protein FlgC	Motility
48	flgE	Flagellar hook protein FlgE	Motility
49	flhB	Flagellar biosynthesis protein FlhB	Motility
50	flgF	Flagellar basal-body rod protein FlgF	Motility
51	flgG	Flagellar basal-body rod protein FlgG	Motility
52	fleN/flhG	Flagellar synthesis regulator FleN	Motility
53	cheV	Chemotaxis protein CheV (EC 2.7.3.-)	Motility
54	flgH	Flagellar L-ring protein FlgH	Motility
55	fliI	Flagellum-specific ATP synthase FliI	Motility
56	fliS	Flagellar biosynthesis protein FliS	Motility
57	fleS/flrB	Flagellar sensor histidine kinase FleS	Motility
58	flaD	Flagellin protein FlaD	Motility
59	flaC	Flagellin protein FlaD	Motility
60	fliG	Flagellar motor switch protein FliG	Motility
61	fliF	Flagellar M-ring protein FliF	Motility
62	flhA	Flagellar biosynthesis protein FlhA	Motility
63	cheY	Chemotaxis regulator - transmits chemoreceptor signals to flagellar motor components CheY	Motility
64	flaB	Flagellin protein FlaA	Motility
65	fleR/flrC	Flagellar two-component response regulator FleR	Motility
66	motA	Flagellar motor rotation protein MotA	Motility
67	motY	Sodium-type flagellar protein MotY	Motility
68	fliL	Flagellar basal body-associated protein FliL	Motility
69	fliP	Flagellar biosynthesis protein FliP	Motility
70	csrA	Carbon storage regulator	Regulation
71	rpoS	RNA polymerase sigma factor RpoS	Regulation

Table S4. Presence–absence variation (pav) matrix of accessory genes among *V. anguillarum* strains

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
1	COG1396	[K] COG1396 Predicted transcriptional regulators	6284	0	0	0	0	1	1	Strain-specific
2			6265	0	0	0	0	1	1	Strain-specific
3	COG1192	[D] COG1192 ATPases involved in chromosome partitioning	6032	0	0	0	0	1	1	Strain-specific
4			6028	0	0	0	0	1	1	Strain-specific
5	COG2207	[K] COG2207 AraC-type DNA-binding domain-containing proteins	6021	0	0	0	0	1	1	Strain-specific
6	COG3436	[L] COG3436 Transposase and inactivated derivatives	6008	0	0	0	0	1	1	Strain-specific
7	COG4974	[L] COG4974 Site-specific recombinase XerD	6006	0	0	0	0	1	1	Strain-specific
8			5743	0	0	0	0	1	1	Strain-specific
9	COG0863	[L] COG0863 DNA modification methylase	5738	0	0	0	0	1	1	Strain-specific
10			5720	0	0	0	0	1	1	Strain-specific
11			5429	0	0	0	0	1	1	Strain-specific
12			5427	0	0	0	0	1	1	Strain-specific
13	COG4974	[L] COG4974 Site-specific recombinase XerD	5411	0	0	0	0	1	1	Strain-specific
14	COG0732	[V] COG0732 Restriction endonuclease S subunits	5405	0	0	0	0	1	1	Strain-specific
15	COG0847	[L] COG0847 DNA polymerase III, epsilon subunit and related 3'-5' exonucleases	5372	0	0	0	0	1	1	Strain-specific
16			5371	0	0	0	0	1	1	Strain-specific
17			5368	0	0	0	0	1	1	Strain-specific
18	COG4938	[S] COG4938 Uncharacterized conserved protein	5367	0	0	0	0	1	1	Strain-specific
19			5245	0	0	0	0	1	1	Strain-specific
20			5239	0	0	0	0	1	1	Strain-specific
21			5238	0	0	0	0	1	1	Strain-specific
22	COG3593	[L] COG3593 Predicted ATP-dependent endonuclease of the OLD family	5234	0	0	0	0	1	1	Strain-specific
23			5106	0	0	0	0	1	1	Strain-specific
24	COG2378	[K] COG2378 Predicted transcriptional regulator	5103	0	0	0	0	1	1	Strain-specific
25			5101	0	0	0	0	1	1	Strain-specific
26			5098	0	0	0	0	1	1	Strain-specific
27			5097	0	0	0	0	1	1	Strain-specific
28			5093	0	0	0	0	1	1	Strain-specific
29			4956	0	0	0	0	1	1	Strain-specific
30			4951	0	0	0	0	1	1	Strain-specific
31	COG5933		4949	0	0	0	0	1	1	Strain-specific
32			4944	0	0	0	0	1	1	Strain-specific
33			4943	0	0	0	0	1	1	Strain-specific
34			4937	0	0	0	0	1	1	Strain-specific
35	COG4823	[V] COG4823 Abortive infection bacteriophage resistance protein	4798	0	0	0	0	1	1	Strain-specific
36	COG4974	[L] COG4974 Site-specific recombinase XerD	4796	0	0	0	0	1	1	Strain-specific
37	COG3436	[L] COG3436 Transposase and inactivated derivatives	4795	0	0	0	0	1	1	Strain-specific
38			4791	0	0	0	0	1	1	Strain-specific
39	COG5936		4789	0	0	0	0	1	1	Strain-specific
40	COG0433	[R] COG0433 Predicted ATPase	4782	0	0	0	0	1	1	Strain-specific
41			4620	0	0	0	0	1	1	Strain-specific
42	COG5935		4618	0	0	0	0	1	1	Strain-specific
43			4465	0	0	0	0	1	1	Strain-specific
44			4464	0	0	0	0	1	1	Strain-specific
45			4461	0	0	0	0	1	1	Strain-specific
46			4454	0	0	0	0	1	1	Strain-specific
47			4452	0	0	0	0	1	1	Strain-specific
48	COG0732	[V] COG0732 Restriction endonuclease S subunits	4440	0	0	0	0	1	1	Strain-specific
49			4308	0	0	0	0	1	1	Strain-specific
50	COG4974	[L] COG4974 Site-specific recombinase XerD	4307	0	0	0	0	1	1	Strain-specific
51			4306	0	0	0	0	1	1	Strain-specific
52			4304	0	0	0	0	1	1	Strain-specific
53	COG3344	[L] COG3344 Retron-type reverse transcriptase	4303	0	0	0	0	1	1	Strain-specific
54			4301	0	0	0	0	1	1	Strain-specific
55	COG1506	[E] COG1506 Dipeptidyl aminopeptidases/acylaminoacyl-peptidases	4207	0	0	0	0	1	1	Strain-specific
56			4205	0	0	0	0	1	1	Strain-specific
57	COG1061	[KL] COG1061 DNA or RNA helicases of superfamily II	4201	0	0	0	0	1	1	Strain-specific
58			4196	0	0	0	0	1	1	Strain-specific
59	COG1247	[M] COG1247 Sortase and related acyltransferases	4097	0	0	0	0	1	1	Strain-specific
60			4096	0	0	0	0	1	1	Strain-specific
61			3873	0	0	0	0	1	1	Strain-specific
62			3872	0	0	0	0	1	1	Strain-specific
63			3870	0	0	0	0	1	1	Strain-specific
64			3751	0	0	0	0	1	1	Strain-specific
65			3750	0	0	0	0	1	1	Strain-specific
66			3748	0	0	0	0	1	1	Strain-specific
67			3746	0	0	0	0	1	1	Strain-specific
68	COG2199	[T] COG2199 FOG: GGDEF domain	3745	0	0	0	0	1	1	Strain-specific
69			3630	0	0	0	0	1	1	Strain-specific
70	COG0582	[L] COG0582 Integrase	3628	0	0	0	0	1	1	Strain-specific
71	COG0286	[V] COG0286 Type I restriction-modification system methyltransferase subunit	3626	0	0	0	0	1	1	Strain-specific
72	COG3328	[L] COG3328 Transposase and inactivated derivatives	3624	0	0	0	0	1	1	Strain-specific
73	COG3914	[O] COG3914 Predicted O-linked N-acetylglucosamine transferase, SPINDLY family	3623	0	0	0	0	1	1	Strain-specific
74			3510	0	0	0	0	1	1	Strain-specific
75			3509	0	0	0	0	1	1	Strain-specific
76			3506	0	0	0	0	1	1	Strain-specific
77	COG0732	[V] COG0732 Restriction endonuclease S subunits	3505	0	0	0	0	1	1	Strain-specific
78	COG1670	[J] COG1670 Acetyltransferases, including N-acetylases of ribosomal proteins	3276	0	0	0	0	1	1	Strain-specific
79	COG4974	[L] COG4974 Site-specific recombinase XerD	3264	0	0	0	0	1	1	Strain-specific
80	COG5640	[O] COG5640 Secreted trypsin-like serine protease	3257	0	0	0	0	1	1	Strain-specific
81	COG5304	[S] COG5304 Uncharacterized protein conserved in bacteria	3094	0	0	0	0	1	1	Strain-specific
82	COG3210	[U] COG3210 Large exoproteins involved in heme utilization or adhesion	3079	0	0	0	0	1	1	Strain-specific
83	COG1106	[R] COG1106 Predicted ATPases	3077	0	0	0	0	1	1	Strain-specific
84			3073	0	0	0	0	1	1	Strain-specific
85	COG2929	[S] COG2929 Uncharacterized protein conserved in bacteria	2905	0	0	0	0	1	1	Strain-specific
86			2899	0	0	0	0	1	1	Strain-specific
87			2891	0	0	0	0	1	1	Strain-specific
88			2723	0	0	0	0	1	1	Strain-specific
89	COG1072	[H] COG1072 Panthothenate kinase	2716	0	0	0	0	1	1	Strain-specific
90	COG3547	[L] COG3547 Transposase and inactivated derivatives	2708	0	0	0	0	1	1	Strain-specific
91	COG2842	[R] COG2842 Uncharacterized ATPase, putative transposase	2702	0	0	0	0	1	1	Strain-specific
92			2518	0	0	0	0	1	1	Strain-specific
93			2517	0	0	0	0	1	1	Strain-specific
94	COG3341	[R] COG3341 Predicted double-stranded RNA/RNA-DNA hybrid binding protein	2510	0	0	0	0	1	1	Strain-specific
95	COG1672	[R] COG1672 Predicted ATPase (AAA+ superfamily)	2503	0	0	0	0	1	1	Strain-specific
96			2496	0	0	0	0	1	1	Strain-specific
97			2074	0	0	0	0	1	1	Strain-specific
98	COG0834	[ET] COG0834 ABC-type amino acid transport/signal transduction systems, periplas	2072	0	0	0	0	1	1	Strain-specific
99			2065	0	0	0	0	1	1	Strain-specific
100	COG4942	[D] COG4942 Membrane-bound metallopeptidase	2063	0	0	0	0	1	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
101	COG4974	[L] COG4974 Site-specific recombinase XerD	2037	0	0	0	0	1	1	Strain-specific
102			1692	0	0	0	0	1	1	Strain-specific
103	COG1670	[J] COG1670 Acetyltransferases, including N-acetylases of ribosomal proteins	1681	0	0	0	0	1	1	Strain-specific
104			1674	0	0	0	0	1	1	Strain-specific
105	COG1051	[F] COG1051 ADP-ribose pyrophosphatase	1374	0	0	0	0	1	1	Strain-specific
106	COG0433	[R] COG0433 Predicted ATPase	1371	0	0	0	0	1	1	Strain-specific
107	COG0642	[T] COG0642 Signal transduction histidine kinase	1365	0	0	0	0	1	1	Strain-specific
108			1264	0	0	0	0	1	1	Strain-specific
109	COG3437	[KT] COG3437 Response regulator containing a CheY-like receiver domain and an H	1257	0	0	0	0	1	1	Strain-specific
110			1148	0	0	0	0	1	1	Strain-specific
111			1144	0	0	0	0	1	1	Strain-specific
112			1142	0	0	0	0	1	1	Strain-specific
113	COG5346	[S] COG5346 Predicted membrane protein	1028	0	0	0	0	1	1	Strain-specific
114			1023	0	0	0	0	1	1	Strain-specific
115	COG5124	[DR] COG5124 Protein predicted to be involved in meiotic recombination	1021	0	0	0	0	1	1	Strain-specific
116			1017	0	0	0	0	1	1	Strain-specific
117	COG0286	[V] COG0286 Type I restriction-modification system methyltransferase subunit	1014	0	0	0	0	1	1	Strain-specific
118	COG4974	[L] COG4974 Site-specific recombinase XerD	868	0	0	0	0	1	1	Strain-specific
119			816	0	0	0	0	1	1	Strain-specific
120	COG1247	[M] COG1247 Sortase and related acyltransferases	406	0	0	0	0	1	1	Strain-specific
121			405	0	0	0	0	1	1	Strain-specific
122			404	0	0	0	0	1	1	Strain-specific
123	COG4974	[L] COG4974 Site-specific recombinase XerD	396	0	0	0	0	1	1	Strain-specific
124			335	0	0	0	0	1	1	Strain-specific
125	COG3505	[U] COG3505 Type IV secretory pathway, VirD4 components	6223	0	0	0	1	0	1	Strain-specific
126	COG1061	[KL] COG1061 DNA or RNA helicases of superfamily II	6222	0	0	0	1	0	1	Strain-specific
127			6220	0	0	0	1	0	1	Strain-specific
128			6218	0	0	0	1	0	1	Strain-specific
129	COG0582	[L] COG0582 Integrase	6217	0	0	0	1	0	1	Strain-specific
130	COG0433	[R] COG0433 Predicted ATPase	6216	0	0	0	1	0	1	Strain-specific
131	COG3064	[M] COG3064 Membrane protein involved in colicin uptake	6215	0	0	0	1	0	1	Strain-specific
132			6213	0	0	0	1	0	1	Strain-specific
133	COG0789	[K] COG0789 Predicted transcriptional regulators	6212	0	0	0	1	0	1	Strain-specific
134			6211	0	0	0	1	0	1	Strain-specific
135	COG1737	[K] COG1737 Transcriptional regulators	6209	0	0	0	1	0	1	Strain-specific
136	COG4974	[L] COG4974 Site-specific recombinase XerD	6206	0	0	0	1	0	1	Strain-specific
137			6203	0	0	0	1	0	1	Strain-specific
138	COG1535	[Q] COG1535 Isochorismate hydrolase	6202	0	0	0	1	0	1	Strain-specific
139			6201	0	0	0	1	0	1	Strain-specific
140	COG2246	[S] COG2246 Predicted membrane protein	6187	0	0	0	1	0	1	Strain-specific
141	COG5640	[O] COG5640 Secreted trypsin-like serine protease	6185	0	0	0	1	0	1	Strain-specific
142	COG0582	[L] COG0582 Integrase	6177	0	0	0	1	0	1	Strain-specific
143	COG3436	[L] COG3436 Transposase and inactivated derivatives	6171	0	0	0	1	0	1	Strain-specific
144	COG1247	[M] COG1247 Sortase and related acyltransferases	6158	0	0	0	1	0	1	Strain-specific
145	COG2842	[R] COG2842 Uncharacterized ATPase, putative transposase	6157	0	0	0	1	0	1	Strain-specific
146	COG0764	[I] COG0764 3-hydroxymyristoyl/3-hydroxydecanoyl-(acyl carrier protein) dehydrat	5962	0	0	0	1	0	1	Strain-specific
147	COG4771	[P] COG4771 Outer membrane receptor for ferrienterochelin and colicins	5961	0	0	0	1	0	1	Strain-specific
148			5960	0	0	0	1	0	1	Strain-specific
149	COG1075	[R] COG1075 Predicted acetyltransferases and hydrolases with the alpha/beta hydr	5958	0	0	0	1	0	1	Strain-specific
150			5955	0	0	0	1	0	1	Strain-specific
151	COG1357	[S] COG1357 Uncharacterized low-complexity proteins	5954	0	0	0	1	0	1	Strain-specific
152			5953	0	0	0	1	0	1	Strain-specific
153	COG0730	[R] COG0730 Predicted permeases	5952	0	0	0	1	0	1	Strain-specific
154	COG3010	[G] COG3010 Putative N-acetylmannosamine-6-phosphate epimerase	5950	0	0	0	1	0	1	Strain-specific
155			5949	0	0	0	1	0	1	Strain-specific
156			5948	0	0	0	1	0	1	Strain-specific
157			5947	0	0	0	1	0	1	Strain-specific
158	COG4278	[S] COG4278 Uncharacterized conserved protein	5943	0	0	0	1	0	1	Strain-specific
159	COG2842	[R] COG2842 Uncharacterized ATPase, putative transposase	5940	0	0	0	1	0	1	Strain-specific
160			5939	0	0	0	1	0	1	Strain-specific
161	COG3293	[L] COG3293 Transposase and inactivated derivatives	5938	0	0	0	1	0	1	Strain-specific
162			5930	0	0	0	1	0	1	Strain-specific
163			5923	0	0	0	1	0	1	Strain-specific
164			5913	0	0	0	1	0	1	Strain-specific
165	COG3436	[L] COG3436 Transposase and inactivated derivatives	5907	0	0	0	1	0	1	Strain-specific
166			5896	0	0	0	1	0	1	Strain-specific
167	COG1475	[K] COG1475 Predicted transcriptional regulators	5890	0	0	0	1	0	1	Strain-specific
168			5889	0	0	0	1	0	1	Strain-specific
169	COG0433	[R] COG0433 Predicted ATPase	5669	0	0	0	1	0	1	Strain-specific
170	COG2801	[L] COG2801 Transposase and inactivated derivatives	5668	0	0	0	1	0	1	Strain-specific
171			5667	0	0	0	1	0	1	Strain-specific
172			5666	0	0	0	1	0	1	Strain-specific
173	COG3183	[V] COG3183 Predicted restriction endonuclease	5665	0	0	0	1	0	1	Strain-specific
174			5664	0	0	0	1	0	1	Strain-specific
175	COG0178	[L] COG0178 Excinuclease ATPase subunit	5663	0	0	0	1	0	1	Strain-specific
176			5659	0	0	0	1	0	1	Strain-specific
177	COG1112	[L] COG1112 Superfamily I DNA and RNA helicases and helicase subunits	5658	0	0	0	1	0	1	Strain-specific
178	COG1192	[D] COG1192 ATPases involved in chromosome partitioning	5655	0	0	0	1	0	1	Strain-specific
179			5650	0	0	0	1	0	1	Strain-specific
180			5649	0	0	0	1	0	1	Strain-specific
181			5644	0	0	0	1	0	1	Strain-specific
182	COG3550	[R] COG3550 Uncharacterized protein related to capsule biosynthesis enzymes	5624	0	0	0	1	0	1	Strain-specific
183	COG0270	[L] COG0270 Site-specific DNA methylase	5620	0	0	0	1	0	1	Strain-specific
184			5613	0	0	0	1	0	1	Strain-specific
185			5607	0	0	0	1	0	1	Strain-specific
186			5595	0	0	0	1	0	1	Strain-specific
187	COG1162	[R] COG1162 Predicted GTPases	5588	0	0	0	1	0	1	Strain-specific
188			5587	0	0	0	1	0	1	Strain-specific
189	COG1764	[O] COG1764 Predicted redox protein, regulator of disulfide bond formation	5345	0	0	0	1	0	1	Strain-specific
190	COG4283	[S] COG4283 Uncharacterized conserved protein	5344	0	0	0	1	0	1	Strain-specific
191			5343	0	0	0	1	0	1	Strain-specific
192	COG2274	[V] COG2274 ABC-type bacteriocin/lantibiotic exporters, contain an N-terminal dou	5342	0	0	0	1	0	1	Strain-specific
193	COG2871	[C] COG2871 Na+-transporting NADH:ubiquinone oxidoreductase, subunit NqrF	5341	0	0	0	1	0	1	Strain-specific
194			5340	0	0	0	1	0	1	Strain-specific
195	COG1196	[D] COG1196 Chromosome segregation ATPases	5339	0	0	0	1	0	1	Strain-specific
196			5338	0	0	0	1	0	1	Strain-specific
197			5337	0	0	0	1	0	1	Strain-specific
198			5336	0	0	0	1	0	1	Strain-specific
199			5335	0	0	0	1	0	1	Strain-specific
200			5333	0	0	0	1	0	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
201			5331	0	0	0	1	0	1	Strain-specific
202			5330	0	0	0	1	0	1	Strain-specific
203	COG1114	[E] COG1114 Branched-chain amino acid permeases	5328	0	0	0	1	0	1	Strain-specific
204	COG1020	[Q] COG1020 Non-ribosomal peptide synthetase modules and related proteins	5323	0	0	0	1	0	1	Strain-specific
205			5322	0	0	0	1	0	1	Strain-specific
206	COG4626	[R] COG4626 Phage terminase-like protein, large subunit	5319	0	0	0	1	0	1	Strain-specific
207	COG3436	[L] COG3436 Transposase and inactivated derivatives	5315	0	0	0	1	0	1	Strain-specific
208	COG1247	[M] COG1247 Sortase and related acyltransferases	5314	0	0	0	1	0	1	Strain-specific
209	COG1674	[D] COG1674 DNA segregation ATPase FtsK/SpoIIIE and related proteins	5312	0	0	0	1	0	1	Strain-specific
210	COG1961	[L] COG1961 Site-specific recombinases, DNA invertase Pin homologs	5216	0	0	0	1	0	1	Strain-specific
211			5214	0	0	0	1	0	1	Strain-specific
212			5213	0	0	0	1	0	1	Strain-specific
213	COG3593	[L] COG3593 Predicted ATP-dependent endonuclease of the OLD family	5212	0	0	0	1	0	1	Strain-specific
214	COG0543	[HC] COG0543 2-polyprenylphenol hydroxylase and related flavodoxin oxidoreduct	5211	0	0	0	1	0	1	Strain-specific
215	COG3081	[R] COG3081 Nucleoid-associated protein	5210	0	0	0	1	0	1	Strain-specific
216	COG0438	[M] COG0438 Glycosyltransferase	5208	0	0	0	1	0	1	Strain-specific
217			5206	0	0	0	1	0	1	Strain-specific
218	COG0021	[G] COG0021 Transketolase	5205	0	0	0	1	0	1	Strain-specific
219			5204	0	0	0	1	0	1	Strain-specific
220	COG3747	[L] COG3747 Phage terminase, small subunit	5200	0	0	0	1	0	1	Strain-specific
221			5199	0	0	0	1	0	1	Strain-specific
222			5197	0	0	0	1	0	1	Strain-specific
223	COG1396	[K] COG1396 Predicted transcriptional regulators	5194	0	0	0	1	0	1	Strain-specific
224	COG3208	[Q] COG3208 Predicted thioesterase involved in non-ribosomal peptide biosynthes	5192	0	0	0	1	0	1	Strain-specific
225			5188	0	0	0	1	0	1	Strain-specific
226	COG0300	[R] COG0300 Short-chain dehydrogenases of various substrate specificities	5187	0	0	0	1	0	1	Strain-specific
227			5183	0	0	0	1	0	1	Strain-specific
228	COG4974	[L] COG4974 Site-specific recombinase XerD	5065	0	0	0	1	0	1	Strain-specific
229			5062	0	0	0	1	0	1	Strain-specific
230	COG1246	[E] COG1246 N-acetylglutamate synthase and related acetyltransferases	5061	0	0	0	1	0	1	Strain-specific
231	COG2373	[R] COG2373 Large extracellular alpha-helical protein	5058	0	0	0	1	0	1	Strain-specific
232	COG4974	[L] COG4974 Site-specific recombinase XerD	5057	0	0	0	1	0	1	Strain-specific
233			5056	0	0	0	1	0	1	Strain-specific
234	COG0451	[MG] COG0451 Nucleoside-diphosphate-sugar epimerases	5053	0	0	0	1	0	1	Strain-specific
235	COG1858	[P] COG1858 Cytochrome c peroxidase	5052	0	0	0	1	0	1	Strain-specific
236	COG4973	[L] COG4973 Site-specific recombinase XerC	5051	0	0	0	1	0	1	Strain-specific
237			5049	0	0	0	1	0	1	Strain-specific
238	COG4626	[R] COG4626 Phage terminase-like protein, large subunit	5047	0	0	0	1	0	1	Strain-specific
239			5046	0	0	0	1	0	1	Strain-specific
240	COG0583	[K] COG0583 Transcriptional regulator	5044	0	0	0	1	0	1	Strain-specific
241	COG2842	[R] COG2842 Uncharacterized ATPase, putative transposase	5043	0	0	0	1	0	1	Strain-specific
242	COG3344	[L] COG3344 Retron-type reverse transcriptase	5042	0	0	0	1	0	1	Strain-specific
243			5040	0	0	0	1	0	1	Strain-specific
244			5035	0	0	0	1	0	1	Strain-specific
245			5034	0	0	0	1	0	1	Strain-specific
246			4912	0	0	0	1	0	1	Strain-specific
247	COG0664	[T] COG0664 cAMP-binding proteins - catabolite gene activator and regulatory subu	4911	0	0	0	1	0	1	Strain-specific
248			4905	0	0	0	1	0	1	Strain-specific
249			4904	0	0	0	1	0	1	Strain-specific
250			4903	0	0	0	1	0	1	Strain-specific
251	COG3550	[R] COG3550 Uncharacterized protein related to capsule biosynthesis enzymes	4902	0	0	0	1	0	1	Strain-specific
252			4901	0	0	0	1	0	1	Strain-specific
253			4897	0	0	0	1	0	1	Strain-specific
254			4896	0	0	0	1	0	1	Strain-specific
255	COG4096	[V] COG4096 Type I site-specific restriction-modification system, R (restriction) subu	4895	0	0	0	1	0	1	Strain-specific
256			4893	0	0	0	1	0	1	Strain-specific
257			4891	0	0	0	1	0	1	Strain-specific
258			4888	0	0	0	1	0	1	Strain-specific
259	COG0656	[R] COG0656 Aldo/keto reductases, related to diketogulonate reductase	4886	0	0	0	1	0	1	Strain-specific
260	COG0859	[M] COG0859 ADP-heptose:LPS heptosyltransferase	4885	0	0	0	1	0	1	Strain-specific
261			4884	0	0	0	1	0	1	Strain-specific
262	COG2003	[L] COG2003 DNA repair proteins	4883	0	0	0	1	0	1	Strain-specific
263	COG5640	[O] COG5640 Secreted trypsin-like serine protease	4882	0	0	0	1	0	1	Strain-specific
264	COG0840	[NT] COG0840 Methyl-accepting chemotaxis protein	4751	0	0	0	1	0	1	Strain-specific
265			4749	0	0	0	1	0	1	Strain-specific
266	COG3077	[L] COG3077 DNA-damage-inducible protein J	4748	0	0	0	1	0	1	Strain-specific
267			4747	0	0	0	1	0	1	Strain-specific
268			4746	0	0	0	1	0	1	Strain-specific
269			4744	0	0	0	1	0	1	Strain-specific
270	COG3409	[M] COG3409 Putative peptidoglycan-binding domain-containing protein	4742	0	0	0	1	0	1	Strain-specific
271	COG0021	[G] COG0021 Transketolase	4741	0	0	0	1	0	1	Strain-specific
272			4738	0	0	0	1	0	1	Strain-specific
273			4736	0	0	0	1	0	1	Strain-specific
274	COG0583	[K] COG0583 Transcriptional regulator	4730	0	0	0	1	0	1	Strain-specific
275	COG2274	[V] COG2274 ABC-type bacteriocin/lantibiotic exporters, contain an N-terminal dou	4729	0	0	0	1	0	1	Strain-specific
276			4726	0	0	0	1	0	1	Strain-specific
277	COG0346	[E] COG0346 Lactoylglutathione lyase and related lyases	4723	0	0	0	1	0	1	Strain-specific
278			4722	0	0	0	1	0	1	Strain-specific
279	COG2003	[L] COG2003 DNA repair proteins	4720	0	0	0	1	0	1	Strain-specific
280	COG0656	[R] COG0656 Aldo/keto reductases, related to diketogulonate reductase	4716	0	0	0	1	0	1	Strain-specific
281	COG2003	[L] COG2003 DNA repair proteins	4715	0	0	0	1	0	1	Strain-specific
282	COG1134	[GM] COG1134 ABC-type polysaccharide/polyol phosphate transport system, ATPas	4585	0	0	0	1	0	1	Strain-specific
283			4584	0	0	0	1	0	1	Strain-specific
284			4582	0	0	0	1	0	1	Strain-specific
285	COG0025	[P] COG0025 NhaP-type Na+/H+ and K+/H+ antiporters	4581	0	0	0	1	0	1	Strain-specific
286	COG3307	[M] COG3307 Lipid A core - O-antigen ligase and related enzymes	4580	0	0	0	1	0	1	Strain-specific
287	COG3500	[R] COG3500 Phage protein D	4578	0	0	0	1	0	1	Strain-specific
288			4577	0	0	0	1	0	1	Strain-specific
289	COG0535	[R] COG0535 Predicted Fe-S oxidoreductases	4573	0	0	0	1	0	1	Strain-specific
290	COG0463	[M] COG0463 Glycosyltransferases involved in cell wall biogenesis	4569	0	0	0	1	0	1	Strain-specific
291	COG3055	[S] COG3055 Uncharacterized protein conserved in bacteria	4566	0	0	0	1	0	1	Strain-specific
292	COG3658	[C] COG3658 Cytochrome b	4562	0	0	0	1	0	1	Strain-specific
293	COG0286	[V] COG0286 Type I restriction-modification system methyltransferase subunit	4561	0	0	0	1	0	1	Strain-specific
294			4560	0	0	0	1	0	1	Strain-specific
295			4557	0	0	0	1	0	1	Strain-specific
296			4550	0	0	0	1	0	1	Strain-specific
297			4402	0	0	0	1	0	1	Strain-specific
298	COG4823	[V] COG4823 Abortive infection bacteriophage resistance protein	4401	0	0	0	1	0	1	Strain-specific
299			4398	0	0	0	1	0	1	Strain-specific
300			4397	0	0	0	1	0	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
301	COG4974	[L] COG4974 Site-specific recombinase XerD	4395	0	0	0	1	0	1	Strain-specific
302	COG4096	[V] COG4096 Type I site-specific restriction-modification system, R (restriction) subu	4394	0	0	0	1	0	1	Strain-specific
303			4393	0	0	0	1	0	1	Strain-specific
304	COG1737	[K] COG1737 Transcriptional regulators	4391	0	0	0	1	0	1	Strain-specific
305			4390	0	0	0	1	0	1	Strain-specific
306	COG4146	[R] COG4146 Predicted symporter	4388	0	0	0	1	0	1	Strain-specific
307			4387	0	0	0	1	0	1	Strain-specific
308	COG2801	[L] COG2801 Transposase and inactivated derivatives	4386	0	0	0	1	0	1	Strain-specific
309	COG0417	[L] COG0417 DNA polymerase elongation subunit (family B)	4385	0	0	0	1	0	1	Strain-specific
310			4381	0	0	0	1	0	1	Strain-specific
311	COG0417	[L] COG0417 DNA polymerase elongation subunit (family B)	4380	0	0	0	1	0	1	Strain-specific
312	COG1476	[K] COG1476 Predicted transcriptional regulators	4294	0	0	0	1	0	1	Strain-specific
313	COG5640	[O] COG5640 Secreted trypsin-like serine protease	4293	0	0	0	1	0	1	Strain-specific
314			4292	0	0	0	1	0	1	Strain-specific
315			4291	0	0	0	1	0	1	Strain-specific
316			4290	0	0	0	1	0	1	Strain-specific
317			4289	0	0	0	1	0	1	Strain-specific
318	COG0732	[V] COG0732 Restriction endonuclease S subunits	4288	0	0	0	1	0	1	Strain-specific
319	COG3883	[S] COG3883 Uncharacterized protein conserved in bacteria	4287	0	0	0	1	0	1	Strain-specific
320			4285	0	0	0	1	0	1	Strain-specific
321			4282	0	0	0	1	0	1	Strain-specific
322	COG2963	[L] COG2963 Transposase and inactivated derivatives	4279	0	0	0	1	0	1	Strain-specific
323	COG0021	[G] COG0021 Transketolase	4278	0	0	0	1	0	1	Strain-specific
324	COG1088	[M] COG1088 dTDP-D-glucose 4,6-dehydratase	4277	0	0	0	1	0	1	Strain-specific
325			4187	0	0	0	1	0	1	Strain-specific
326			4185	0	0	0	1	0	1	Strain-specific
327			4183	0	0	0	1	0	1	Strain-specific
328			4182	0	0	0	1	0	1	Strain-specific
329	COG0270	[L] COG0270 Site-specific DNA methylase	4181	0	0	0	1	0	1	Strain-specific
330			4180	0	0	0	1	0	1	Strain-specific
331	COG0673	[R] COG0673 Predicted dehydrogenases and related proteins	4179	0	0	0	1	0	1	Strain-specific
332	COG0399	[M] COG0399 Predicted pyridoxal phosphate-dependent enzyme apparently involv	4178	0	0	0	1	0	1	Strain-specific
333	COG3505	[U] COG3505 Type IV secretory pathway, VirD4 components	4177	0	0	0	1	0	1	Strain-specific
334	COG3715	[G] COG3715 Phosphotransferase system, mannose/fructose/N-acetylgalactosamin	4174	0	0	0	1	0	1	Strain-specific
335	COG3547	[L] COG3547 Transposase and inactivated derivatives	4172	0	0	0	1	0	1	Strain-specific
336			4171	0	0	0	1	0	1	Strain-specific
337			4169	0	0	0	1	0	1	Strain-specific
338			4168	0	0	0	1	0	1	Strain-specific
339	COG2207	[K] COG2207 AraC-type DNA-binding domain-containing proteins	4166	0	0	0	1	0	1	Strain-specific
340			4076	0	0	0	1	0	1	Strain-specific
341	COG4974	[L] COG4974 Site-specific recombinase XerD	4073	0	0	0	1	0	1	Strain-specific
342	COG1061	[KL] COG1061 DNA or RNA helicases of superfamily II	4072	0	0	0	1	0	1	Strain-specific
343			4071	0	0	0	1	0	1	Strain-specific
344			4069	0	0	0	1	0	1	Strain-specific
345			4068	0	0	0	1	0	1	Strain-specific
346			4067	0	0	0	1	0	1	Strain-specific
347	COG1519	[M] COG1519 3-deoxy-D-manno-octulosonic-acid transferase	4066	0	0	0	1	0	1	Strain-specific
348			4065	0	0	0	1	0	1	Strain-specific
349			4064	0	0	0	1	0	1	Strain-specific
350	COG3344	[L] COG3344 Retron-type reverse transcriptase	4062	0	0	0	1	0	1	Strain-specific
351	COG1446	[E] COG1446 Asparaginase	4058	0	0	0	1	0	1	Strain-specific
352			4056	0	0	0	1	0	1	Strain-specific
353			4055	0	0	0	1	0	1	Strain-specific
354	COG3613	[F] COG3613 Nucleoside 2-deoxyribosyltransferase	4054	0	0	0	1	0	1	Strain-specific
355	COG0625	[O] COG0625 Glutathione S-transferase	4053	0	0	0	1	0	1	Strain-specific
356			4052	0	0	0	1	0	1	Strain-specific
357			3975	0	0	0	1	0	1	Strain-specific
358	COG1132	[V] COG1132 ABC-type multidrug transport system, ATPase and permease compone	3972	0	0	0	1	0	1	Strain-specific
359	COG0507	[L] COG0507 ATP-dependent exoDNAse (exonuclease V), alpha subunit - helicase su	3971	0	0	0	1	0	1	Strain-specific
360	COG4605	[P] COG4605 ABC-type enterochelin transport system, permease component	3970	0	0	0	1	0	1	Strain-specific
361	COG0863	[L] COG0863 DNA modification methylase	3968	0	0	0	1	0	1	Strain-specific
362	COG3950	[R] COG3950 Predicted ATP-binding protein involved in virulence	3966	0	0	0	1	0	1	Strain-specific
363			3965	0	0	0	1	0	1	Strain-specific
364	COG0583	[K] COG0583 Transcriptional regulator	3964	0	0	0	1	0	1	Strain-specific
365			3963	0	0	0	1	0	1	Strain-specific
366	COG0286	[V] COG0286 Type I restriction-modification system methyltransferase subunit	3962	0	0	0	1	0	1	Strain-specific
367	COG3436	[L] COG3436 Transposase and inactivated derivatives	3961	0	0	0	1	0	1	Strain-specific
368	COG0553	[KL] COG0553 Superfamily II DNA/RNA helicases, SNF2 family	3958	0	0	0	1	0	1	Strain-specific
369	COG0583	[K] COG0583 Transcriptional regulator	3955	0	0	0	1	0	1	Strain-specific
370	COG0740	[OU] COG0740 Protease subunit of ATP-dependent Clp proteases	3954	0	0	0	1	0	1	Strain-specific
371			3952	0	0	0	1	0	1	Strain-specific
372			3950	0	0	0	1	0	1	Strain-specific
373	COG0603	[R] COG0603 Predicted PP-loop superfamily ATPase	3949	0	0	0	1	0	1	Strain-specific
374	COG1246	[E] COG1246 N-acetylglutamate synthase and related acetyltransferases	3948	0	0	0	1	0	1	Strain-specific
375	COG1629	[P] COG1629 Outer membrane receptor proteins, mostly Fe transport	3947	0	0	0	1	0	1	Strain-specific
376			3849	0	0	0	1	0	1	Strain-specific
377	COG4166	[E] COG4166 ABC-type oligopeptide transport system, periplasmic component	3846	0	0	0	1	0	1	Strain-specific
378	COG4607	[P] COG4607 ABC-type enterochelin transport system, periplasmic component	3845	0	0	0	1	0	1	Strain-specific
379	COG0050	[J] COG0050 GTPases - translation elongation factors	3844	0	0	0	1	0	1	Strain-specific
380			3842	0	0	0	1	0	1	Strain-specific
381			3840	0	0	0	1	0	1	Strain-specific
382			3838	0	0	0	1	0	1	Strain-specific
383	COG0011	[S] COG0011 Uncharacterized conserved protein	3837	0	0	0	1	0	1	Strain-specific
384	COG0338	[L] COG0338 Site-specific DNA methylase	3836	0	0	0	1	0	1	Strain-specific
385			3834	0	0	0	1	0	1	Strain-specific
386	COG2114	[T] COG2114 Adenylate cyclase, family 3 (some proteins contain HAMP domain)	3833	0	0	0	1	0	1	Strain-specific
387	COG0732	[V] COG0732 Restriction endonuclease S subunits	3832	0	0	0	1	0	1	Strain-specific
388	COG4771	[P] COG4771 Outer membrane receptor for ferrienterochelin and colicins	3831	0	0	0	1	0	1	Strain-specific
389	COG2826	[L] COG2826 Transposase and inactivated derivatives, IS30 family	3830	0	0	0	1	0	1	Strain-specific
390	COG4646	[KL] COG4646 DNA methylase	3829	0	0	0	1	0	1	Strain-specific
391	COG1192	[D] COG1192 ATPases involved in chromosome partitioning	3828	0	0	0	1	0	1	Strain-specific
392			3827	0	0	0	1	0	1	Strain-specific
393	COG1974	[KT] COG1974 SOS-response transcriptional repressors (RecA-mediated autopeptid	3729	0	0	0	1	0	1	Strain-specific
394			3727	0	0	0	1	0	1	Strain-specific
395			3725	0	0	0	1	0	1	Strain-specific
396	COG1112	[L] COG1112 Superfamily I DNA and RNA helicases and helicase subunits	3723	0	0	0	1	0	1	Strain-specific
397			3721	0	0	0	1	0	1	Strain-specific
398			3720	0	0	0	1	0	1	Strain-specific
399	COG0191	[G] COG0191 Fructose/tagatose bisphosphate aldolase	3719	0	0	0	1	0	1	Strain-specific
400			3718	0	0	0	1	0	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
401			3716	0	0	0	1	0	1	Strain-specific
402	COG1114	[E] COG1114 Branched-chain amino acid permeases	3715	0	0	0	1	0	1	Strain-specific
403			3713	0	0	0	1	0	1	Strain-specific
404	COG4973	[L] COG4973 Site-specific recombinase XerC	3712	0	0	0	1	0	1	Strain-specific
405			3708	0	0	0	1	0	1	Strain-specific
406	COG0722	[E] COG0722 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase	3707	0	0	0	1	0	1	Strain-specific
407	COG2804	[NU] COG2804 Type II secretory pathway, ATPase PulE/Tfp pilus assembly pathway	3706	0	0	0	1	0	1	Strain-specific
408			3702	0	0	0	1	0	1	Strain-specific
409			3610	0	0	0	1	0	1	Strain-specific
410			3609	0	0	0	1	0	1	Strain-specific
411			3607	0	0	0	1	0	1	Strain-specific
412	COG0399	[M] COG0399 Predicted pyridoxal phosphate-dependent enzyme apparently involv	3604	0	0	0	1	0	1	Strain-specific
413	COG3040	[M] COG3040 Bacterial lipocalin	3602	0	0	0	1	0	1	Strain-specific
414	COG3596	[R] COG3596 Predicted GTPase	3601	0	0	0	1	0	1	Strain-specific
415			3600	0	0	0	1	0	1	Strain-specific
416	COG1940	[KG] COG1940 Transcriptional regulator/sugar kinase	3599	0	0	0	1	0	1	Strain-specific
417			3598	0	0	0	1	0	1	Strain-specific
418			3595	0	0	0	1	0	1	Strain-specific
419	COG3307	[M] COG3307 Lipid A core - O-antigen ligase and related enzymes	3593	0	0	0	1	0	1	Strain-specific
420	COG0028	[EH] COG0028 Thiamine pyrophosphate-requiring enzymes [acetolactate synthase,	3592	0	0	0	1	0	1	Strain-specific
421	COG1012	[C] COG1012 NAD-dependent aldehyde dehydrogenases	3587	0	0	0	1	0	1	Strain-specific
422	COG4695	[S] COG4695 Phage-related protein	3583	0	0	0	1	0	1	Strain-specific
423			3579	0	0	0	1	0	1	Strain-specific
424			3484	0	0	0	1	0	1	Strain-specific
425	COG3972	[R] COG3972 Superfamily I DNA and RNA helicases	3483	0	0	0	1	0	1	Strain-specific
426			3482	0	0	0	1	0	1	Strain-specific
427			3481	0	0	0	1	0	1	Strain-specific
428			3480	0	0	0	1	0	1	Strain-specific
429			3479	0	0	0	1	0	1	Strain-specific
430			3478	0	0	0	1	0	1	Strain-specific
431	COG0399	[M] COG0399 Predicted pyridoxal phosphate-dependent enzyme apparently involv	3477	0	0	0	1	0	1	Strain-specific
432	COG4409	[G] COG4409 Neuraminidase (sialidase)	3476	0	0	0	1	0	1	Strain-specific
433			3475	0	0	0	1	0	1	Strain-specific
434	COG2856	[E] COG2856 Predicted Zn peptidase	3474	0	0	0	1	0	1	Strain-specific
435			3472	0	0	0	1	0	1	Strain-specific
436	COG0389	[L] COG0389 Nucleotidyltransferase/DNA polymerase involved in DNA repair	3469	0	0	0	1	0	1	Strain-specific
437	COG4604	[P] COG4604 ABC-type enterochelin transport system, ATPase component	3467	0	0	0	1	0	1	Strain-specific
438			3466	0	0	0	1	0	1	Strain-specific
439			3465	0	0	0	1	0	1	Strain-specific
440			3463	0	0	0	1	0	1	Strain-specific
441	COG1328	[F] COG1328 Oxygen-sensitive ribonucleoside-triphosphate reductase	3458	0	0	0	1	0	1	Strain-specific
442	COG3104	[E] COG3104 Dipeptide/tripeptide permease	3451	0	0	0	1	0	1	Strain-specific
443			3449	0	0	0	1	0	1	Strain-specific
444			3395	0	0	0	1	0	1	Strain-specific
445	COG0641	[R] COG0641 Arylsulfatase regulator (Fe-S oxidoreductase)	3391	0	0	0	1	0	1	Strain-specific
446			3386	0	0	0	1	0	1	Strain-specific
447	COG2378	[K] COG2378 Predicted transcriptional regulator	3383	0	0	0	1	0	1	Strain-specific
448			3379	0	0	0	1	0	1	Strain-specific
449			3370	0	0	0	1	0	1	Strain-specific
450			3367	0	0	0	1	0	1	Strain-specific
451			3366	0	0	0	1	0	1	Strain-specific
452			3225	0	0	0	1	0	1	Strain-specific
453			3224	0	0	0	1	0	1	Strain-specific
454	COG1737	[K] COG1737 Transcriptional regulators	3223	0	0	0	1	0	1	Strain-specific
455			3221	0	0	0	1	0	1	Strain-specific
456			3220	0	0	0	1	0	1	Strain-specific
457	COG1020	[Q] COG1020 Non-ribosomal peptide synthetase modules and related proteins	3219	0	0	0	1	0	1	Strain-specific
458			3217	0	0	0	1	0	1	Strain-specific
459	COG0346	[E] COG0346 Lactoylglutathione lyase and related lyases	3216	0	0	0	1	0	1	Strain-specific
460			3213	0	0	0	1	0	1	Strain-specific
461	COG3972	[R] COG3972 Superfamily I DNA and RNA helicases	3212	0	0	0	1	0	1	Strain-specific
462			3211	0	0	0	1	0	1	Strain-specific
463			3209	0	0	0	1	0	1	Strain-specific
464	COG0859	[M] COG0859 ADP-heptose:LPS heptosyltransferase	3208	0	0	0	1	0	1	Strain-specific
465			3207	0	0	0	1	0	1	Strain-specific
466			3205	0	0	0	1	0	1	Strain-specific
467	COG2253	[S] COG2253 Uncharacterized conserved protein	3204	0	0	0	1	0	1	Strain-specific
468			3200	0	0	0	1	0	1	Strain-specific
469	COG1112	[L] COG1112 Superfamily I DNA and RNA helicases and helicase subunits	3199	0	0	0	1	0	1	Strain-specific
470			3196	0	0	0	1	0	1	Strain-specific
471			3189	0	0	0	1	0	1	Strain-specific
472	COG0603	[R] COG0603 Predicted PP-loop superfamily ATPase	3187	0	0	0	1	0	1	Strain-specific
473	COG3158	[P] COG3158 K+ transporter	3183	0	0	0	1	0	1	Strain-specific
474			3182	0	0	0	1	0	1	Strain-specific
475	COG3385	[L] COG3385 FOG: Transposase and inactivated derivatives	3041	0	0	0	1	0	1	Strain-specific
476			3040	0	0	0	1	0	1	Strain-specific
477	COG1961	[L] COG1961 Site-specific recombinases, DNA invertase Pin homologs	3039	0	0	0	1	0	1	Strain-specific
478			3037	0	0	0	1	0	1	Strain-specific
479	COG5322	[R] COG5322 Predicted dehydrogenase	3036	0	0	0	1	0	1	Strain-specific
480			3034	0	0	0	1	0	1	Strain-specific
481	COG4646	[KL] COG4646 DNA methylase	3033	0	0	0	1	0	1	Strain-specific
482	COG1082	[G] COG1082 Sugar phosphate isomerases/epimerases	3030	0	0	0	1	0	1	Strain-specific
483	COG4191	[T] COG4191 Signal transduction histidine kinase regulating C4-dicarboxylate trans	3028	0	0	0	1	0	1	Strain-specific
484	COG3774	[M] COG3774 Mannosyltransferase OCH1 and related enzymes	3027	0	0	0	1	0	1	Strain-specific
485			3026	0	0	0	1	0	1	Strain-specific
486	COG4626	[R] COG4626 Phage terminase-like protein, large subunit	3024	0	0	0	1	0	1	Strain-specific
487	COG3344	[L] COG3344 Retron-type reverse transcriptase	3023	0	0	0	1	0	1	Strain-specific
488	COG1196	[D] COG1196 Chromosome segregation ATPases	3020	0	0	0	1	0	1	Strain-specific
489			3019	0	0	0	1	0	1	Strain-specific
490	COG3081	[R] COG3081 Nucleoid-associated protein	3016	0	0	0	1	0	1	Strain-specific
491			3014	0	0	0	1	0	1	Strain-specific
492	COG1714	[S] COG1714 Predicted membrane protein/domain	3001	0	0	0	1	0	1	Strain-specific
493			3000	0	0	0	1	0	1	Strain-specific
494	COG0818	[M] COG0818 Diacylglycerol kinase	2859	0	0	0	1	0	1	Strain-specific
495	COG1670	[J] COG1670 Acetyltransferases, including N-acetylases of ribosomal proteins	2858	0	0	0	1	0	1	Strain-specific
496	COG3550	[R] COG3550 Uncharacterized protein related to capsule biosynthesis enzymes	2856	0	0	0	1	0	1	Strain-specific
497			2855	0	0	0	1	0	1	Strain-specific
498			2853	0	0	0	1	0	1	Strain-specific
499	COG3774	[M] COG3774 Mannosyltransferase OCH1 and related enzymes	2852	0	0	0	1	0	1	Strain-specific
500	COG0472	[M] COG0472 UDP-N-acetylmuramyl pentapeptide phosphotransferase/UDP-N-ace	2850	0	0	0	1	0	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
501	COG0162	[J] COG0162 Tyrosyl-tRNA synthetase	2848	0	0	0	1	0	1	Strain-specific
502	COG1041	[L] COG1041 Predicted DNA modification methylase	2846	0	0	0	1	0	1	Strain-specific
503	COG3203	[M] COG3203 Outer membrane protein (porin)	2845	0	0	0	1	0	1	Strain-specific
504	COG3945	[S] COG3945 Uncharacterized conserved protein	2842	0	0	0	1	0	1	Strain-specific
505	COG1566	[V] COG1566 Multidrug resistance efflux pump	2840	0	0	0	1	0	1	Strain-specific
506			2832	0	0	0	1	0	1	Strain-specific
507			2829	0	0	0	1	0	1	Strain-specific
508	COG4604	[P] COG4604 ABC-type enterochelin transport system, ATPase component	2818	0	0	0	1	0	1	Strain-specific
509			2813	0	0	0	1	0	1	Strain-specific
510			2812	0	0	0	1	0	1	Strain-specific
511			2667	0	0	0	1	0	1	Strain-specific
512			2666	0	0	0	1	0	1	Strain-specific
513	COG2378	[K] COG2378 Predicted transcriptional regulator	2665	0	0	0	1	0	1	Strain-specific
514			2662	0	0	0	1	0	1	Strain-specific
515			2661	0	0	0	1	0	1	Strain-specific
516	COG0399	[M] COG0399 Predicted pyridoxal phosphate-dependent enzyme apparently involv	2659	0	0	0	1	0	1	Strain-specific
517	COG4771	[P] COG4771 Outer membrane receptor for ferrienterochelin and colicins	2657	0	0	0	1	0	1	Strain-specific
518			2656	0	0	0	1	0	1	Strain-specific
519	COG0507	[L] COG0507 ATP-dependent exoDNAse (exonuclease V), alpha subunit - helicase su	2654	0	0	0	1	0	1	Strain-specific
520			2653	0	0	0	1	0	1	Strain-specific
521	COG2801	[L] COG2801 Transposase and inactivated derivatives	2652	0	0	0	1	0	1	Strain-specific
522			2650	0	0	0	1	0	1	Strain-specific
523	COG2932	[K] COG2932 Predicted transcriptional regulator	2649	0	0	0	1	0	1	Strain-specific
524			2640	0	0	0	1	0	1	Strain-specific
525			2635	0	0	0	1	0	1	Strain-specific
526	COG2842	[R] COG2842 Uncharacterized ATPase, putative transposase	2627	0	0	0	1	0	1	Strain-specific
527	COG0625	[O] COG0625 Glutathione S-transferase	2623	0	0	0	1	0	1	Strain-specific
528			2462	0	0	0	1	0	1	Strain-specific
529			2461	0	0	0	1	0	1	Strain-specific
530	COG0629	[L] COG0629 Single-stranded DNA-binding protein	2460	0	0	0	1	0	1	Strain-specific
531	COG2148	[M] COG2148 Sugar transferases involved in lipopolysaccharide synthesis	2458	0	0	0	1	0	1	Strain-specific
532	COG1519	[M] COG1519 3-deoxy-D-manno-octulosonic-acid transferase	2457	0	0	0	1	0	1	Strain-specific
533	COG0125	[F] COG0125 Thymidylate kinase	2456	0	0	0	1	0	1	Strain-specific
534			2455	0	0	0	1	0	1	Strain-specific
535			2454	0	0	0	1	0	1	Strain-specific
536	COG1123	[R] COG1123 ATPase components of various ABC-type transport systems, contain d	2452	0	0	0	1	0	1	Strain-specific
537			2451	0	0	0	1	0	1	Strain-specific
538			2450	0	0	0	1	0	1	Strain-specific
539	COG1407	[R] COG1407 Predicted ICC-like phosphoesterases	2449	0	0	0	1	0	1	Strain-specific
540	COG0583	[K] COG0583 Transcriptional regulator	2448	0	0	0	1	0	1	Strain-specific
541			2446	0	0	0	1	0	1	Strain-specific
542	COG3575	[S] COG3575 Uncharacterized protein conserved in bacteria	2445	0	0	0	1	0	1	Strain-specific
543	COG3324	[R] COG3324 Predicted enzyme related to lactoylglutathione lyase	2443	0	0	0	1	0	1	Strain-specific
544	COG0463	[M] COG0463 Glycosyltransferases involved in cell wall biogenesis	2434	0	0	0	1	0	1	Strain-specific
545			2432	0	0	0	1	0	1	Strain-specific
546	COG1446	[E] COG1446 Asparaginase	2417	0	0	0	1	0	1	Strain-specific
547	COG4190	[K] COG4190 Predicted transcriptional regulator	2413	0	0	0	1	0	1	Strain-specific
548	COG3945	[S] COG3945 Uncharacterized conserved protein	2411	0	0	0	1	0	1	Strain-specific
549			2328	0	0	0	1	0	1	Strain-specific
550	COG4096	[V] COG4096 Type I site-specific restriction-modification system, R (restriction) subu	2327	0	0	0	1	0	1	Strain-specific
551			2324	0	0	0	1	0	1	Strain-specific
552	COG0625	[O] COG0625 Glutathione S-transferase	2323	0	0	0	1	0	1	Strain-specific
553	COG3324	[R] COG3324 Predicted enzyme related to lactoylglutathione lyase	2322	0	0	0	1	0	1	Strain-specific
554	COG3501	[S] COG3501 Uncharacterized protein conserved in bacteria	2319	0	0	0	1	0	1	Strain-specific
555	COG2404	[R] COG2404 Predicted phosphohydrolase (DHH superfamily)	2318	0	0	0	1	0	1	Strain-specific
556	COG0789	[K] COG0789 Predicted transcriptional regulators	2315	0	0	0	1	0	1	Strain-specific
557			2307	0	0	0	1	0	1	Strain-specific
558	COG3645	[S] COG3645 Uncharacterized phage-encoded protein	2300	0	0	0	1	0	1	Strain-specific
559	COG0607	[P] COG0607 Rhodanese-related sulfurtransferase	2280	0	0	0	1	0	1	Strain-specific
560	COG1961	[L] COG1961 Site-specific recombinases, DNA invertase Pin homologs	2274	0	0	0	1	0	1	Strain-specific
561			2269	0	0	0	1	0	1	Strain-specific
562			2265	0	0	0	1	0	1	Strain-specific
563			2264	0	0	0	1	0	1	Strain-specific
564			2251	0	0	0	1	0	1	Strain-specific
565			2245	0	0	0	1	0	1	Strain-specific
566	COG3505	[U] COG3505 Type IV secretory pathway, VirD4 components	2243	0	0	0	1	0	1	Strain-specific
567	COG3110	[S] COG3110 Uncharacterized protein conserved in bacteria	1958	0	0	0	1	0	1	Strain-specific
568			1956	0	0	0	1	0	1	Strain-specific
569	COG4453	[S] COG4453 Uncharacterized protein conserved in bacteria	1954	0	0	0	1	0	1	Strain-specific
570			1952	0	0	0	1	0	1	Strain-specific
571	COG3593	[L] COG3593 Predicted ATP-dependent endonuclease of the OLD family	1950	0	0	0	1	0	1	Strain-specific
572	COG1216	[R] COG1216 Predicted glycosyltransferases	1948	0	0	0	1	0	1	Strain-specific
573			1945	0	0	0	1	0	1	Strain-specific
574	COG4096	[V] COG4096 Type I site-specific restriction-modification system, R (restriction) subu	1940	0	0	0	1	0	1	Strain-specific
575	COG0863	[L] COG0863 DNA modification methylase	1937	0	0	0	1	0	1	Strain-specific
576	COG0191	[G] COG0191 Fructose/tagatose bisphosphate aldolase	1931	0	0	0	1	0	1	Strain-specific
577			1911	0	0	0	1	0	1	Strain-specific
578	COG1087	[M] COG1087 UDP-glucose 4-epimerase	1910	0	0	0	1	0	1	Strain-specific
579	COG4644	[L] COG4644 Transposase and inactivated derivatives, TnpA family	1904	0	0	0	1	0	1	Strain-specific
580	COG1199	[KL] COG1199 Rad3-related DNA helicases	1898	0	0	0	1	0	1	Strain-specific
581			1892	0	0	0	1	0	1	Strain-specific
582	COG1004	[M] COG1004 Predicted UDP-glucose 6-dehydrogenase	1877	0	0	0	1	0	1	Strain-specific
583	COG2113	[E] COG2113 ABC-type proline/glycine betaine transport systems, periplasmic comp	1867	0	0	0	1	0	1	Strain-specific
584			1865	0	0	0	1	0	1	Strain-specific
585	COG4026	[R] COG4026 Uncharacterized protein containing TOPRIM domain, potential nuclea	1559	0	0	0	1	0	1	Strain-specific
586	COG0286	[V] COG0286 Type I restriction-modification system methyltransferase subunit	1557	0	0	0	1	0	1	Strain-specific
587	COG1246	[E] COG1246 N-acetylglutamate synthase and related acetyltransferases	1555	0	0	0	1	0	1	Strain-specific
588	COG0399	[M] COG0399 Predicted pyridoxal phosphate-dependent enzyme apparently involv	1553	0	0	0	1	0	1	Strain-specific
589	COG3547	[L] COG3547 Transposase and inactivated derivatives	1552	0	0	0	1	0	1	Strain-specific
590	COG0732	[V] COG0732 Restriction endonuclease S subunits	1548	0	0	0	1	0	1	Strain-specific
591	COG3344	[L] COG3344 Retron-type reverse transcriptase	1539	0	0	0	1	0	1	Strain-specific
592			1537	0	0	0	1	0	1	Strain-specific
593	COG1820	[G] COG1820 N-acetylglucosamine-6-phosphate deacetylase	1531	0	0	0	1	0	1	Strain-specific
594	COG1974	[KT] COG1974 SOS-response transcriptional repressors (RecA-mediated autopeptid	1510	0	0	0	1	0	1	Strain-specific
595	COG2194	[R] COG2194 Predicted membrane-associated, metal-dependent hydrolase	1502	0	0	0	1	0	1	Strain-specific
596	COG3344	[L] COG3344 Retron-type reverse transcriptase	1493	0	0	0	1	0	1	Strain-specific
597			1491	0	0	0	1	0	1	Strain-specific
598	COG0328	[L] COG0328 Ribonuclease HI	1478	0	0	0	1	0	1	Strain-specific
599	COG2113	[E] COG2113 ABC-type proline/glycine betaine transport systems, periplasmic comp	1473	0	0	0	1	0	1	Strain-specific
600			1469	0	0	0	1	0	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
601			1358	0	0	0	1	0	1	Strain-specific
602			1357	0	0	0	1	0	1	Strain-specific
603	COG3436	[L] COG3436 Transposase and inactivated derivatives	1356	0	0	0	1	0	1	Strain-specific
604	COG4646	[KL] COG4646 DNA methylase	1355	0	0	0	1	0	1	Strain-specific
605	COG0863	[L] COG0863 DNA modification methylase	1354	0	0	0	1	0	1	Strain-specific
606			1352	0	0	0	1	0	1	Strain-specific
607	COG4973	[L] COG4973 Site-specific recombinase XerC	1351	0	0	0	1	0	1	Strain-specific
608			1349	0	0	0	1	0	1	Strain-specific
609	COG0433	[R] COG0433 Predicted ATPase	1348	0	0	0	1	0	1	Strain-specific
610	COG1002	[V] COG1002 Type II restriction enzyme, methylase subunits	1347	0	0	0	1	0	1	Strain-specific
611	COG0725	[P] COG0725 ABC-type molybdate transport system, periplasmic component	1343	0	0	0	1	0	1	Strain-specific
612			1342	0	0	0	1	0	1	Strain-specific
613	COG3593	[L] COG3593 Predicted ATP-dependent endonuclease of the OLD family	1340	0	0	0	1	0	1	Strain-specific
614	COG3904	[S] COG3904 Predicted periplasmic protein	1339	0	0	0	1	0	1	Strain-specific
615	COG4974	[L] COG4974 Site-specific recombinase XerD	1338	0	0	0	1	0	1	Strain-specific
616			1337	0	0	0	1	0	1	Strain-specific
617			1331	0	0	0	1	0	1	Strain-specific
618	COG2130	[R] COG2130 Putative NADP-dependent oxidoreductases	1330	0	0	0	1	0	1	Strain-specific
619			1248	0	0	0	1	0	1	Strain-specific
620			1247	0	0	0	1	0	1	Strain-specific
621			1246	0	0	0	1	0	1	Strain-specific
622			1245	0	0	0	1	0	1	Strain-specific
623			1242	0	0	0	1	0	1	Strain-specific
624	COG4974	[L] COG4974 Site-specific recombinase XerD	1241	0	0	0	1	0	1	Strain-specific
625			1237	0	0	0	1	0	1	Strain-specific
626	COG0582	[L] COG0582 Integrase	1236	0	0	0	1	0	1	Strain-specific
627			1234	0	0	0	1	0	1	Strain-specific
628			1231	0	0	0	1	0	1	Strain-specific
629	COG0399	[M] COG0399 Predicted pyridoxal phosphate-dependent enzyme apparently involv	1230	0	0	0	1	0	1	Strain-specific
630	COG1670	[J] COG1670 Acetyltransferases, including N-acetylases of ribosomal proteins	1229	0	0	0	1	0	1	Strain-specific
631	COG0593	[L] COG0593 ATPase involved in DNA replication initiation	1228	0	0	0	1	0	1	Strain-specific
632	COG1209	[M] COG1209 dTDP-glucose pyrophosphorylase	1226	0	0	0	1	0	1	Strain-specific
633	COG1566	[V] COG1566 Multidrug resistance efflux pump	1224	0	0	0	1	0	1	Strain-specific
634	COG3409	[M] COG3409 Putative peptidoglycan-binding domain-containing protein	1223	0	0	0	1	0	1	Strain-specific
635			1220	0	0	0	1	0	1	Strain-specific
636	COG3469	[G] COG3469 Chitinase	1219	0	0	0	1	0	1	Strain-specific
637	COG3328	[L] COG3328 Transposase and inactivated derivatives	1217	0	0	0	1	0	1	Strain-specific
638			1126	0	0	0	1	0	1	Strain-specific
639	COG4096	[V] COG4096 Type I site-specific restriction-modification system, R (restriction) subu	1125	0	0	0	1	0	1	Strain-specific
640			1124	0	0	0	1	0	1	Strain-specific
641	COG4974	[L] COG4974 Site-specific recombinase XerD	1123	0	0	0	1	0	1	Strain-specific
642			1122	0	0	0	1	0	1	Strain-specific
643			1121	0	0	0	1	0	1	Strain-specific
644	COG0640	[K] COG0640 Predicted transcriptional regulators	1120	0	0	0	1	0	1	Strain-specific
645	COG1082	[G] COG1082 Sugar phosphate isomerases/epimerases	1119	0	0	0	1	0	1	Strain-specific
646			1116	0	0	0	1	0	1	Strain-specific
647			1115	0	0	0	1	0	1	Strain-specific
648	COG3344	[L] COG3344 Retron-type reverse transcriptase	1110	0	0	0	1	0	1	Strain-specific
649			1109	0	0	0	1	0	1	Strain-specific
650			1107	0	0	0	1	0	1	Strain-specific
651			1106	0	0	0	1	0	1	Strain-specific
652	COG0433	[R] COG0433 Predicted ATPase	1105	0	0	0	1	0	1	Strain-specific
653	COG3550	[R] COG3550 Uncharacterized protein related to capsule biosynthesis enzymes	1103	0	0	0	1	0	1	Strain-specific
654	COG5492	[N] COG5492 Bacterial surface proteins containing Ig-like domains	1102	0	0	0	1	0	1	Strain-specific
655			1101	0	0	0	1	0	1	Strain-specific
656	COG4974	[L] COG4974 Site-specific recombinase XerD	1099	0	0	0	1	0	1	Strain-specific
657	COG2226	[H] COG2226 Methylase involved in ubiquinone/menaquinone biosynthesis	1098	0	0	0	1	0	1	Strain-specific
658			1097	0	0	0	1	0	1	Strain-specific
659			1003	0	0	0	1	0	1	Strain-specific
660	COG3499	[R] COG3499 Phage protein U	1002	0	0	0	1	0	1	Strain-specific
661			1001	0	0	0	1	0	1	Strain-specific
662			719	0	0	0	1	0	1	Strain-specific
663	COG4605	[P] COG4605 ABC-type enterochelin transport system, permease component	715	0	0	0	1	0	1	Strain-specific
664			714	0	0	0	1	0	1	Strain-specific
665	COG2194	[R] COG2194 Predicted membrane-associated, metal-dependent hydrolase	713	0	0	0	1	0	1	Strain-specific
666	COG3886	[L] COG3886 Predicted HKD family nuclease	712	0	0	0	1	0	1	Strain-specific
667	COG3206	[M] COG3206 Uncharacterized protein involved in exopolysaccharide biosynthesis	711	0	0	0	1	0	1	Strain-specific
668			701	0	0	0	1	0	1	Strain-specific
669			693	0	0	0	1	0	1	Strain-specific
670			686	0	0	0	1	0	1	Strain-specific
671	COG3486	[Q] COG3486 Lysine/ornithine N-monooxygenase	666	0	0	0	1	0	1	Strain-specific
672			657	0	0	0	1	0	1	Strain-specific
673			648	0	0	0	1	0	1	Strain-specific
674	COG4421	[G] COG4421 Capsular polysaccharide biosynthesis protein	639	0	0	0	1	0	1	Strain-specific
675	COG0697	[GER] COG0697 Permeases of the drug/metabolite transporter (DMT) superfamily	620	0	0	0	1	0	1	Strain-specific
676			225	0	0	0	1	0	1	Strain-specific
677	COG0859	[M] COG0859 ADP-heptose:LPS heptosyltransferase	222	0	0	0	1	0	1	Strain-specific
678	COG0438	[M] COG0438 Glycosyltransferase	221	0	0	0	1	0	1	Strain-specific
679			220	0	0	0	1	0	1	Strain-specific
680	COG1225	[O] COG1225 Peroxiredoxin	219	0	0	0	1	0	1	Strain-specific
681	COG0834	[ET] COG0834 ABC-type amino acid transport/signal transduction systems, periplas	218	0	0	0	1	0	1	Strain-specific
682	COG0178	[L] COG0178 Excinuclease ATPase subunit	213	0	0	0	1	0	1	Strain-specific
683	COG0433	[R] COG0433 Predicted ATPase	207	0	0	0	1	0	1	Strain-specific
684			192	0	0	0	1	0	1	Strain-specific
685	COG1061	[KL] COG1061 DNA or RNA helicases of superfamily II	168	0	0	0	1	0	1	Strain-specific
686			158	0	0	0	1	0	1	Strain-specific
687			118	0	0	0	1	0	1	Strain-specific
688			6259	0	1	0	0	0	1	Strain-specific
689			6256	0	1	0	0	0	1	Strain-specific
690	COG4973	[L] COG4973 Site-specific recombinase XerC	6002	0	1	0	0	0	1	Strain-specific
691			5966	0	1	0	0	0	1	Strain-specific
692	COG4989	[R] COG4989 Predicted oxidoreductase	5359	0	1	0	0	0	1	Strain-specific
693	COG2333	[R] COG2333 Predicted hydrolase (metallo-beta-lactamase superfamily)	5350	0	1	0	0	0	1	Strain-specific
694			5348	0	1	0	0	0	1	Strain-specific
695	COG0840	[NT] COG0840 Methyl-accepting chemotaxis protein	5220	0	1	0	0	0	1	Strain-specific
696	COG4823	[V] COG4823 Abortive infection bacteriophage resistance protein	5219	0	1	0	0	0	1	Strain-specific
697			5218	0	1	0	0	0	1	Strain-specific
698			5066	0	1	0	0	0	1	Strain-specific
699	COG1196	[D] COG1196 Chromosome segregation ATPases	4931	0	1	0	0	0	1	Strain-specific
700	COG4942	[D] COG4942 Membrane-bound metallopeptidase	4777	0	1	0	0	0	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
701			4753	0	1	0	0	0	1	Strain-specific
702	COG4974	[L] COG4974 Site-specific recombinase XerD	4607	0	1	0	0	0	1	Strain-specific
703			4602	0	1	0	0	0	1	Strain-specific
704			4591	0	1	0	0	0	1	Strain-specific
705	COG1396	[K] COG1396 Predicted transcriptional regulators	4431	0	1	0	0	0	1	Strain-specific
706			4423	0	1	0	0	0	1	Strain-specific
707			4410	0	1	0	0	0	1	Strain-specific
708	COG1061	[KL] COG1061 DNA or RNA helicases of superfamily II	4299	0	1	0	0	0	1	Strain-specific
709			4297	0	1	0	0	0	1	Strain-specific
710			4192	0	1	0	0	0	1	Strain-specific
711	COG0582	[L] COG0582 Integrase	4191	0	1	0	0	0	1	Strain-specific
712	COG3311	[K] COG3311 Predicted transcriptional regulator	4085	0	1	0	0	0	1	Strain-specific
713			4084	0	1	0	0	0	1	Strain-specific
714	COG4974	[L] COG4974 Site-specific recombinase XerD	4078	0	1	0	0	0	1	Strain-specific
715			3978	0	1	0	0	0	1	Strain-specific
716			3976	0	1	0	0	0	1	Strain-specific
717			3859	0	1	0	0	0	1	Strain-specific
718	COG1401	[V] COG1401 GTPase subunit of restriction endonuclease	3854	0	1	0	0	0	1	Strain-specific
719	COG5483	[S] COG5483 Uncharacterized conserved protein	3850	0	1	0	0	0	1	Strain-specific
720			3738	0	1	0	0	0	1	Strain-specific
721	COG4268	[V] COG4268 McrBC 5-methylcytosine restriction system component	3737	0	1	0	0	0	1	Strain-specific
722			3734	0	1	0	0	0	1	Strain-specific
723	COG0419	[L] COG0419 ATPase involved in DNA repair	3733	0	1	0	0	0	1	Strain-specific
724			3730	0	1	0	0	0	1	Strain-specific
725			3619	0	1	0	0	0	1	Strain-specific
726	COG0232	[F] COG0232 dGTP triphosphohydrolase	3615	0	1	0	0	0	1	Strain-specific
727	COG3344	[L] COG3344 Retron-type reverse transcriptase	3614	0	1	0	0	0	1	Strain-specific
728			3613	0	1	0	0	0	1	Strain-specific
729	COG4974	[L] COG4974 Site-specific recombinase XerD	3612	0	1	0	0	0	1	Strain-specific
730			3611	0	1	0	0	0	1	Strain-specific
731	COG5340	[K] COG5340 Predicted transcriptional regulator	3495	0	1	0	0	0	1	Strain-specific
732	COG0847	[L] COG0847 DNA polymerase III, epsilon subunit and related 3'-5' exonucleases	3491	0	1	0	0	0	1	Strain-specific
733	COG3593	[L] COG3593 Predicted ATP-dependent endonuclease of the OLD family	3490	0	1	0	0	0	1	Strain-specific
734			3488	0	1	0	0	0	1	Strain-specific
735	COG1196	[D] COG1196 Chromosome segregation ATPases	3486	0	1	0	0	0	1	Strain-specific
736			3249	0	1	0	0	0	1	Strain-specific
737			3247	0	1	0	0	0	1	Strain-specific
738			3230	0	1	0	0	0	1	Strain-specific
739			3226	0	1	0	0	0	1	Strain-specific
740	COG0270	[L] COG0270 Site-specific DNA methylase	3071	0	1	0	0	0	1	Strain-specific
741			3053	0	1	0	0	0	1	Strain-specific
742			2882	0	1	0	0	0	1	Strain-specific
743			2871	0	1	0	0	0	1	Strain-specific
744	COG3081	[R] COG3081 Nucleoid-associated protein	2692	0	1	0	0	0	1	Strain-specific
745	COG2931	[Q] COG2931 RTX toxins and related Ca2+-binding proteins	2680	0	1	0	0	0	1	Strain-specific
746			2489	0	1	0	0	0	1	Strain-specific
747			2474	0	1	0	0	0	1	Strain-specific
748	COG0507	[L] COG0507 ATP-dependent exoDNAse (exonuclease V), alpha subunit - helicase su	2471	0	1	0	0	0	1	Strain-specific
749			2385	0	1	0	0	0	1	Strain-specific
750	COG2246	[S] COG2246 Predicted membrane protein	2379	0	1	0	0	0	1	Strain-specific
751			2348	0	1	0	0	0	1	Strain-specific
752			1977	0	1	0	0	0	1	Strain-specific
753	COG2964	[S] COG2964 Uncharacterized protein conserved in bacteria	1578	0	1	0	0	0	1	Strain-specific
754	COG0847	[L] COG0847 DNA polymerase III, epsilon subunit and related 3'-5' exonucleases	1361	0	1	0	0	0	1	Strain-specific
755			1252	0	1	0	0	0	1	Strain-specific
756	COG1961	[L] COG1961 Site-specific recombinases, DNA invertase Pin homologs	1250	0	1	0	0	0	1	Strain-specific
757	COG3883	[S] COG3883 Uncharacterized protein conserved in bacteria	1134	0	1	0	0	0	1	Strain-specific
758	COG0582	[L] COG0582 Integrase	1129	0	1	0	0	0	1	Strain-specific
759			1008	0	1	0	0	0	1	Strain-specific
760	COG1061	[KL] COG1061 DNA or RNA helicases of superfamily II	1005	0	1	0	0	0	1	Strain-specific
761			789	0	1	0	0	0	1	Strain-specific
762			788	0	1	0	0	0	1	Strain-specific
763			6338	0	0	1	0	0	1	Strain-specific
764	COG0727	[R] COG0727 Predicted Fe-S-cluster oxidoreductase	6334	0	0	1	0	0	1	Strain-specific
765	COG0308	[E] COG0308 Aminopeptidase N	6332	0	0	1	0	0	1	Strain-specific
766	COG0643	[NT] COG0643 Chemotaxis protein histidine kinase and related kinases	6327	0	0	1	0	0	1	Strain-specific
767			6325	0	0	1	0	0	1	Strain-specific
768	COG1061	[KL] COG1061 DNA or RNA helicases of superfamily II	6322	0	0	1	0	0	1	Strain-specific
769	COG0640	[K] COG0640 Predicted transcriptional regulators	6316	0	0	1	0	0	1	Strain-specific
770			6315	0	0	1	0	0	1	Strain-specific
771	COG5652	[S] COG5652 Predicted integral membrane protein	6309	0	0	1	0	0	1	Strain-specific
772			6077	0	0	1	0	0	1	Strain-specific
773	COG5281	[S] COG5281 Phage-related minor tail protein	6075	0	0	1	0	0	1	Strain-specific
774			6074	0	0	1	0	0	1	Strain-specific
775	COG3516	[S] COG3516 Uncharacterized protein conserved in bacteria	6063	0	0	1	0	0	1	Strain-specific
776			6062	0	0	1	0	0	1	Strain-specific
777			6060	0	0	1	0	0	1	Strain-specific
778			6047	0	0	1	0	0	1	Strain-specific
779			5805	0	0	1	0	0	1	Strain-specific
780	COG3311	[K] COG3311 Predicted transcriptional regulator	5801	0	0	1	0	0	1	Strain-specific
781			5796	0	0	1	0	0	1	Strain-specific
782	COG4088	[F] COG4088 Predicted nucleotide kinase	5794	0	0	1	0	0	1	Strain-specific
783	COG3250	[G] COG3250 Beta-galactosidase/beta-glucuronidase	5789	0	0	1	0	0	1	Strain-specific
784			5787	0	0	1	0	0	1	Strain-specific
785	COG0389	[L] COG0389 Nucleotidyltransferase/DNA polymerase involved in DNA repair	5786	0	0	1	0	0	1	Strain-specific
786			5778	0	0	1	0	0	1	Strain-specific
787			5767	0	0	1	0	0	1	Strain-specific
788			5763	0	0	1	0	0	1	Strain-specific
789			5495	0	0	1	0	0	1	Strain-specific
790	COG4974	[L] COG4974 Site-specific recombinase XerD	5494	0	0	1	0	0	1	Strain-specific
791	COG0582	[L] COG0582 Integrase	5493	0	0	1	0	0	1	Strain-specific
792	COG1955	[NU] COG1955 Archaeal flagella assembly protein J	5491	0	0	1	0	0	1	Strain-specific
793	COG4733	[S] COG4733 Phage-related protein, tail component	5487	0	0	1	0	0	1	Strain-specific
794			5486	0	0	1	0	0	1	Strain-specific
795			5485	0	0	1	0	0	1	Strain-specific
796			5480	0	0	1	0	0	1	Strain-specific
797	COG3250	[G] COG3250 Beta-galactosidase/beta-glucuronidase	5479	0	0	1	0	0	1	Strain-specific
798			5476	0	0	1	0	0	1	Strain-specific
799	COG2356	[L] COG2356 Endonuclease I	5474	0	0	1	0	0	1	Strain-specific
800			5394	0	0	1	0	0	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
801	COG1157	[NU] COG1157 Flagellar biosynthesis/type III secretory pathway ATPase	5391	0	0	1	0	0	1	Strain-specific
802			5389	0	0	1	0	0	1	Strain-specific
803	COG2175	[Q] COG2175 Probable taurine catabolism dioxygenase	5387	0	0	1	0	0	1	Strain-specific
804			5386	0	0	1	0	0	1	Strain-specific
805	COG1278	[K] COG1278 Cold shock proteins	5385	0	0	1	0	0	1	Strain-specific
806	COG1401	[V] COG1401 GTPase subunit of restriction endonuclease	5384	0	0	1	0	0	1	Strain-specific
807	COG0471	[P] COG0471 Di- and tricarboxylate transporters	5383	0	0	1	0	0	1	Strain-specific
808	COG3886	[L] COG3886 Predicted HKD family nuclease	5378	0	0	1	0	0	1	Strain-specific
809	COG2433	[S] COG2433 Uncharacterized conserved protein	5262	0	0	1	0	0	1	Strain-specific
810			5261	0	0	1	0	0	1	Strain-specific
811			5260	0	0	1	0	0	1	Strain-specific
812	COG4278	[S] COG4278 Uncharacterized conserved protein	5259	0	0	1	0	0	1	Strain-specific
813	COG3673	[S] COG3673 Uncharacterized conserved protein	5258	0	0	1	0	0	1	Strain-specific
814			5256	0	0	1	0	0	1	Strain-specific
815			5255	0	0	1	0	0	1	Strain-specific
816			5250	0	0	1	0	0	1	Strain-specific
817	COG1020	[Q] COG1020 Non-ribosomal peptide synthetase modules and related proteins	5249	0	0	1	0	0	1	Strain-specific
818	COG1476	[K] COG1476 Predicted transcriptional regulators	5135	0	0	1	0	0	1	Strain-specific
819			5134	0	0	1	0	0	1	Strain-specific
820			5133	0	0	1	0	0	1	Strain-specific
821	COG3747	[L] COG3747 Phage terminase, small subunit	5132	0	0	1	0	0	1	Strain-specific
822	COG0300	[R] COG0300 Short-chain dehydrogenases of various substrate specificities	5130	0	0	1	0	0	1	Strain-specific
823	COG2095	[U] COG2095 Multiple antibiotic transporter	5129	0	0	1	0	0	1	Strain-specific
824			5127	0	0	1	0	0	1	Strain-specific
825			5126	0	0	1	0	0	1	Strain-specific
826	COG1020	[Q] COG1020 Non-ribosomal peptide synthetase modules and related proteins	5125	0	0	1	0	0	1	Strain-specific
827	COG4974	[L] COG4974 Site-specific recombinase XerD	5122	0	0	1	0	0	1	Strain-specific
828			5120	0	0	1	0	0	1	Strain-specific
829			5119	0	0	1	0	0	1	Strain-specific
830	COG0207	[F] COG0207 Thymidylate synthase	5113	0	0	1	0	0	1	Strain-specific
831	COG2963	[L] COG2963 Transposase and inactivated derivatives	4975	0	0	1	0	0	1	Strain-specific
832	COG3642	[T] COG3642 Mn2+-dependent serine/threonine protein kinase	4974	0	0	1	0	0	1	Strain-specific
833	COG2801	[L] COG2801 Transposase and inactivated derivatives	4973	0	0	1	0	0	1	Strain-specific
834			4971	0	0	1	0	0	1	Strain-specific
835	COG4626	[R] COG4626 Phage terminase-like protein, large subunit	4970	0	0	1	0	0	1	Strain-specific
836	COG0207	[F] COG0207 Thymidylate synthase	4967	0	0	1	0	0	1	Strain-specific
837			4966	0	0	1	0	0	1	Strain-specific
838	COG4564	[T] COG4564 Signal transduction histidine kinase	4959	0	0	1	0	0	1	Strain-specific
839	COG2963	[L] COG2963 Transposase and inactivated derivatives	4826	0	0	1	0	0	1	Strain-specific
840	COG4646	[KL] COG4646 DNA methylase	4821	0	0	1	0	0	1	Strain-specific
841			4819	0	0	1	0	0	1	Strain-specific
842	COG3344	[L] COG3344 Retron-type reverse transcriptase	4818	0	0	1	0	0	1	Strain-specific
843			4813	0	0	1	0	0	1	Strain-specific
844			4806	0	0	1	0	0	1	Strain-specific
845	COG0840	[NT] COG0840 Methyl-accepting chemotaxis protein	4802	0	0	1	0	0	1	Strain-specific
846			4800	0	0	1	0	0	1	Strain-specific
847			4658	0	0	1	0	0	1	Strain-specific
848	COG3081	[R] COG3081 Nucleoid-associated protein	4657	0	0	1	0	0	1	Strain-specific
849			4656	0	0	1	0	0	1	Strain-specific
850	COG1961	[L] COG1961 Site-specific recombinases, DNA invertase Pin homologs	4654	0	0	1	0	0	1	Strain-specific
851	COG4695	[S] COG4695 Phage-related protein	4653	0	0	1	0	0	1	Strain-specific
852	COG2333	[R] COG2333 Predicted hydrolase (metallo-beta-lactamase superfamily)	4651	0	0	1	0	0	1	Strain-specific
853			4646	0	0	1	0	0	1	Strain-specific
854	COG2932	[K] COG2932 Predicted transcriptional regulator	4642	0	0	1	0	0	1	Strain-specific
855			4636	0	0	1	0	0	1	Strain-specific
856			4633	0	0	1	0	0	1	Strain-specific
857			4499	0	0	1	0	0	1	Strain-specific
858			4498	0	0	1	0	0	1	Strain-specific
859	COG1092	[R] COG1092 Predicted SAM-dependent methyltransferases	4497	0	0	1	0	0	1	Strain-specific
860			4495	0	0	1	0	0	1	Strain-specific
861			4494	0	0	1	0	0	1	Strain-specific
862	COG0740	[OU] COG0740 Protease subunit of ATP-dependent Clp proteases	4493	0	0	1	0	0	1	Strain-specific
863	COG2199	[T] COG2199 FOG: GGDEF domain	4489	0	0	1	0	0	1	Strain-specific
864			4480	0	0	1	0	0	1	Strain-specific
865			4479	0	0	1	0	0	1	Strain-specific
866			4333	0	0	1	0	0	1	Strain-specific
867	COG2801	[L] COG2801 Transposase and inactivated derivatives	4332	0	0	1	0	0	1	Strain-specific
868			4331	0	0	1	0	0	1	Strain-specific
869	COG1566	[V] COG1566 Multidrug resistance efflux pump	4329	0	0	1	0	0	1	Strain-specific
870			4328	0	0	1	0	0	1	Strain-specific
871	COG0189	[HJ] COG0189 Glutathione synthase/Ribosomal protein S6 modification enzyme (gl	4325	0	0	1	0	0	1	Strain-specific
872	COG0697	[GER] COG0697 Permeases of the drug/metabolite transporter (DMT) superfamily	4324	0	0	1	0	0	1	Strain-specific
873			4323	0	0	1	0	0	1	Strain-specific
874	COG4974	[L] COG4974 Site-specific recombinase XerD	4319	0	0	1	0	0	1	Strain-specific
875			4317	0	0	1	0	0	1	Strain-specific
876	COG3531	[O] COG3531 Predicted protein-disulfide isomerase	4316	0	0	1	0	0	1	Strain-specific
877			4315	0	0	1	0	0	1	Strain-specific
878			4228	0	0	1	0	0	1	Strain-specific
879			4227	0	0	1	0	0	1	Strain-specific
880	COG3081	[R] COG3081 Nucleoid-associated protein	4225	0	0	1	0	0	1	Strain-specific
881			4222	0	0	1	0	0	1	Strain-specific
882			4219	0	0	1	0	0	1	Strain-specific
883	COG0730	[R] COG0730 Predicted permeases	4218	0	0	1	0	0	1	Strain-specific
884	COG0471	[P] COG0471 Di- and tricarboxylate transporters	4216	0	0	1	0	0	1	Strain-specific
885			4215	0	0	1	0	0	1	Strain-specific
886			4214	0	0	1	0	0	1	Strain-specific
887	COG1629	[P] COG1629 Outer membrane receptor proteins, mostly Fe transport	4213	0	0	1	0	0	1	Strain-specific
888	COG0534	[V] COG0534 Na+-driven multidrug efflux pump	4212	0	0	1	0	0	1	Strain-specific
889	COG0534	[V] COG0534 Na+-driven multidrug efflux pump	4209	0	0	1	0	0	1	Strain-specific
890			4121	0	0	1	0	0	1	Strain-specific
891			4120	0	0	1	0	0	1	Strain-specific
892			4116	0	0	1	0	0	1	Strain-specific
893			4113	0	0	1	0	0	1	Strain-specific
894			4110	0	0	1	0	0	1	Strain-specific
895	COG1770	[E] COG1770 Protease II	4108	0	0	1	0	0	1	Strain-specific
896	COG1961	[L] COG1961 Site-specific recombinases, DNA invertase Pin homologs	4107	0	0	1	0	0	1	Strain-specific
897			4104	0	0	1	0	0	1	Strain-specific
898	COG0534	[V] COG0534 Na+-driven multidrug efflux pump	4100	0	0	1	0	0	1	Strain-specific
899			4009	0	0	1	0	0	1	Strain-specific
900			4008	0	0	1	0	0	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
901			4003	0	0	1	0	0	1	Strain-specific
902			4002	0	0	1	0	0	1	Strain-specific
903			4001	0	0	1	0	0	1	Strain-specific
904	COG0290	[J] COG0290 Translation initiation factor 3 (IF-3)	3998	0	0	1	0	0	1	Strain-specific
905			3997	0	0	1	0	0	1	Strain-specific
906	COG2932	[K] COG2932 Predicted transcriptional regulator	3892	0	0	1	0	0	1	Strain-specific
907	COG3415	[L] COG3415 Transposase and inactivated derivatives	3891	0	0	1	0	0	1	Strain-specific
908			3888	0	0	1	0	0	1	Strain-specific
909			3885	0	0	1	0	0	1	Strain-specific
910	COG0446	[R] COG0446 Uncharacterized NAD(FAD)-dependent dehydrogenases	3884	0	0	1	0	0	1	Strain-specific
911			3883	0	0	1	0	0	1	Strain-specific
912			3878	0	0	1	0	0	1	Strain-specific
913	COG0847	[L] COG0847 DNA polymerase III, epsilon subunit and related 3'-5' exonucleases	3877	0	0	1	0	0	1	Strain-specific
914	COG0612	[R] COG0612 Predicted Zn-dependent peptidases	3875	0	0	1	0	0	1	Strain-specific
915	COG0110	[R] COG0110 Acetyltransferase (isoleucine patch superfamily)	3874	0	0	1	0	0	1	Strain-specific
916			3768	0	0	1	0	0	1	Strain-specific
917			3767	0	0	1	0	0	1	Strain-specific
918	COG0272	[L] COG0272 NAD-dependent DNA ligase (contains BRCT domain type II)	3766	0	0	1	0	0	1	Strain-specific
919			3764	0	0	1	0	0	1	Strain-specific
920			3763	0	0	1	0	0	1	Strain-specific
921	COG2203	[T] COG2203 FOG: GAF domain	3761	0	0	1	0	0	1	Strain-specific
922	COG3926	[R] COG3926 Putative secretion activating protein	3760	0	0	1	0	0	1	Strain-specific
923			3758	0	0	1	0	0	1	Strain-specific
924	COG0153	[G] COG0153 Galactokinase	3756	0	0	1	0	0	1	Strain-specific
925	COG5633	[R] COG5633 Predicted periplasmic lipoprotein	3647	0	0	1	0	0	1	Strain-specific
926			3641	0	0	1	0	0	1	Strain-specific
927	COG0178	[L] COG0178 Excinuclease ATPase subunit	3640	0	0	1	0	0	1	Strain-specific
928			3639	0	0	1	0	0	1	Strain-specific
929			3634	0	0	1	0	0	1	Strain-specific
930	COG1450	[NU] COG1450 Type II secretory pathway, component PulD	3525	0	0	1	0	0	1	Strain-specific
931	COG3547	[L] COG3547 Transposase and inactivated derivatives	3524	0	0	1	0	0	1	Strain-specific
932	COG0641	[R] COG0641 Arylsulfatase regulator (Fe-S oxidoreductase)	3523	0	0	1	0	0	1	Strain-specific
933	COG0270	[L] COG0270 Site-specific DNA methylase	3521	0	0	1	0	0	1	Strain-specific
934			3517	0	0	1	0	0	1	Strain-specific
935	COG2963	[L] COG2963 Transposase and inactivated derivatives	3308	0	0	1	0	0	1	Strain-specific
936	COG2931	[Q] COG2931 RTX toxins and related Ca2+-binding proteins	3306	0	0	1	0	0	1	Strain-specific
937	COG4640	[S] COG4640 Predicted membrane protein	3304	0	0	1	0	0	1	Strain-specific
938			3303	0	0	1	0	0	1	Strain-specific
939	COG4653	[R] COG4653 Predicted phage phi-C31 gp36 major capsid-like protein	3302	0	0	1	0	0	1	Strain-specific
940	COG3613	[F] COG3613 Nucleoside 2-deoxyribosyltransferase	3299	0	0	1	0	0	1	Strain-specific
941			3298	0	0	1	0	0	1	Strain-specific
942			3297	0	0	1	0	0	1	Strain-specific
943			3296	0	0	1	0	0	1	Strain-specific
944			3293	0	0	1	0	0	1	Strain-specific
945	COG4974	[L] COG4974 Site-specific recombinase XerD	3287	0	0	1	0	0	1	Strain-specific
946			3284	0	0	1	0	0	1	Strain-specific
947	COG3322	[T] COG3322 Predicted periplasmic ligand-binding sensor domain	3281	0	0	1	0	0	1	Strain-specific
948			3120	0	0	1	0	0	1	Strain-specific
949			3119	0	0	1	0	0	1	Strain-specific
950	COG0419	[L] COG0419 ATPase involved in DNA repair	3117	0	0	1	0	0	1	Strain-specific
951			3115	0	0	1	0	0	1	Strain-specific
952			3111	0	0	1	0	0	1	Strain-specific
953	COG1974	[KT] COG1974 SOS-response transcriptional repressors (RecA-mediated autopeptid	3109	0	0	1	0	0	1	Strain-specific
954			3098	0	0	1	0	0	1	Strain-specific
955	COG2199	[T] COG2199 FOG: GGDEF domain	3097	0	0	1	0	0	1	Strain-specific
956	COG1378	[K] COG1378 Predicted transcriptional regulators	2941	0	0	1	0	0	1	Strain-specific
957			2939	0	0	1	0	0	1	Strain-specific
958			2938	0	0	1	0	0	1	Strain-specific
959			2936	0	0	1	0	0	1	Strain-specific
960			2935	0	0	1	0	0	1	Strain-specific
961			2934	0	0	1	0	0	1	Strain-specific
962			2933	0	0	1	0	0	1	Strain-specific
963			2931	0	0	1	0	0	1	Strain-specific
964			2929	0	0	1	0	0	1	Strain-specific
965			2927	0	0	1	0	0	1	Strain-specific
966	COG5293	[S] COG5293 Uncharacterized protein conserved in bacteria	2926	0	0	1	0	0	1	Strain-specific
967			2921	0	0	1	0	0	1	Strain-specific
968			2919	0	0	1	0	0	1	Strain-specific
969	COG5433	[L] COG5433 Transposase	2914	0	0	1	0	0	1	Strain-specific
970			2748	0	0	1	0	0	1	Strain-specific
971			2747	0	0	1	0	0	1	Strain-specific
972			2742	0	0	1	0	0	1	Strain-specific
973	COG2862	[S] COG2862 Predicted membrane protein	2741	0	0	1	0	0	1	Strain-specific
974	COG0840	[NT] COG0840 Methyl-accepting chemotaxis protein	2740	0	0	1	0	0	1	Strain-specific
975			2739	0	0	1	0	0	1	Strain-specific
976			2736	0	0	1	0	0	1	Strain-specific
977	COG0204	[I] COG0204 1-acyl-sn-glycerol-3-phosphate acyltransferase	2730	0	0	1	0	0	1	Strain-specific
978			2553	0	0	1	0	0	1	Strain-specific
979			2552	0	0	1	0	0	1	Strain-specific
980			2551	0	0	1	0	0	1	Strain-specific
981	COG2862	[S] COG2862 Predicted membrane protein	2547	0	0	1	0	0	1	Strain-specific
982			2546	0	0	1	0	0	1	Strain-specific
983	COG1061	[KL] COG1061 DNA or RNA helicases of superfamily II	2544	0	0	1	0	0	1	Strain-specific
984			2543	0	0	1	0	0	1	Strain-specific
985			2538	0	0	1	0	0	1	Strain-specific
986			2536	0	0	1	0	0	1	Strain-specific
987			2535	0	0	1	0	0	1	Strain-specific
988	COG4625	[S] COG4625 Uncharacterized protein with a C-terminal OMP (outer membrane pro	2534	0	0	1	0	0	1	Strain-specific
989			2527	0	0	1	0	0	1	Strain-specific
990			2150	0	0	1	0	0	1	Strain-specific
991			2149	0	0	1	0	0	1	Strain-specific
992			2148	0	0	1	0	0	1	Strain-specific
993	COG0687	[E] COG0687 Spermidine/putrescine-binding periplasmic protein	2143	0	0	1	0	0	1	Strain-specific
994			2135	0	0	1	0	0	1	Strain-specific
995	COG1479	[S] COG1479 Uncharacterized conserved protein	2133	0	0	1	0	0	1	Strain-specific
996	COG3344	[L] COG3344 Retron-type reverse transcriptase	2128	0	0	1	0	0	1	Strain-specific
997	COG3180	[R] COG3180 Putative ammonia monooxygenase	2124	0	0	1	0	0	1	Strain-specific
998			2123	0	0	1	0	0	1	Strain-specific
999	COG1725	[K] COG1725 Predicted transcriptional regulators	2113	0	0	1	0	0	1	Strain-specific
1000	COG1979	[C] COG1979 Uncharacterized oxidoreductases, Fe-dependent alcohol dehydrogena	2108	0	0	1	0	0	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
1001			2098	0	0	1	0	0	1	Strain-specific
1002			1764	0	0	1	0	0	1	Strain-specific
1003	COG3645	[S] COG3645 Uncharacterized phage-encoded protein	1763	0	0	1	0	0	1	Strain-specific
1004	COG0050	[J] COG0050 GTPases - translation elongation factors	1757	0	0	1	0	0	1	Strain-specific
1005	COG4974	[L] COG4974 Site-specific recombinase XerD	1755	0	0	1	0	0	1	Strain-specific
1006			1751	0	0	1	0	0	1	Strain-specific
1007	COG2095	[U] COG2095 Multiple antibiotic transporter	1749	0	0	1	0	0	1	Strain-specific
1008	COG0286	[V] COG0286 Type I restriction-modification system methyltransferase subunit	1748	0	0	1	0	0	1	Strain-specific
1009			1745	0	0	1	0	0	1	Strain-specific
1010			1743	0	0	1	0	0	1	Strain-specific
1011			1740	0	0	1	0	0	1	Strain-specific
1012	COG3180	[R] COG3180 Putative ammonia monooxygenase	1739	0	0	1	0	0	1	Strain-specific
1013	COG3415	[L] COG3415 Transposase and inactivated derivatives	1738	0	0	1	0	0	1	Strain-specific
1014			1735	0	0	1	0	0	1	Strain-specific
1015			1721	0	0	1	0	0	1	Strain-specific
1016			1705	0	0	1	0	0	1	Strain-specific
1017			1700	0	0	1	0	0	1	Strain-specific
1018	COG4733	[S] COG4733 Phage-related protein, tail component	1394	0	0	1	0	0	1	Strain-specific
1019	COG4672	[S] COG4672 Phage-related protein	1393	0	0	1	0	0	1	Strain-specific
1020	COG2244	[R] COG2244 Membrane protein involved in the export of O-antigen and teichoic a	1391	0	0	1	0	0	1	Strain-specific
1021			1390	0	0	1	0	0	1	Strain-specific
1022			1389	0	0	1	0	0	1	Strain-specific
1023			1388	0	0	1	0	0	1	Strain-specific
1024			1386	0	0	1	0	0	1	Strain-specific
1025			1385	0	0	1	0	0	1	Strain-specific
1026	COG3774	[M] COG3774 Mannosyltransferase OCH1 and related enzymes	1383	0	0	1	0	0	1	Strain-specific
1027	COG1015	[G] COG1015 Phosphopentomutase	1382	0	0	1	0	0	1	Strain-specific
1028	COG1763	[H] COG1763 Molybdopterin-guanine dinucleotide biosynthesis protein	1381	0	0	1	0	0	1	Strain-specific
1029			1380	0	0	1	0	0	1	Strain-specific
1030			1282	0	0	1	0	0	1	Strain-specific
1031			1280	0	0	1	0	0	1	Strain-specific
1032	COG0507	[L] COG0507 ATP-dependent exoDNAse (exonuclease V), alpha subunit - helicase su	1278	0	0	1	0	0	1	Strain-specific
1033	COG2378	[K] COG2378 Predicted transcriptional regulator	1276	0	0	1	0	0	1	Strain-specific
1034	COG1061	[KL] COG1061 DNA or RNA helicases of superfamily II	1274	0	0	1	0	0	1	Strain-specific
1035			1272	0	0	1	0	0	1	Strain-specific
1036			1271	0	0	1	0	0	1	Strain-specific
1037			1164	0	0	1	0	0	1	Strain-specific
1038			1163	0	0	1	0	0	1	Strain-specific
1039			1162	0	0	1	0	0	1	Strain-specific
1040	COG0606	[O] COG0606 Predicted ATPase with chaperone activity	1161	0	0	1	0	0	1	Strain-specific
1041			1160	0	0	1	0	0	1	Strain-specific
1042			1159	0	0	1	0	0	1	Strain-specific
1043	COG1328	[F] COG1328 Oxygen-sensitive ribonucleoside-triphosphate reductase	1157	0	0	1	0	0	1	Strain-specific
1044			1154	0	0	1	0	0	1	Strain-specific
1045	COG2274	[V] COG2274 ABC-type bacteriocin/lantibiotic exporters, contain an N-terminal dou	1046	0	0	1	0	0	1	Strain-specific
1046			1045	0	0	1	0	0	1	Strain-specific
1047			1043	0	0	1	0	0	1	Strain-specific
1048	COG0210	[L] COG0210 Superfamily I DNA and RNA helicases	1041	0	0	1	0	0	1	Strain-specific
1049			1039	0	0	1	0	0	1	Strain-specific
1050	COG4974	[L] COG4974 Site-specific recombinase XerD	1038	0	0	1	0	0	1	Strain-specific
1051			1035	0	0	1	0	0	1	Strain-specific
1052			964	0	0	1	0	0	1	Strain-specific
1053			963	0	0	1	0	0	1	Strain-specific
1054			962	0	0	1	0	0	1	Strain-specific
1055			955	0	0	1	0	0	1	Strain-specific
1056			950	0	0	1	0	0	1	Strain-specific
1057			946	0	0	1	0	0	1	Strain-specific
1058	COG2801	[L] COG2801 Transposase and inactivated derivatives	942	0	0	1	0	0	1	Strain-specific
1059			937	0	0	1	0	0	1	Strain-specific
1060			910	0	0	1	0	0	1	Strain-specific
1061			902	0	0	1	0	0	1	Strain-specific
1062			511	0	0	1	0	0	1	Strain-specific
1063	COG4723	[S] COG4723 Phage-related protein, tail component	509	0	0	1	0	0	1	Strain-specific
1064			504	0	0	1	0	0	1	Strain-specific
1065	COG0050	[J] COG0050 GTPases - translation elongation factors	502	0	0	1	0	0	1	Strain-specific
1066			501	0	0	1	0	0	1	Strain-specific
1067			497	0	0	1	0	0	1	Strain-specific
1068	COG2963	[L] COG2963 Transposase and inactivated derivatives	494	0	0	1	0	0	1	Strain-specific
1069	COG2317	[E] COG2317 Zn-dependent carboxypeptidase	491	0	0	1	0	0	1	Strain-specific
1070	COG1396	[K] COG1396 Predicted transcriptional regulators	489	0	0	1	0	0	1	Strain-specific
1071			484	0	0	1	0	0	1	Strain-specific
1072			475	0	0	1	0	0	1	Strain-specific
1073			6394	1	0	0	0	0	1	Strain-specific
1074	COG0841	[V] COG0841 Cation/multidrug efflux pump	6393	1	0	0	0	0	1	Strain-specific
1075			6392	1	0	0	0	0	1	Strain-specific
1076	COG3306	[M] COG3306 Glycosyltransferase involved in LPS biosynthesis	6391	1	0	0	0	0	1	Strain-specific
1077	COG0456	[R] COG0456 Acetyltransferases	6390	1	0	0	0	0	1	Strain-specific
1078			6388	1	0	0	0	0	1	Strain-specific
1079			6387	1	0	0	0	0	1	Strain-specific
1080	COG4928	[R] COG4928 Predicted P-loop ATPase	6386	1	0	0	0	0	1	Strain-specific
1081			6385	1	0	0	0	0	1	Strain-specific
1082	COG3436	[L] COG3436 Transposase and inactivated derivatives	6383	1	0	0	0	0	1	Strain-specific
1083	COG1129	[G] COG1129 ABC-type sugar transport system, ATPase component	6382	1	0	0	0	0	1	Strain-specific
1084	COG1715	[V] COG1715 Restriction endonuclease	6381	1	0	0	0	0	1	Strain-specific
1085			6378	1	0	0	0	0	1	Strain-specific
1086			6377	1	0	0	0	0	1	Strain-specific
1087			6376	1	0	0	0	0	1	Strain-specific
1088			6370	1	0	0	0	0	1	Strain-specific
1089	COG3762	[S] COG3762 Predicted membrane protein	6368	1	0	0	0	0	1	Strain-specific
1090	COG3440	[V] COG3440 Predicted restriction endonuclease	6367	1	0	0	0	0	1	Strain-specific
1091			6366	1	0	0	0	0	1	Strain-specific
1092			6365	1	0	0	0	0	1	Strain-specific
1093			6364	1	0	0	0	0	1	Strain-specific
1094			6363	1	0	0	0	0	1	Strain-specific
1095			6362	1	0	0	0	0	1	Strain-specific
1096			6361	1	0	0	0	0	1	Strain-specific
1097			6360	1	0	0	0	0	1	Strain-specific
1098			6359	1	0	0	0	0	1	Strain-specific
1099			6358	1	0	0	0	0	1	Strain-specific
1100	COG0582	[L] COG0582 Integrase	6357	1	0	0	0	0	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
1101	COG1737	[K] COG1737 Transcriptional regulators	6356	1	0	0	0	0	1	Strain-specific
1102			6353	1	0	0	0	0	1	Strain-specific
1103	COG0697	[GER] COG0697 Permeases of the drug/metabolite transporter (DMT) superfamily	6352	1	0	0	0	0	1	Strain-specific
1104			6351	1	0	0	0	0	1	Strain-specific
1105	COG0178	[L] COG0178 Excinuclease ATPase subunit	6350	1	0	0	0	0	1	Strain-specific
1106			6348	1	0	0	0	0	1	Strain-specific
1107			6346	1	0	0	0	0	1	Strain-specific
1108	COG0841	[V] COG0841 Cation/multidrug efflux pump	6342	1	0	0	0	0	1	Strain-specific
1109			6340	1	0	0	0	0	1	Strain-specific
1110	COG2194	[R] COG2194 Predicted membrane-associated, metal-dependent hydrolase	6339	1	0	0	0	0	1	Strain-specific
1111			6155	1	0	0	0	0	1	Strain-specific
1112	COG0235	[G] COG0235 Ribulose-5-phosphate 4-epimerase and related epimerases and aldol	6153	1	0	0	0	0	1	Strain-specific
1113			6149	1	0	0	0	0	1	Strain-specific
1114	COG1388	[M] COG1388 FOG: LysM repeat	6148	1	0	0	0	0	1	Strain-specific
1115	COG3041	[S] COG3041 Uncharacterized protein conserved in bacteria	6145	1	0	0	0	0	1	Strain-specific
1116			6142	1	0	0	0	0	1	Strain-specific
1117	COG3962	[E] COG3962 Acetolactate synthase	6140	1	0	0	0	0	1	Strain-specific
1118	COG4646	[KL] COG4646 DNA methylase	6139	1	0	0	0	0	1	Strain-specific
1119			6138	1	0	0	0	0	1	Strain-specific
1120	COG0859	[M] COG0859 ADP-heptose:LPS heptosyltransferase	6137	1	0	0	0	0	1	Strain-specific
1121	COG0399	[M] COG0399 Predicted pyridoxal phosphate-dependent enzyme apparently involv	6136	1	0	0	0	0	1	Strain-specific
1122			6131	1	0	0	0	0	1	Strain-specific
1123	COG3038	[C] COG3038 Cytochrome B561	6130	1	0	0	0	0	1	Strain-specific
1124			6128	1	0	0	0	0	1	Strain-specific
1125	COG1172	[G] COG1172 Ribose/xylose/arabinose/galactoside ABC-type transport systems, per	6127	1	0	0	0	0	1	Strain-specific
1126			6126	1	0	0	0	0	1	Strain-specific
1127	COG2207	[K] COG2207 AraC-type DNA-binding domain-containing proteins	6124	1	0	0	0	0	1	Strain-specific
1128	COG3593	[L] COG3593 Predicted ATP-dependent endonuclease of the OLD family	6123	1	0	0	0	0	1	Strain-specific
1129	COG2199	[T] COG2199 FOG: GGDEF domain	6122	1	0	0	0	0	1	Strain-specific
1130			6121	1	0	0	0	0	1	Strain-specific
1131	COG1061	[KL] COG1061 DNA or RNA helicases of superfamily II	6120	1	0	0	0	0	1	Strain-specific
1132	COG0433	[R] COG0433 Predicted ATPase	6117	1	0	0	0	0	1	Strain-specific
1133			6115	1	0	0	0	0	1	Strain-specific
1134	COG5635	[T] COG5635 Predicted NTPase (NACHT family)	6113	1	0	0	0	0	1	Strain-specific
1135			6112	1	0	0	0	0	1	Strain-specific
1136	COG3506	[S] COG3506 Uncharacterized conserved protein	6111	1	0	0	0	0	1	Strain-specific
1137			6110	1	0	0	0	0	1	Strain-specific
1138			6108	1	0	0	0	0	1	Strain-specific
1139			6107	1	0	0	0	0	1	Strain-specific
1140	COG1028	[IQR] COG1028 Dehydrogenases with different specificities (related to short-chain a	6106	1	0	0	0	0	1	Strain-specific
1141			6103	1	0	0	0	0	1	Strain-specific
1142	COG5935		6102	1	0	0	0	0	1	Strain-specific
1143			6101	1	0	0	0	0	1	Strain-specific
1144			6100	1	0	0	0	0	1	Strain-specific
1145			6098	1	0	0	0	0	1	Strain-specific
1146	COG0110	[R] COG0110 Acetyltransferase (isoleucine patch superfamily)	6097	1	0	0	0	0	1	Strain-specific
1147			6095	1	0	0	0	0	1	Strain-specific
1148			6094	1	0	0	0	0	1	Strain-specific
1149			6093	1	0	0	0	0	1	Strain-specific
1150	COG2131	[F] COG2131 Deoxycytidylate deaminase	6092	1	0	0	0	0	1	Strain-specific
1151	COG2026	[JD] COG2026 Cytotoxic translational repressor of toxin-antitoxin stability system	6091	1	0	0	0	0	1	Strain-specific
1152	COG0732	[V] COG0732 Restriction endonuclease S subunits	6090	1	0	0	0	0	1	Strain-specific
1153	COG0286	[V] COG0286 Type I restriction-modification system methyltransferase subunit	6085	1	0	0	0	0	1	Strain-specific
1154	COG2842	[R] COG2842 Uncharacterized ATPase, putative transposase	6084	1	0	0	0	0	1	Strain-specific
1155	COG4771	[P] COG4771 Outer membrane receptor for ferrienterochelin and colicins	6083	1	0	0	0	0	1	Strain-specific
1156			6082	1	0	0	0	0	1	Strain-specific
1157	COG3593	[L] COG3593 Predicted ATP-dependent endonuclease of the OLD family	6081	1	0	0	0	0	1	Strain-specific
1158	COG0180	[J] COG0180 Tryptophanyl-tRNA synthetase	6079	1	0	0	0	0	1	Strain-specific
1159	COG5009	[M] COG5009 Membrane carboxypeptidase/penicillin-binding protein	5887	1	0	0	0	0	1	Strain-specific
1160	COG0270	[L] COG0270 Site-specific DNA methylase	5884	1	0	0	0	0	1	Strain-specific
1161	COG0394	[T] COG0394 Protein-tyrosine-phosphatase	5883	1	0	0	0	0	1	Strain-specific
1162			5877	1	0	0	0	0	1	Strain-specific
1163			5876	1	0	0	0	0	1	Strain-specific
1164			5871	1	0	0	0	0	1	Strain-specific
1165	COG4646	[KL] COG4646 DNA methylase	5870	1	0	0	0	0	1	Strain-specific
1166			5869	1	0	0	0	0	1	Strain-specific
1167	COG4128	[R] COG4128 Zonula occludens toxin	5867	1	0	0	0	0	1	Strain-specific
1168	COG2148	[M] COG2148 Sugar transferases involved in lipopolysaccharide synthesis	5866	1	0	0	0	0	1	Strain-specific
1169			5865	1	0	0	0	0	1	Strain-specific
1170	COG0270	[L] COG0270 Site-specific DNA methylase	5863	1	0	0	0	0	1	Strain-specific
1171	COG1541	[H] COG1541 Coenzyme F390 synthetase	5862	1	0	0	0	0	1	Strain-specific
1172			5861	1	0	0	0	0	1	Strain-specific
1173			5860	1	0	0	0	0	1	Strain-specific
1174	COG3550	[R] COG3550 Uncharacterized protein related to capsule biosynthesis enzymes	5858	1	0	0	0	0	1	Strain-specific
1175	COG1946	[I] COG1946 Acyl-CoA thioesterase	5856	1	0	0	0	0	1	Strain-specific
1176	COG2931	[Q] COG2931 RTX toxins and related Ca2+-binding proteins	5855	1	0	0	0	0	1	Strain-specific
1177			5853	1	0	0	0	0	1	Strain-specific
1178			5851	1	0	0	0	0	1	Strain-specific
1179			5846	1	0	0	0	0	1	Strain-specific
1180	COG2333	[R] COG2333 Predicted hydrolase (metallo-beta-lactamase superfamily)	5843	1	0	0	0	0	1	Strain-specific
1181	COG3181	[S] COG3181 Uncharacterized protein conserved in bacteria	5842	1	0	0	0	0	1	Strain-specific
1182	COG1512	[R] COG1512 Beta-propeller domains of methanol dehydrogenase type	5841	1	0	0	0	0	1	Strain-specific
1183	COG1309	[K] COG1309 Transcriptional regulator	5840	1	0	0	0	0	1	Strain-specific
1184	COG3593	[L] COG3593 Predicted ATP-dependent endonuclease of the OLD family	5839	1	0	0	0	0	1	Strain-specific
1185			5836	1	0	0	0	0	1	Strain-specific
1186			5835	1	0	0	0	0	1	Strain-specific
1187			5834	1	0	0	0	0	1	Strain-specific
1188	COG0178	[L] COG0178 Excinuclease ATPase subunit	5829	1	0	0	0	0	1	Strain-specific
1189	COG2318	[S] COG2318 Uncharacterized protein conserved in bacteria	5828	1	0	0	0	0	1	Strain-specific
1190			5827	1	0	0	0	0	1	Strain-specific
1191			5826	1	0	0	0	0	1	Strain-specific
1192	COG0840	[NT] COG0840 Methyl-accepting chemotaxis protein	5825	1	0	0	0	0	1	Strain-specific
1193	COG1898	[M] COG1898 dTDP-4-dehydrorhamnose 3,5-epimerase and related enzymes	5824	1	0	0	0	0	1	Strain-specific
1194	COG0456	[R] COG0456 Acetyltransferases	5823	1	0	0	0	0	1	Strain-specific
1195			5822	1	0	0	0	0	1	Strain-specific
1196			5821	1	0	0	0	0	1	Strain-specific
1197	COG0732	[V] COG0732 Restriction endonuclease S subunits	5820	1	0	0	0	0	1	Strain-specific
1198			5819	1	0	0	0	0	1	Strain-specific
1199	COG4913	[S] COG4913 Uncharacterized protein conserved in bacteria	5818	1	0	0	0	0	1	Strain-specific
1200	COG3593	[L] COG3593 Predicted ATP-dependent endonuclease of the OLD family	5813	1	0	0	0	0	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
1201			5812	1	0	0	0	0	1	Strain-specific
1202			5809	1	0	0	0	0	1	Strain-specific
1203	COG0451	[MG] COG0451 Nucleoside-diphosphate-sugar epimerases	5808	1	0	0	0	0	1	Strain-specific
1204	COG1004	[M] COG1004 Predicted UDP-glucose 6-dehydrogenase	5807	1	0	0	0	0	1	Strain-specific
1205	COG3501	[S] COG3501 Uncharacterized protein conserved in bacteria	5585	1	0	0	0	0	1	Strain-specific
1206	COG3547	[L] COG3547 Transposase and inactivated derivatives	5584	1	0	0	0	0	1	Strain-specific
1207	COG5009	[M] COG5009 Membrane carboxypeptidase/penicillin-binding protein	5581	1	0	0	0	0	1	Strain-specific
1208			5580	1	0	0	0	0	1	Strain-specific
1209			5575	1	0	0	0	0	1	Strain-specific
1210	COG0366	[G] COG0366 Glycosidases	5570	1	0	0	0	0	1	Strain-specific
1211	COG0270	[L] COG0270 Site-specific DNA methylase	5569	1	0	0	0	0	1	Strain-specific
1212	COG3081	[R] COG3081 Nucleoid-associated protein	5568	1	0	0	0	0	1	Strain-specific
1213	COG2199	[T] COG2199 FOG: GGDEF domain	5567	1	0	0	0	0	1	Strain-specific
1214			5565	1	0	0	0	0	1	Strain-specific
1215	COG1596	[M] COG1596 Periplasmic protein involved in polysaccharide export	5564	1	0	0	0	0	1	Strain-specific
1216			5561	1	0	0	0	0	1	Strain-specific
1217	COG3385	[L] COG3385 FOG: Transposase and inactivated derivatives	5559	1	0	0	0	0	1	Strain-specific
1218	COG3549	[R] COG3549 Plasmid maintenance system killer protein	5557	1	0	0	0	0	1	Strain-specific
1219	COG5421	[L] COG5421 Transposase	5553	1	0	0	0	0	1	Strain-specific
1220	COG1715	[V] COG1715 Restriction endonuclease	5550	1	0	0	0	0	1	Strain-specific
1221	COG1127	[Q] COG1127 ABC-type transport system involved in resistance to organic solvents,	5546	1	0	0	0	0	1	Strain-specific
1222	COG1595	[K] COG1595 DNA-directed RNA polymerase specialized sigma subunit, sigma24 hor	5542	1	0	0	0	0	1	Strain-specific
1223			5540	1	0	0	0	0	1	Strain-specific
1224			5539	1	0	0	0	0	1	Strain-specific
1225	COG2931	[Q] COG2931 RTX toxins and related Ca2+-binding proteins	5538	1	0	0	0	0	1	Strain-specific
1226	COG1512	[R] COG1512 Beta-propeller domains of methanol dehydrogenase type	5537	1	0	0	0	0	1	Strain-specific
1227			5536	1	0	0	0	0	1	Strain-specific
1228			5535	1	0	0	0	0	1	Strain-specific
1229			5534	1	0	0	0	0	1	Strain-specific
1230			5533	1	0	0	0	0	1	Strain-specific
1231			5532	1	0	0	0	0	1	Strain-specific
1232	COG1452	[M] COG1452 Organic solvent tolerance protein OstA	5529	1	0	0	0	0	1	Strain-specific
1233	COG1609	[K] COG1609 Transcriptional regulators	5528	1	0	0	0	0	1	Strain-specific
1234	COG2095	[U] COG2095 Multiple antibiotic transporter	5527	1	0	0	0	0	1	Strain-specific
1235			5526	1	0	0	0	0	1	Strain-specific
1236			5525	1	0	0	0	0	1	Strain-specific
1237	COG4646	[KL] COG4646 DNA methylase	5524	1	0	0	0	0	1	Strain-specific
1238	COG2905	[T] COG2905 Predicted signal-transduction protein containing cAMP-binding and C	5523	1	0	0	0	0	1	Strain-specific
1239			5522	1	0	0	0	0	1	Strain-specific
1240			5521	1	0	0	0	0	1	Strain-specific
1241	COG5161	[A] COG5161 Pre-mRNA cleavage and polyadenylation specificity factor	5520	1	0	0	0	0	1	Strain-specific
1242			5519	1	0	0	0	0	1	Strain-specific
1243			5518	1	0	0	0	0	1	Strain-specific
1244			5517	1	0	0	0	0	1	Strain-specific
1245			5516	1	0	0	0	0	1	Strain-specific
1246	COG4823	[V] COG4823 Abortive infection bacteriophage resistance protein	5513	1	0	0	0	0	1	Strain-specific
1247	COG4943	[T] COG4943 Predicted signal transduction protein containing sensor and EAL doma	5512	1	0	0	0	0	1	Strain-specific
1248	COG4974	[L] COG4974 Site-specific recombinase XerD	5508	1	0	0	0	0	1	Strain-specific
1249			5507	1	0	0	0	0	1	Strain-specific
1250			5506	1	0	0	0	0	1	Strain-specific
1251	COG0084	[L] COG0084 Mg-dependent DNase	5502	1	0	0	0	0	1	Strain-specific
1252	COG1216	[R] COG1216 Predicted glycosyltransferases	5501	1	0	0	0	0	1	Strain-specific
1253	COG0251	[J] COG0251 Putative translation initiation inhibitor, yjgF family	5500	1	0	0	0	0	1	Strain-specific
1254			5395	1	0	0	0	0	1	Strain-specific
1255	COG0631	[T] COG0631 Serine/threonine protein phosphatase	5310	1	0	0	0	0	1	Strain-specific
1256			5309	1	0	0	0	0	1	Strain-specific
1257			5308	1	0	0	0	0	1	Strain-specific
1258			5307	1	0	0	0	0	1	Strain-specific
1259	COG0673	[R] COG0673 Predicted dehydrogenases and related proteins	5306	1	0	0	0	0	1	Strain-specific
1260			5305	1	0	0	0	0	1	Strain-specific
1261	COG0840	[NT] COG0840 Methyl-accepting chemotaxis protein	5304	1	0	0	0	0	1	Strain-specific
1262	COG0367	[E] COG0367 Asparagine synthase (glutamine-hydrolyzing)	5303	1	0	0	0	0	1	Strain-specific
1263			5302	1	0	0	0	0	1	Strain-specific
1264	COG1044	[M] COG1044 UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	5301	1	0	0	0	0	1	Strain-specific
1265	COG4974	[L] COG4974 Site-specific recombinase XerD	5300	1	0	0	0	0	1	Strain-specific
1266	COG1883	[C] COG1883 Na+-transporting methylmalonyl-CoA/oxaloacetate decarboxylase, be	5298	1	0	0	0	0	1	Strain-specific
1267			5297	1	0	0	0	0	1	Strain-specific
1268	COG1328	[F] COG1328 Oxygen-sensitive ribonucleoside-triphosphate reductase	5296	1	0	0	0	0	1	Strain-specific
1269			5291	1	0	0	0	0	1	Strain-specific
1270	COG0476	[H] COG0476 Dinucleotide-utilizing enzymes involved in molybdopterin and thiami	5290	1	0	0	0	0	1	Strain-specific
1271			5289	1	0	0	0	0	1	Strain-specific
1272			5287	1	0	0	0	0	1	Strain-specific
1273			5286	1	0	0	0	0	1	Strain-specific
1274			5285	1	0	0	0	0	1	Strain-specific
1275			5284	1	0	0	0	0	1	Strain-specific
1276			5283	1	0	0	0	0	1	Strain-specific
1277	COG1961	[L] COG1961 Site-specific recombinases, DNA invertase Pin homologs	5282	1	0	0	0	0	1	Strain-specific
1278			5281	1	0	0	0	0	1	Strain-specific
1279			5280	1	0	0	0	0	1	Strain-specific
1280	COG3311	[K] COG3311 Predicted transcriptional regulator	5279	1	0	0	0	0	1	Strain-specific
1281			5278	1	0	0	0	0	1	Strain-specific
1282	COG2334	[R] COG2334 Putative homoserine kinase type II (protein kinase fold)	5277	1	0	0	0	0	1	Strain-specific
1283			5276	1	0	0	0	0	1	Strain-specific
1284			5275	1	0	0	0	0	1	Strain-specific
1285	COG3897	[R] COG3897 Predicted methyltransferase	5274	1	0	0	0	0	1	Strain-specific
1286	COG0489	[D] COG0489 ATPases involved in chromosome partitioning	5272	1	0	0	0	0	1	Strain-specific
1287	COG1451	[R] COG1451 Predicted metal-dependent hydrolase	5271	1	0	0	0	0	1	Strain-specific
1288	COG3344	[L] COG3344 Retron-type reverse transcriptase	5270	1	0	0	0	0	1	Strain-specific
1289	COG2194	[R] COG2194 Predicted membrane-associated, metal-dependent hydrolase	5269	1	0	0	0	0	1	Strain-specific
1290			5268	1	0	0	0	0	1	Strain-specific
1291	COG3658	[C] COG3658 Cytochrome b	5267	1	0	0	0	0	1	Strain-specific
1292	COG1502	[I] COG1502 Phosphatidylserine/phosphatidylglycerophosphate/cardiolipin syntha	5266	1	0	0	0	0	1	Strain-specific
1293	COG1922	[M] COG1922 Teichoic acid biosynthesis proteins	5265	1	0	0	0	0	1	Strain-specific
1294	COG4795	[U] COG4795 Type II secretory pathway, component PulJ	5181	1	0	0	0	0	1	Strain-specific
1295			5180	1	0	0	0	0	1	Strain-specific
1296			5179	1	0	0	0	0	1	Strain-specific
1297	COG0697	[GER] COG0697 Permeases of the drug/metabolite transporter (DMT) superfamily	5178	1	0	0	0	0	1	Strain-specific
1298			5177	1	0	0	0	0	1	Strain-specific
1299	COG3306	[M] COG3306 Glycosyltransferase involved in LPS biosynthesis	5176	1	0	0	0	0	1	Strain-specific
1300			5175	1	0	0	0	0	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
1301			5174	1	0	0	0	0	1	Strain-specific
1302	COG0438	[M] COG0438 Glycosyltransferase	5173	1	0	0	0	0	1	Strain-specific
1303	COG3344	[L] COG3344 Retron-type reverse transcriptase	5172	1	0	0	0	0	1	Strain-specific
1304	COG0472	[M] COG0472 UDP-N-acetylmuramyl pentapeptide phosphotransferase/UDP-N-ace	5171	1	0	0	0	0	1	Strain-specific
1305	COG2801	[L] COG2801 Transposase and inactivated derivatives	5170	1	0	0	0	0	1	Strain-specific
1306			5168	1	0	0	0	0	1	Strain-specific
1307	COG1922	[M] COG1922 Teichoic acid biosynthesis proteins	5167	1	0	0	0	0	1	Strain-specific
1308	COG3309	[S] COG3309 Uncharacterized virulence-associated protein D	5166	1	0	0	0	0	1	Strain-specific
1309			5165	1	0	0	0	0	1	Strain-specific
1310	COG2207	[K] COG2207 AraC-type DNA-binding domain-containing proteins	5164	1	0	0	0	0	1	Strain-specific
1311	COG2333	[R] COG2333 Predicted hydrolase (metallo-beta-lactamase superfamily)	5163	1	0	0	0	0	1	Strain-specific
1312			5162	1	0	0	0	0	1	Strain-specific
1313			5161	1	0	0	0	0	1	Strain-specific
1314			5159	1	0	0	0	0	1	Strain-specific
1315	COG0610	[V] COG0610 Type I site-specific restriction-modification system, R (restriction) subu	5158	1	0	0	0	0	1	Strain-specific
1316			5157	1	0	0	0	0	1	Strain-specific
1317	COG0235	[G] COG0235 Ribulose-5-phosphate 4-epimerase and related epimerases and aldol	5156	1	0	0	0	0	1	Strain-specific
1318	COG3335	[L] COG3335 Transposase and inactivated derivatives	5155	1	0	0	0	0	1	Strain-specific
1319	COG0583	[K] COG0583 Transcriptional regulator	5154	1	0	0	0	0	1	Strain-specific
1320			5153	1	0	0	0	0	1	Strain-specific
1321			5152	1	0	0	0	0	1	Strain-specific
1322			5151	1	0	0	0	0	1	Strain-specific
1323	COG2161	[D] COG2161 Antitoxin of toxin-antitoxin stability system	5150	1	0	0	0	0	1	Strain-specific
1324			5149	1	0	0	0	0	1	Strain-specific
1325			5148	1	0	0	0	0	1	Strain-specific
1326	COG4694	[S] COG4694 Uncharacterized protein conserved in bacteria	5147	1	0	0	0	0	1	Strain-specific
1327	COG0596	[R] COG0596 Predicted hydrolases or acyltransferases (alpha/beta hydrolase superf	5146	1	0	0	0	0	1	Strain-specific
1328	COG1280	[E] COG1280 Putative threonine efflux protein	5145	1	0	0	0	0	1	Strain-specific
1329			5144	1	0	0	0	0	1	Strain-specific
1330	COG0610	[V] COG0610 Type I site-specific restriction-modification system, R (restriction) subu	5143	1	0	0	0	0	1	Strain-specific
1331	COG4688	[S] COG4688 Uncharacterized protein conserved in bacteria	5142	1	0	0	0	0	1	Strain-specific
1332			5141	1	0	0	0	0	1	Strain-specific
1333	COG1437	[F] COG1437 Adenylate cyclase, class 2 (thermophilic)	5140	1	0	0	0	0	1	Strain-specific
1334	COG0270	[L] COG0270 Site-specific DNA methylase	5139	1	0	0	0	0	1	Strain-specific
1335			5138	1	0	0	0	0	1	Strain-specific
1336	COG1114	[E] COG1114 Branched-chain amino acid permeases	5137	1	0	0	0	0	1	Strain-specific
1337			5136	1	0	0	0	0	1	Strain-specific
1338	COG0399	[M] COG0399 Predicted pyridoxal phosphate-dependent enzyme apparently involv	5032	1	0	0	0	0	1	Strain-specific
1339	COG0789	[K] COG0789 Predicted transcriptional regulators	5031	1	0	0	0	0	1	Strain-specific
1340			5030	1	0	0	0	0	1	Strain-specific
1341	COG0178	[L] COG0178 Excinuclease ATPase subunit	5029	1	0	0	0	0	1	Strain-specific
1342			5028	1	0	0	0	0	1	Strain-specific
1343	COG3658	[C] COG3658 Cytochrome b	5027	1	0	0	0	0	1	Strain-specific
1344			5025	1	0	0	0	0	1	Strain-specific
1345	COG1134	[GM] COG1134 ABC-type polysaccharide/polyol phosphate transport system, ATPa	5024	1	0	0	0	0	1	Strain-specific
1346			5023	1	0	0	0	0	1	Strain-specific
1347			5022	1	0	0	0	0	1	Strain-specific
1348	COG3344	[L] COG3344 Retron-type reverse transcriptase	5021	1	0	0	0	0	1	Strain-specific
1349	COG1175	[G] COG1175 ABC-type sugar transport systems, permease components	5020	1	0	0	0	0	1	Strain-specific
1350	COG1082	[G] COG1082 Sugar phosphate isomerases/epimerases	5019	1	0	0	0	0	1	Strain-specific
1351	COG2378	[K] COG2378 Predicted transcriptional regulator	5018	1	0	0	0	0	1	Strain-specific
1352	COG0021	[G] COG0021 Transketolase	5017	1	0	0	0	0	1	Strain-specific
1353	COG0840	[NT] COG0840 Methyl-accepting chemotaxis protein	5016	1	0	0	0	0	1	Strain-specific
1354			5013	1	0	0	0	0	1	Strain-specific
1355			5012	1	0	0	0	0	1	Strain-specific
1356	COG0424	[D] COG0424 Nucleotide-binding protein implicated in inhibition of septum format	5011	1	0	0	0	0	1	Strain-specific
1357			5010	1	0	0	0	0	1	Strain-specific
1358	COG4188	[R] COG4188 Predicted dienelactone hydrolase	5009	1	0	0	0	0	1	Strain-specific
1359			5008	1	0	0	0	0	1	Strain-specific
1360	COG0262	[H] COG0262 Dihydrofolate reductase	5007	1	0	0	0	0	1	Strain-specific
1361			5005	1	0	0	0	0	1	Strain-specific
1362			5003	1	0	0	0	0	1	Strain-specific
1363			5002	1	0	0	0	0	1	Strain-specific
1364	COG4146	[R] COG4146 Predicted symporter	5000	1	0	0	0	0	1	Strain-specific
1365			4999	1	0	0	0	0	1	Strain-specific
1366			4998	1	0	0	0	0	1	Strain-specific
1367	COG1479	[S] COG1479 Uncharacterized conserved protein	4997	1	0	0	0	0	1	Strain-specific
1368	COG1246	[E] COG1246 N-acetylglutamate synthase and related acetyltransferases	4996	1	0	0	0	0	1	Strain-specific
1369	COG1704	[S] COG1704 Uncharacterized conserved protein	4995	1	0	0	0	0	1	Strain-specific
1370	COG0612	[R] COG0612 Predicted Zn-dependent peptidases	4994	1	0	0	0	0	1	Strain-specific
1371	COG0207	[F] COG0207 Thymidylate synthase	4993	1	0	0	0	0	1	Strain-specific
1372	COG1403	[V] COG1403 Restriction endonuclease	4992	1	0	0	0	0	1	Strain-specific
1373			4991	1	0	0	0	0	1	Strain-specific
1374			4990	1	0	0	0	0	1	Strain-specific
1375			4989	1	0	0	0	0	1	Strain-specific
1376			4988	1	0	0	0	0	1	Strain-specific
1377			4987	1	0	0	0	0	1	Strain-specific
1378	COG0645	[R] COG0645 Predicted kinase	4986	1	0	0	0	0	1	Strain-specific
1379	COG3311	[K] COG3311 Predicted transcriptional regulator	4985	1	0	0	0	0	1	Strain-specific
1380			4984	1	0	0	0	0	1	Strain-specific
1381	COG2865	[K] COG2865 Predicted transcriptional regulator containing an HTH domain and an	4983	1	0	0	0	0	1	Strain-specific
1382	COG1205	[R] COG1205 Distinct helicase family with a unique C-terminal domain including a r	4982	1	0	0	0	0	1	Strain-specific
1383			4981	1	0	0	0	0	1	Strain-specific
1384			4980	1	0	0	0	0	1	Strain-specific
1385	COG0389	[L] COG0389 Nucleotidyltransferase/DNA polymerase involved in DNA repair	4978	1	0	0	0	0	1	Strain-specific
1386			4977	1	0	0	0	0	1	Strain-specific
1387			4976	1	0	0	0	0	1	Strain-specific
1388			4872	1	0	0	0	0	1	Strain-specific
1389			4868	1	0	0	0	0	1	Strain-specific
1390	COG0286	[V] COG0286 Type I restriction-modification system methyltransferase subunit	4867	1	0	0	0	0	1	Strain-specific
1391	COG1596	[M] COG1596 Periplasmic protein involved in polysaccharide export	4866	1	0	0	0	0	1	Strain-specific
1392	COG3655	[K] COG3655 Predicted transcriptional regulator	4865	1	0	0	0	0	1	Strain-specific
1393			4863	1	0	0	0	0	1	Strain-specific
1394			4862	1	0	0	0	0	1	Strain-specific
1395			4861	1	0	0	0	0	1	Strain-specific
1396	COG0582	[L] COG0582 Integrase	4860	1	0	0	0	0	1	Strain-specific
1397	COG4166	[E] COG4166 ABC-type oligopeptide transport system, periplasmic component	4859	1	0	0	0	0	1	Strain-specific
1398	COG3505	[U] COG3505 Type IV secretory pathway, VirD4 components	4858	1	0	0	0	0	1	Strain-specific
1399	COG2856	[E] COG2856 Predicted Zn peptidase	4857	1	0	0	0	0	1	Strain-specific
1400	COG0488	[R] COG0488 ATPase components of ABC transporters with duplicated ATPase dom	4856	1	0	0	0	0	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
1401	COG3385	[L] COG3385 FOG: Transposase and inactivated derivatives	4854	1	0	0	0	0	1	Strain-specific
1402	COG5635	[T] COG5635 Predicted NTPase (NACHT family)	4853	1	0	0	0	0	1	Strain-specific
1403	COG1061	[KL] COG1061 DNA or RNA helicases of superfamily II	4851	1	0	0	0	0	1	Strain-specific
1404			4849	1	0	0	0	0	1	Strain-specific
1405			4848	1	0	0	0	0	1	Strain-specific
1406			4847	1	0	0	0	0	1	Strain-specific
1407			4846	1	0	0	0	0	1	Strain-specific
1408			4845	1	0	0	0	0	1	Strain-specific
1409	COG1512	[R] COG1512 Beta-propeller domains of methanol dehydrogenase type	4844	1	0	0	0	0	1	Strain-specific
1410	COG4718	[S] COG4718 Phage-related protein	4843	1	0	0	0	0	1	Strain-specific
1411	COG3747	[L] COG3747 Phage terminase, small subunit	4842	1	0	0	0	0	1	Strain-specific
1412	COG1846	[K] COG1846 Transcriptional regulators	4841	1	0	0	0	0	1	Strain-specific
1413	COG2801	[L] COG2801 Transposase and inactivated derivatives	4839	1	0	0	0	0	1	Strain-specific
1414	COG3324	[R] COG3324 Predicted enzyme related to lactoylglutathione lyase	4838	1	0	0	0	0	1	Strain-specific
1415			4837	1	0	0	0	0	1	Strain-specific
1416			4836	1	0	0	0	0	1	Strain-specific
1417	COG0613	[R] COG0613 Predicted metal-dependent phosphoesterases (PHP family)	4834	1	0	0	0	0	1	Strain-specific
1418			4833	1	0	0	0	0	1	Strain-specific
1419			4831	1	0	0	0	0	1	Strain-specific
1420	COG1012	[C] COG1012 NAD-dependent aldehyde dehydrogenases	4830	1	0	0	0	0	1	Strain-specific
1421			4829	1	0	0	0	0	1	Strain-specific
1422			4714	1	0	0	0	0	1	Strain-specific
1423			4710	1	0	0	0	0	1	Strain-specific
1424	COG1132	[V] COG1132 ABC-type multidrug transport system, ATPase and permease compone	4709	1	0	0	0	0	1	Strain-specific
1425	COG3658	[C] COG3658 Cytochrome b	4708	1	0	0	0	0	1	Strain-specific
1426	COG3333	[S] COG3333 Uncharacterized protein conserved in bacteria	4706	1	0	0	0	0	1	Strain-specific
1427	COG3550	[R] COG3550 Uncharacterized protein related to capsule biosynthesis enzymes	4705	1	0	0	0	0	1	Strain-specific
1428			4703	1	0	0	0	0	1	Strain-specific
1429	COG0730	[R] COG0730 Predicted permeases	4701	1	0	0	0	0	1	Strain-specific
1430	COG0286	[V] COG0286 Type I restriction-modification system methyltransferase subunit	4700	1	0	0	0	0	1	Strain-specific
1431			4698	1	0	0	0	0	1	Strain-specific
1432	COG0438	[M] COG0438 Glycosyltransferase	4697	1	0	0	0	0	1	Strain-specific
1433	COG0625	[O] COG0625 Glutathione S-transferase	4696	1	0	0	0	0	1	Strain-specific
1434	COG1132	[V] COG1132 ABC-type multidrug transport system, ATPase and permease compone	4695	1	0	0	0	0	1	Strain-specific
1435	COG0491	[R] COG0491 Zn-dependent hydrolases, including glyoxylases	4694	1	0	0	0	0	1	Strain-specific
1436	COG3328	[L] COG3328 Transposase and inactivated derivatives	4693	1	0	0	0	0	1	Strain-specific
1437			4692	1	0	0	0	0	1	Strain-specific
1438	COG5340	[K] COG5340 Predicted transcriptional regulator	4691	1	0	0	0	0	1	Strain-specific
1439	COG4823	[V] COG4823 Abortive infection bacteriophage resistance protein	4690	1	0	0	0	0	1	Strain-specific
1440			4688	1	0	0	0	0	1	Strain-specific
1441			4686	1	0	0	0	0	1	Strain-specific
1442			4684	1	0	0	0	0	1	Strain-specific
1443	COG3037	[S] COG3037 Uncharacterized protein conserved in bacteria	4683	1	0	0	0	0	1	Strain-specific
1444	COG0845	[M] COG0845 Membrane-fusion protein	4682	1	0	0	0	0	1	Strain-specific
1445			4681	1	0	0	0	0	1	Strain-specific
1446			4680	1	0	0	0	0	1	Strain-specific
1447			4679	1	0	0	0	0	1	Strain-specific
1448			4678	1	0	0	0	0	1	Strain-specific
1449	COG0270	[L] COG0270 Site-specific DNA methylase	4677	1	0	0	0	0	1	Strain-specific
1450	COG1051	[F] COG1051 ADP-ribose pyrophosphatase	4676	1	0	0	0	0	1	Strain-specific
1451			4675	1	0	0	0	0	1	Strain-specific
1452			4674	1	0	0	0	0	1	Strain-specific
1453	COG4672	[S] COG4672 Phage-related protein	4673	1	0	0	0	0	1	Strain-specific
1454			4672	1	0	0	0	0	1	Strain-specific
1455	COG4891	[S] COG4891 Uncharacterized conserved protein	4671	1	0	0	0	0	1	Strain-specific
1456			4670	1	0	0	0	0	1	Strain-specific
1457			4669	1	0	0	0	0	1	Strain-specific
1458			4668	1	0	0	0	0	1	Strain-specific
1459			4666	1	0	0	0	0	1	Strain-specific
1460	COG0732	[V] COG0732 Restriction endonuclease S subunits	4665	1	0	0	0	0	1	Strain-specific
1461	COG0286	[V] COG0286 Type I restriction-modification system methyltransferase subunit	4664	1	0	0	0	0	1	Strain-specific
1462	COG4775	[M] COG4775 Outer membrane protein/protective antigen OMA87	4663	1	0	0	0	0	1	Strain-specific
1463			4662	1	0	0	0	0	1	Strain-specific
1464	COG0381	[M] COG0381 UDP-N-acetylglucosamine 2-epimerase	4661	1	0	0	0	0	1	Strain-specific
1465			4660	1	0	0	0	0	1	Strain-specific
1466	COG4096	[V] COG4096 Type I site-specific restriction-modification system, R (restriction) subu	4659	1	0	0	0	0	1	Strain-specific
1467	COG3344	[L] COG3344 Retron-type reverse transcriptase	4548	1	0	0	0	0	1	Strain-specific
1468			4545	1	0	0	0	0	1	Strain-specific
1469			4541	1	0	0	0	0	1	Strain-specific
1470	COG3754	[M] COG3754 Lipopolysaccharide biosynthesis protein	4540	1	0	0	0	0	1	Strain-specific
1471	COG3501	[S] COG3501 Uncharacterized protein conserved in bacteria	4539	1	0	0	0	0	1	Strain-specific
1472	COG5935		4538	1	0	0	0	0	1	Strain-specific
1473			4537	1	0	0	0	0	1	Strain-specific
1474	COG3501	[S] COG3501 Uncharacterized protein conserved in bacteria	4536	1	0	0	0	0	1	Strain-specific
1475	COG0463	[M] COG0463 Glycosyltransferases involved in cell wall biogenesis	4535	1	0	0	0	0	1	Strain-specific
1476	COG1309	[K] COG1309 Transcriptional regulator	4534	1	0	0	0	0	1	Strain-specific
1477			4533	1	0	0	0	0	1	Strain-specific
1478	COG0583	[K] COG0583 Transcriptional regulator	4532	1	0	0	0	0	1	Strain-specific
1479	COG3328	[L] COG3328 Transposase and inactivated derivatives	4531	1	0	0	0	0	1	Strain-specific
1480			4529	1	0	0	0	0	1	Strain-specific
1481			4527	1	0	0	0	0	1	Strain-specific
1482	COG4974	[L] COG4974 Site-specific recombinase XerD	4525	1	0	0	0	0	1	Strain-specific
1483			4524	1	0	0	0	0	1	Strain-specific
1484	COG4589	[R] COG4589 Predicted CDP-diglyceride synthetase/phosphatidate cytidylyltransfer	4522	1	0	0	0	0	1	Strain-specific
1485			4521	1	0	0	0	0	1	Strain-specific
1486	COG1129	[G] COG1129 ABC-type sugar transport system, ATPase component	4520	1	0	0	0	0	1	Strain-specific
1487	COG4694	[S] COG4694 Uncharacterized protein conserved in bacteria	4519	1	0	0	0	0	1	Strain-specific
1488	COG2801	[L] COG2801 Transposase and inactivated derivatives	4517	1	0	0	0	0	1	Strain-specific
1489	COG1670	[J] COG1670 Acetyltransferases, including N-acetylases of ribosomal proteins	4516	1	0	0	0	0	1	Strain-specific
1490			4515	1	0	0	0	0	1	Strain-specific
1491	COG4128	[R] COG4128 Zonula occludens toxin	4514	1	0	0	0	0	1	Strain-specific
1492	COG0175	[EH] COG0175 3'-phosphoadenosine 5'-phosphosulfate sulfotransferase (PAPS redu	4513	1	0	0	0	0	1	Strain-specific
1493			4512	1	0	0	0	0	1	Strain-specific
1494			4511	1	0	0	0	0	1	Strain-specific
1495	COG2801	[L] COG2801 Transposase and inactivated derivatives	4510	1	0	0	0	0	1	Strain-specific
1496	COG3153	[R] COG3153 Predicted acetyltransferase	4509	1	0	0	0	0	1	Strain-specific
1497	COG0286	[V] COG0286 Type I restriction-modification system methyltransferase subunit	4507	1	0	0	0	0	1	Strain-specific
1498			4505	1	0	0	0	0	1	Strain-specific
1499			4504	1	0	0	0	0	1	Strain-specific
1500			4503	1	0	0	0	0	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
1501			4501	1	0	0	0	0	1	Strain-specific
1502	COG3037	[S] COG3037 Uncharacterized protein conserved in bacteria	4378	1	0	0	0	0	1	Strain-specific
1503	COG4565	[KT] COG4565 Response regulator of citrate/malate metabolism	4377	1	0	0	0	0	1	Strain-specific
1504	COG1518	[L] COG1518 Uncharacterized protein predicted to be involved in DNA repair	4375	1	0	0	0	0	1	Strain-specific
1505	COG1541	[H] COG1541 Coenzyme F390 synthetase	4374	1	0	0	0	0	1	Strain-specific
1506			4373	1	0	0	0	0	1	Strain-specific
1507			4372	1	0	0	0	0	1	Strain-specific
1508			4371	1	0	0	0	0	1	Strain-specific
1509			4370	1	0	0	0	0	1	Strain-specific
1510	COG3645	[S] COG3645 Uncharacterized phage-encoded protein	4369	1	0	0	0	0	1	Strain-specific
1511	COG1506	[E] COG1506 Dipeptidyl aminopeptidases/acylaminoacyl-peptidases	4368	1	0	0	0	0	1	Strain-specific
1512	COG1879	[G] COG1879 ABC-type sugar transport system, periplasmic component	4367	1	0	0	0	0	1	Strain-specific
1513	COG3501	[S] COG3501 Uncharacterized protein conserved in bacteria	4366	1	0	0	0	0	1	Strain-specific
1514			4365	1	0	0	0	0	1	Strain-specific
1515	COG3833	[G] COG3833 ABC-type maltose transport systems, permease component	4364	1	0	0	0	0	1	Strain-specific
1516	COG0419	[L] COG0419 ATPase involved in DNA repair	4363	1	0	0	0	0	1	Strain-specific
1517	COG4771	[P] COG4771 Outer membrane receptor for ferrienterochelin and colicins	4361	1	0	0	0	0	1	Strain-specific
1518	COG1172	[G] COG1172 Ribose/xylose/arabinose/galactoside ABC-type transport systems, per	4360	1	0	0	0	0	1	Strain-specific
1519	COG3593	[L] COG3593 Predicted ATP-dependent endonuclease of the OLD family	4359	1	0	0	0	0	1	Strain-specific
1520			4358	1	0	0	0	0	1	Strain-specific
1521	COG3774	[M] COG3774 Mannosyltransferase OCH1 and related enzymes	4357	1	0	0	0	0	1	Strain-specific
1522			4356	1	0	0	0	0	1	Strain-specific
1523			4355	1	0	0	0	0	1	Strain-specific
1524			4354	1	0	0	0	0	1	Strain-specific
1525	COG3765	[M] COG3765 Chain length determinant protein	4353	1	0	0	0	0	1	Strain-specific
1526			4352	1	0	0	0	0	1	Strain-specific
1527			4351	1	0	0	0	0	1	Strain-specific
1528			4350	1	0	0	0	0	1	Strain-specific
1529			4349	1	0	0	0	0	1	Strain-specific
1530	COG1216	[R] COG1216 Predicted glycosyltransferases	4348	1	0	0	0	0	1	Strain-specific
1531			4347	1	0	0	0	0	1	Strain-specific
1532			4346	1	0	0	0	0	1	Strain-specific
1533			4344	1	0	0	0	0	1	Strain-specific
1534			4343	1	0	0	0	0	1	Strain-specific
1535	COG4771	[P] COG4771 Outer membrane receptor for ferrienterochelin and colicins	4342	1	0	0	0	0	1	Strain-specific
1536	COG1061	[KL] COG1061 DNA or RNA helicases of superfamily II	4341	1	0	0	0	0	1	Strain-specific
1537			4338	1	0	0	0	0	1	Strain-specific
1538	COG1061	[KL] COG1061 DNA or RNA helicases of superfamily II	4337	1	0	0	0	0	1	Strain-specific
1539	COG1396	[K] COG1396 Predicted transcriptional regulators	4336	1	0	0	0	0	1	Strain-specific
1540			4335	1	0	0	0	0	1	Strain-specific
1541			4334	1	0	0	0	0	1	Strain-specific
1542			4276	1	0	0	0	0	1	Strain-specific
1543	COG5932		4275	1	0	0	0	0	1	Strain-specific
1544			4274	1	0	0	0	0	1	Strain-specific
1545	COG0825	[I] COG0825 Acetyl-CoA carboxylase alpha subunit	4273	1	0	0	0	0	1	Strain-specific
1546	COG0399	[M] COG0399 Predicted pyridoxal phosphate-dependent enzyme apparently involv	4272	1	0	0	0	0	1	Strain-specific
1547			4271	1	0	0	0	0	1	Strain-specific
1548	COG0534	[V] COG0534 Na+-driven multidrug efflux pump	4270	1	0	0	0	0	1	Strain-specific
1549			4269	1	0	0	0	0	1	Strain-specific
1550			4268	1	0	0	0	0	1	Strain-specific
1551			4267	1	0	0	0	0	1	Strain-specific
1552	COG0583	[K] COG0583 Transcriptional regulator	4266	1	0	0	0	0	1	Strain-specific
1553	COG1913	[R] COG1913 Predicted Zn-dependent proteases	4265	1	0	0	0	0	1	Strain-specific
1554			4263	1	0	0	0	0	1	Strain-specific
1555	COG0210	[L] COG0210 Superfamily I DNA and RNA helicases	4261	1	0	0	0	0	1	Strain-specific
1556			4260	1	0	0	0	0	1	Strain-specific
1557	COG1519	[M] COG1519 3-deoxy-D-manno-octulosonic-acid transferase	4258	1	0	0	0	0	1	Strain-specific
1558			4257	1	0	0	0	0	1	Strain-specific
1559	COG1263	[G] COG1263 Phosphotransferase system IIC components, glucose/maltose/N-acety	4256	1	0	0	0	0	1	Strain-specific
1560			4255	1	0	0	0	0	1	Strain-specific
1561	COG2226	[H] COG2226 Methylase involved in ubiquinone/menaquinone biosynthesis	4254	1	0	0	0	0	1	Strain-specific
1562	COG1670	[J] COG1670 Acetyltransferases, including N-acetylases of ribosomal proteins	4253	1	0	0	0	0	1	Strain-specific
1563			4252	1	0	0	0	0	1	Strain-specific
1564	COG4974	[L] COG4974 Site-specific recombinase XerD	4251	1	0	0	0	0	1	Strain-specific
1565			4250	1	0	0	0	0	1	Strain-specific
1566	COG4974	[L] COG4974 Site-specific recombinase XerD	4249	1	0	0	0	0	1	Strain-specific
1567	COG2801	[L] COG2801 Transposase and inactivated derivatives	4248	1	0	0	0	0	1	Strain-specific
1568			4247	1	0	0	0	0	1	Strain-specific
1569	COG3316	[L] COG3316 Transposase and inactivated derivatives	4246	1	0	0	0	0	1	Strain-specific
1570	COG0323	[L] COG0323 DNA mismatch repair enzyme (predicted ATPase)	4245	1	0	0	0	0	1	Strain-specific
1571			4243	1	0	0	0	0	1	Strain-specific
1572	COG4694	[S] COG4694 Uncharacterized protein conserved in bacteria	4242	1	0	0	0	0	1	Strain-specific
1573			4241	1	0	0	0	0	1	Strain-specific
1574			4240	1	0	0	0	0	1	Strain-specific
1575			4239	1	0	0	0	0	1	Strain-specific
1576	COG1396	[K] COG1396 Predicted transcriptional regulators	4238	1	0	0	0	0	1	Strain-specific
1577	COG4175	[E] COG4175 ABC-type proline/glycine betaine transport system, ATPase componer	4237	1	0	0	0	0	1	Strain-specific
1578			4236	1	0	0	0	0	1	Strain-specific
1579			4235	1	0	0	0	0	1	Strain-specific
1580			4234	1	0	0	0	0	1	Strain-specific
1581	COG3597	[S] COG3597 Uncharacterized protein/domain associated with GTPases	4232	1	0	0	0	0	1	Strain-specific
1582			4231	1	0	0	0	0	1	Strain-specific
1583	COG3385	[L] COG3385 FOG: Transposase and inactivated derivatives	4230	1	0	0	0	0	1	Strain-specific
1584			4165	1	0	0	0	0	1	Strain-specific
1585			4164	1	0	0	0	0	1	Strain-specific
1586	COG5943		4163	1	0	0	0	0	1	Strain-specific
1587	COG4421	[G] COG4421 Capsular polysaccharide biosynthesis protein	4162	1	0	0	0	0	1	Strain-specific
1588	COG5933		4161	1	0	0	0	0	1	Strain-specific
1589	COG0631	[T] COG0631 Serine/threonine protein phosphatase	4160	1	0	0	0	0	1	Strain-specific
1590			4159	1	0	0	0	0	1	Strain-specific
1591			4158	1	0	0	0	0	1	Strain-specific
1592			4157	1	0	0	0	0	1	Strain-specific
1593			4156	1	0	0	0	0	1	Strain-specific
1594	COG2197	[TK] COG2197 Response regulator containing a CheY-like receiver domain and an H	4155	1	0	0	0	0	1	Strain-specific
1595			4154	1	0	0	0	0	1	Strain-specific
1596	COG0318	[IQ] COG0318 Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II	4152	1	0	0	0	0	1	Strain-specific
1597			4150	1	0	0	0	0	1	Strain-specific
1598	COG1104	[E] COG1104 Cysteine sulfinate desulfinase/cysteine desulfurase and related enzym	4148	1	0	0	0	0	1	Strain-specific
1599			4147	1	0	0	0	0	1	Strain-specific
1600	COG0110	[R] COG0110 Acetyltransferase (isoleucine patch superfamily)	4146	1	0	0	0	0	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
1601			4145	1	0	0	0	0	1	Strain-specific
1602	COG0286	[V] COG0286 Type I restriction-modification system methyltransferase subunit	4144	1	0	0	0	0	1	Strain-specific
1603	COG2333	[R] COG2333 Predicted hydrolase (metallo-beta-lactamase superfamily)	4143	1	0	0	0	0	1	Strain-specific
1604	COG3547	[L] COG3547 Transposase and inactivated derivatives	4142	1	0	0	0	0	1	Strain-specific
1605			4141	1	0	0	0	0	1	Strain-specific
1606			4139	1	0	0	0	0	1	Strain-specific
1607	COG0673	[R] COG0673 Predicted dehydrogenases and related proteins	4135	1	0	0	0	0	1	Strain-specific
1608	COG0674	[C] COG0674 Pyruvate:ferredoxin oxidoreductase and related 2-oxoacid:ferredoxin	4134	1	0	0	0	0	1	Strain-specific
1609			4133	1	0	0	0	0	1	Strain-specific
1610	COG0641	[R] COG0641 Arylsulfatase regulator (Fe-S oxidoreductase)	4132	1	0	0	0	0	1	Strain-specific
1611	COG0210	[L] COG0210 Superfamily I DNA and RNA helicases	4131	1	0	0	0	0	1	Strain-specific
1612	COG0021	[G] COG0021 Transketolase	4130	1	0	0	0	0	1	Strain-specific
1613			4129	1	0	0	0	0	1	Strain-specific
1614			4128	1	0	0	0	0	1	Strain-specific
1615			4127	1	0	0	0	0	1	Strain-specific
1616			4126	1	0	0	0	0	1	Strain-specific
1617	COG3436	[L] COG3436 Transposase and inactivated derivatives	4125	1	0	0	0	0	1	Strain-specific
1618			4124	1	0	0	0	0	1	Strain-specific
1619			4123	1	0	0	0	0	1	Strain-specific
1620			4122	1	0	0	0	0	1	Strain-specific
1621			4050	1	0	0	0	0	1	Strain-specific
1622	COG5933		4049	1	0	0	0	0	1	Strain-specific
1623			4048	1	0	0	0	0	1	Strain-specific
1624	COG3593	[L] COG3593 Predicted ATP-dependent endonuclease of the OLD family	4047	1	0	0	0	0	1	Strain-specific
1625			4046	1	0	0	0	0	1	Strain-specific
1626			4045	1	0	0	0	0	1	Strain-specific
1627	COG2814	[G] COG2814 Arabinose efflux permease	4044	1	0	0	0	0	1	Strain-specific
1628	COG1192	[D] COG1192 ATPases involved in chromosome partitioning	4043	1	0	0	0	0	1	Strain-specific
1629			4042	1	0	0	0	0	1	Strain-specific
1630			4038	1	0	0	0	0	1	Strain-specific
1631	COG0470	[L] COG0470 ATPase involved in DNA replication	4036	1	0	0	0	0	1	Strain-specific
1632			4035	1	0	0	0	0	1	Strain-specific
1633	COG1263	[G] COG1263 Phosphotransferase system IIC components, glucose/maltose/N-acety	4034	1	0	0	0	0	1	Strain-specific
1634	COG3505	[U] COG3505 Type IV secretory pathway, VirD4 components	4033	1	0	0	0	0	1	Strain-specific
1635			4032	1	0	0	0	0	1	Strain-specific
1636	COG0841	[V] COG0841 Cation/multidrug efflux pump	4030	1	0	0	0	0	1	Strain-specific
1637	COG4469	[R] COG4469 Competence protein	4029	1	0	0	0	0	1	Strain-specific
1638			4028	1	0	0	0	0	1	Strain-specific
1639	COG4588	[R] COG4588 Accessory colonization factor AcfC, contains ABC-type periplasmic dor	4027	1	0	0	0	0	1	Strain-specific
1640	COG0507	[L] COG0507 ATP-dependent exoDNAse (exonuclease V), alpha subunit - helicase su	4026	1	0	0	0	0	1	Strain-specific
1641			4025	1	0	0	0	0	1	Strain-specific
1642	COG0438	[M] COG0438 Glycosyltransferase	4024	1	0	0	0	0	1	Strain-specific
1643			4023	1	0	0	0	0	1	Strain-specific
1644	COG0840	[NT] COG0840 Methyl-accepting chemotaxis protein	4022	1	0	0	0	0	1	Strain-specific
1645	COG2110	[R] COG2110 Predicted phosphatase homologous to the C-terminal domain of histo	4021	1	0	0	0	0	1	Strain-specific
1646			4020	1	0	0	0	0	1	Strain-specific
1647	COG0346	[E] COG0346 Lactoylglutathione lyase and related lyases	4019	1	0	0	0	0	1	Strain-specific
1648			4018	1	0	0	0	0	1	Strain-specific
1649			4016	1	0	0	0	0	1	Strain-specific
1650			4015	1	0	0	0	0	1	Strain-specific
1651			4014	1	0	0	0	0	1	Strain-specific
1652	COG0384	[R] COG0384 Predicted epimerase, PhzC/PhzF homolog	4013	1	0	0	0	0	1	Strain-specific
1653	COG4969	[NU] COG4969 Tfp pilus assembly protein, major pilin PilA	4012	1	0	0	0	0	1	Strain-specific
1654	COG0207	[F] COG0207 Thymidylate synthase	4011	1	0	0	0	0	1	Strain-specific
1655	COG2244	[R] COG2244 Membrane protein involved in the export of O-antigen and teichoic ac	4010	1	0	0	0	0	1	Strain-specific
1656	COG5293	[S] COG5293 Uncharacterized protein conserved in bacteria	3946	1	0	0	0	0	1	Strain-specific
1657	COG3550	[R] COG3550 Uncharacterized protein related to capsule biosynthesis enzymes	3945	1	0	0	0	0	1	Strain-specific
1658	COG0705	[R] COG0705 Uncharacterized membrane protein (homolog of Drosophila rhomboi	3944	1	0	0	0	0	1	Strain-specific
1659			3943	1	0	0	0	0	1	Strain-specific
1660	COG5936		3942	1	0	0	0	0	1	Strain-specific
1661			3941	1	0	0	0	0	1	Strain-specific
1662	COG4640	[S] COG4640 Predicted membrane protein	3940	1	0	0	0	0	1	Strain-specific
1663			3939	1	0	0	0	0	1	Strain-specific
1664			3938	1	0	0	0	0	1	Strain-specific
1665			3936	1	0	0	0	0	1	Strain-specific
1666			3935	1	0	0	0	0	1	Strain-specific
1667	COG0456	[R] COG0456 Acetyltransferases	3934	1	0	0	0	0	1	Strain-specific
1668			3932	1	0	0	0	0	1	Strain-specific
1669	COG2182	[G] COG2182 Maltose-binding periplasmic proteins/domains	3931	1	0	0	0	0	1	Strain-specific
1670			3930	1	0	0	0	0	1	Strain-specific
1671	COG2067	[I] COG2067 Long-chain fatty acid transport protein	3927	1	0	0	0	0	1	Strain-specific
1672	COG3039	[L] COG3039 Transposase and inactivated derivatives, IS5 family	3926	1	0	0	0	0	1	Strain-specific
1673			3924	1	0	0	0	0	1	Strain-specific
1674	COG3179	[R] COG3179 Predicted chitinase	3923	1	0	0	0	0	1	Strain-specific
1675	COG3501	[S] COG3501 Uncharacterized protein conserved in bacteria	3922	1	0	0	0	0	1	Strain-specific
1676	COG1204	[R] COG1204 Superfamily II helicase	3921	1	0	0	0	0	1	Strain-specific
1677			3920	1	0	0	0	0	1	Strain-specific
1678			3919	1	0	0	0	0	1	Strain-specific
1679			3918	1	0	0	0	0	1	Strain-specific
1680			3917	1	0	0	0	0	1	Strain-specific
1681			3916	1	0	0	0	0	1	Strain-specific
1682			3915	1	0	0	0	0	1	Strain-specific
1683	COG3673	[S] COG3673 Uncharacterized conserved protein	3914	1	0	0	0	0	1	Strain-specific
1684			3913	1	0	0	0	0	1	Strain-specific
1685	COG4823	[V] COG4823 Abortive infection bacteriophage resistance protein	3912	1	0	0	0	0	1	Strain-specific
1686	COG1961	[L] COG1961 Site-specific recombinases, DNA invertase Pin homologs	3911	1	0	0	0	0	1	Strain-specific
1687			3910	1	0	0	0	0	1	Strain-specific
1688	COG3668	[R] COG3668 Plasmid stabilization system protein	3909	1	0	0	0	0	1	Strain-specific
1689			3908	1	0	0	0	0	1	Strain-specific
1690			3907	1	0	0	0	0	1	Strain-specific
1691			3906	1	0	0	0	0	1	Strain-specific
1692			3905	1	0	0	0	0	1	Strain-specific
1693	COG2070	[R] COG2070 Dioxygenases related to 2-nitropropane dioxygenase	3904	1	0	0	0	0	1	Strain-specific
1694			3903	1	0	0	0	0	1	Strain-specific
1695			3902	1	0	0	0	0	1	Strain-specific
1696	COG0456	[R] COG0456 Acetyltransferases	3901	1	0	0	0	0	1	Strain-specific
1697			3900	1	0	0	0	0	1	Strain-specific
1698	COG3774	[M] COG3774 Mannosyltransferase OCH1 and related enzymes	3899	1	0	0	0	0	1	Strain-specific
1699			3898	1	0	0	0	0	1	Strain-specific
1700	COG4096	[V] COG4096 Type I site-specific restriction-modification system, R (restriction) subu	3897	1	0	0	0	0	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
1701	COG1670	[J] COG1670 Acetyltransferases, including N-acetylases of ribosomal proteins	3896	1	0	0	0	0	1	Strain-specific
1702			3895	1	0	0	0	0	1	Strain-specific
1703	COG0463	[M] COG0463 Glycosyltransferases involved in cell wall biogenesis	3894	1	0	0	0	0	1	Strain-specific
1704	COG3550	[R] COG3550 Uncharacterized protein related to capsule biosynthesis enzymes	3824	1	0	0	0	0	1	Strain-specific
1705			3823	1	0	0	0	0	1	Strain-specific
1706	COG0642	[T] COG0642 Signal transduction histidine kinase	3822	1	0	0	0	0	1	Strain-specific
1707	COG5935		3820	1	0	0	0	0	1	Strain-specific
1708			3819	1	0	0	0	0	1	Strain-specific
1709			3818	1	0	0	0	0	1	Strain-specific
1710	COG2274	[V] COG2274 ABC-type bacteriocin/lantibiotic exporters, contain an N-terminal dou	3816	1	0	0	0	0	1	Strain-specific
1711			3815	1	0	0	0	0	1	Strain-specific
1712	COG5001	[T] COG5001 Predicted signal transduction protein containing a membrane domain	3814	1	0	0	0	0	1	Strain-specific
1713	COG1328	[F] COG1328 Oxygen-sensitive ribonucleoside-triphosphate reductase	3813	1	0	0	0	0	1	Strain-specific
1714			3812	1	0	0	0	0	1	Strain-specific
1715			3811	1	0	0	0	0	1	Strain-specific
1716	COG1670	[J] COG1670 Acetyltransferases, including N-acetylases of ribosomal proteins	3810	1	0	0	0	0	1	Strain-specific
1717	COG4405	[S] COG4405 Uncharacterized protein conserved in bacteria	3809	1	0	0	0	0	1	Strain-specific
1718	COG0399	[M] COG0399 Predicted pyridoxal phosphate-dependent enzyme apparently involv	3808	1	0	0	0	0	1	Strain-specific
1719	COG1902	[C] COG1902 NADH:flavin oxidoreductases, Old Yellow Enzyme family	3807	1	0	0	0	0	1	Strain-specific
1720	COG2182	[G] COG2182 Maltose-binding periplasmic proteins/domains	3806	1	0	0	0	0	1	Strain-specific
1721			3805	1	0	0	0	0	1	Strain-specific
1722			3804	1	0	0	0	0	1	Strain-specific
1723			3799	1	0	0	0	0	1	Strain-specific
1724			3798	1	0	0	0	0	1	Strain-specific
1725			3797	1	0	0	0	0	1	Strain-specific
1726	COG2199	[T] COG2199 FOG: GGDEF domain	3795	1	0	0	0	0	1	Strain-specific
1727			3794	1	0	0	0	0	1	Strain-specific
1728			3792	1	0	0	0	0	1	Strain-specific
1729	COG1061	[KL] COG1061 DNA or RNA helicases of superfamily II	3791	1	0	0	0	0	1	Strain-specific
1730			3790	1	0	0	0	0	1	Strain-specific
1731			3788	1	0	0	0	0	1	Strain-specific
1732	COG2274	[V] COG2274 ABC-type bacteriocin/lantibiotic exporters, contain an N-terminal dou	3787	1	0	0	0	0	1	Strain-specific
1733	COG0609	[P] COG0609 ABC-type Fe3+-siderophore transport system, permease component	3786	1	0	0	0	0	1	Strain-specific
1734	COG1898	[M] COG1898 dTDP-4-dehydrorhamnose 3,5-epimerase and related enzymes	3785	1	0	0	0	0	1	Strain-specific
1735	COG0110	[R] COG0110 Acetyltransferase (isoleucine patch superfamily)	3784	1	0	0	0	0	1	Strain-specific
1736			3783	1	0	0	0	0	1	Strain-specific
1737			3782	1	0	0	0	0	1	Strain-specific
1738			3781	1	0	0	0	0	1	Strain-specific
1739			3780	1	0	0	0	0	1	Strain-specific
1740	COG1292	[M] COG1292 Choline-glycine betaine transporter	3779	1	0	0	0	0	1	Strain-specific
1741	COG1199	[KL] COG1199 Rad3-related DNA helicases	3778	1	0	0	0	0	1	Strain-specific
1742	COG3501	[S] COG3501 Uncharacterized protein conserved in bacteria	3777	1	0	0	0	0	1	Strain-specific
1743	COG3677	[L] COG3677 Transposase and inactivated derivatives	3776	1	0	0	0	0	1	Strain-specific
1744	COG3950	[R] COG3950 Predicted ATP-binding protein involved in virulence	3775	1	0	0	0	0	1	Strain-specific
1745	COG3333	[S] COG3333 Uncharacterized protein conserved in bacteria	3774	1	0	0	0	0	1	Strain-specific
1746			3773	1	0	0	0	0	1	Strain-specific
1747	COG3829	[KT] COG3829 Transcriptional regulator containing PAS, AAA-type ATPase, and DNA	3772	1	0	0	0	0	1	Strain-specific
1748	COG1474	[LO] COG1474 Cdc6-related protein, AAA superfamily ATPase	3771	1	0	0	0	0	1	Strain-specific
1749			3770	1	0	0	0	0	1	Strain-specific
1750	COG0438	[M] COG0438 Glycosyltransferase	3769	1	0	0	0	0	1	Strain-specific
1751	COG0790	[R] COG0790 FOG: TPR repeat, SEL1 subfamily	3701	1	0	0	0	0	1	Strain-specific
1752	COG2199	[T] COG2199 FOG: GGDEF domain	3700	1	0	0	0	0	1	Strain-specific
1753	COG4974	[L] COG4974 Site-specific recombinase XerD	3699	1	0	0	0	0	1	Strain-specific
1754	COG4653	[R] COG4653 Predicted phage phi-C31 gp36 major capsid-like protein	3696	1	0	0	0	0	1	Strain-specific
1755	COG2801	[L] COG2801 Transposase and inactivated derivatives	3695	1	0	0	0	0	1	Strain-specific
1756			3694	1	0	0	0	0	1	Strain-specific
1757			3693	1	0	0	0	0	1	Strain-specific
1758			3692	1	0	0	0	0	1	Strain-specific
1759			3691	1	0	0	0	0	1	Strain-specific
1760	COG3344	[L] COG3344 Retron-type reverse transcriptase	3690	1	0	0	0	0	1	Strain-specific
1761	COG4733	[S] COG4733 Phage-related protein, tail component	3689	1	0	0	0	0	1	Strain-specific
1762			3688	1	0	0	0	0	1	Strain-specific
1763			3687	1	0	0	0	0	1	Strain-specific
1764			3686	1	0	0	0	0	1	Strain-specific
1765			3685	1	0	0	0	0	1	Strain-specific
1766			3684	1	0	0	0	0	1	Strain-specific
1767			3681	1	0	0	0	0	1	Strain-specific
1768	COG0732	[V] COG0732 Restriction endonuclease S subunits	3680	1	0	0	0	0	1	Strain-specific
1769	COG1530	[J] COG1530 Ribonucleases G and E	3678	1	0	0	0	0	1	Strain-specific
1770			3677	1	0	0	0	0	1	Strain-specific
1771			3674	1	0	0	0	0	1	Strain-specific
1772	COG5499	[K] COG5499 Predicted transcription regulator containing HTH domain	3673	1	0	0	0	0	1	Strain-specific
1773	COG0178	[L] COG0178 Excinuclease ATPase subunit	3672	1	0	0	0	0	1	Strain-specific
1774			3671	1	0	0	0	0	1	Strain-specific
1775			3670	1	0	0	0	0	1	Strain-specific
1776			3669	1	0	0	0	0	1	Strain-specific
1777			3668	1	0	0	0	0	1	Strain-specific
1778	COG0609	[P] COG0609 ABC-type Fe3+-siderophore transport system, permease component	3667	1	0	0	0	0	1	Strain-specific
1779	COG3607	[R] COG3607 Predicted lactoylglutathione lyase	3666	1	0	0	0	0	1	Strain-specific
1780	COG0399	[M] COG0399 Predicted pyridoxal phosphate-dependent enzyme apparently involv	3665	1	0	0	0	0	1	Strain-specific
1781	COG1082	[G] COG1082 Sugar phosphate isomerases/epimerases	3664	1	0	0	0	0	1	Strain-specific
1782	COG0834	[ET] COG0834 ABC-type amino acid transport/signal transduction systems, periplas	3663	1	0	0	0	0	1	Strain-specific
1783			3662	1	0	0	0	0	1	Strain-specific
1784	COG1378	[K] COG1378 Predicted transcriptional regulators	3661	1	0	0	0	0	1	Strain-specific
1785	COG1479	[S] COG1479 Uncharacterized conserved protein	3660	1	0	0	0	0	1	Strain-specific
1786			3659	1	0	0	0	0	1	Strain-specific
1787			3658	1	0	0	0	0	1	Strain-specific
1788			3657	1	0	0	0	0	1	Strain-specific
1789	COG1403	[V] COG1403 Restriction endonuclease	3656	1	0	0	0	0	1	Strain-specific
1790			3655	1	0	0	0	0	1	Strain-specific
1791			3654	1	0	0	0	0	1	Strain-specific
1792	COG1012	[C] COG1012 NAD-dependent aldehyde dehydrogenases	3653	1	0	0	0	0	1	Strain-specific
1793			3652	1	0	0	0	0	1	Strain-specific
1794	COG4774	[P] COG4774 Outer membrane receptor for monomeric catechols	3650	1	0	0	0	0	1	Strain-specific
1795	COG1045	[E] COG1045 Serine acetyltransferase	3649	1	0	0	0	0	1	Strain-specific
1796	COG5016	[C] COG5016 Pyruvate/oxaloacetate carboxyltransferase	3578	1	0	0	0	0	1	Strain-specific
1797	COG2200	[T] COG2200 FOG: EAL domain	3577	1	0	0	0	0	1	Strain-specific
1798			3576	1	0	0	0	0	1	Strain-specific
1799	COG4128	[R] COG4128 Zonula occludens toxin	3573	1	0	0	0	0	1	Strain-specific
1800			3572	1	0	0	0	0	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
1801	COG0572	[F] COG0572 Uridine kinase	3571	1	0	0	0	0	1	Strain-specific
1802	COG2303	[E] COG2303 Choline dehydrogenase and related flavoproteins	3569	1	0	0	0	0	1	Strain-specific
1803	COG4733	[S] COG4733 Phage-related protein, tail component	3568	1	0	0	0	0	1	Strain-specific
1804	COG0550	[L] COG0550 Topoisomerase IA	3567	1	0	0	0	0	1	Strain-specific
1805			3566	1	0	0	0	0	1	Strain-specific
1806			3565	1	0	0	0	0	1	Strain-specific
1807	COG1596	[M] COG1596 Periplasmic protein involved in polysaccharide export	3563	1	0	0	0	0	1	Strain-specific
1808	COG0207	[F] COG0207 Thymidylate synthase	3562	1	0	0	0	0	1	Strain-specific
1809			3559	1	0	0	0	0	1	Strain-specific
1810	COG4667	[R] COG4667 Predicted esterase of the alpha-beta hydrolase superfamily	3558	1	0	0	0	0	1	Strain-specific
1811	COG0424	[D] COG0424 Nucleotide-binding protein implicated in inhibition of septum format	3555	1	0	0	0	0	1	Strain-specific
1812			3554	1	0	0	0	0	1	Strain-specific
1813	COG4783	[R] COG4783 Putative Zn-dependent protease, contains TPR repeats	3552	1	0	0	0	0	1	Strain-specific
1814	COG1419	[N] COG1419 Flagellar GTP-binding protein	3551	1	0	0	0	0	1	Strain-specific
1815			3550	1	0	0	0	0	1	Strain-specific
1816			3549	1	0	0	0	0	1	Strain-specific
1817			3548	1	0	0	0	0	1	Strain-specific
1818			3547	1	0	0	0	0	1	Strain-specific
1819	COG3081	[R] COG3081 Nucleoid-associated protein	3546	1	0	0	0	0	1	Strain-specific
1820			3545	1	0	0	0	0	1	Strain-specific
1821			3544	1	0	0	0	0	1	Strain-specific
1822			3543	1	0	0	0	0	1	Strain-specific
1823			3542	1	0	0	0	0	1	Strain-specific
1824	COG2842	[R] COG2842 Uncharacterized ATPase, putative transposase	3541	1	0	0	0	0	1	Strain-specific
1825			3540	1	0	0	0	0	1	Strain-specific
1826			3538	1	0	0	0	0	1	Strain-specific
1827			3537	1	0	0	0	0	1	Strain-specific
1828			3536	1	0	0	0	0	1	Strain-specific
1829	COG5002	[T] COG5002 Signal transduction histidine kinase	3535	1	0	0	0	0	1	Strain-specific
1830	COG3030	[R] COG3030 Protein affecting phage T7 exclusion by the F plasmid	3534	1	0	0	0	0	1	Strain-specific
1831	COG1061	[KL] COG1061 DNA or RNA helicases of superfamily II	3533	1	0	0	0	0	1	Strain-specific
1832	COG0732	[V] COG0732 Restriction endonuclease S subunits	3532	1	0	0	0	0	1	Strain-specific
1833	COG1114	[E] COG1114 Branched-chain amino acid permeases	3531	1	0	0	0	0	1	Strain-specific
1834	COG4104	[S] COG4104 Uncharacterized conserved protein	3530	1	0	0	0	0	1	Strain-specific
1835	COG0834	[ET] COG0834 ABC-type amino acid transport/signal transduction systems, periplas	3529	1	0	0	0	0	1	Strain-specific
1836	COG4191	[T] COG4191 Signal transduction histidine kinase regulating C4-dicarboxylate trans	3528	1	0	0	0	0	1	Strain-specific
1837	COG0438	[M] COG0438 Glycosyltransferase	3527	1	0	0	0	0	1	Strain-specific
1838	COG0399	[M] COG0399 Predicted pyridoxal phosphate-dependent enzyme apparently involv	3448	1	0	0	0	0	1	Strain-specific
1839	COG3926	[R] COG3926 Putative secretion activating protein	3447	1	0	0	0	0	1	Strain-specific
1840	COG5492	[N] COG5492 Bacterial surface proteins containing Ig-like domains	3446	1	0	0	0	0	1	Strain-specific
1841			3445	1	0	0	0	0	1	Strain-specific
1842			3444	1	0	0	0	0	1	Strain-specific
1843			3442	1	0	0	0	0	1	Strain-specific
1844			3441	1	0	0	0	0	1	Strain-specific
1845			3440	1	0	0	0	0	1	Strain-specific
1846			3439	1	0	0	0	0	1	Strain-specific
1847	COG3415	[L] COG3415 Transposase and inactivated derivatives	3438	1	0	0	0	0	1	Strain-specific
1848			3437	1	0	0	0	0	1	Strain-specific
1849	COG3609	[K] COG3609 Predicted transcriptional regulators containing the CopG/Arc/MetJ DN	3436	1	0	0	0	0	1	Strain-specific
1850			3435	1	0	0	0	0	1	Strain-specific
1851			3434	1	0	0	0	0	1	Strain-specific
1852			3433	1	0	0	0	0	1	Strain-specific
1853	COG3206	[M] COG3206 Uncharacterized protein involved in exopolysaccharide biosynthesis	3431	1	0	0	0	0	1	Strain-specific
1854	COG2842	[R] COG2842 Uncharacterized ATPase, putative transposase	3430	1	0	0	0	0	1	Strain-specific
1855	COG0534	[V] COG0534 Na+-driven multidrug efflux pump	3429	1	0	0	0	0	1	Strain-specific
1856	COG5001	[T] COG5001 Predicted signal transduction protein containing a membrane domain	3426	1	0	0	0	0	1	Strain-specific
1857			3425	1	0	0	0	0	1	Strain-specific
1858	COG1310	[R] COG1310 Predicted metal-dependent protease of the PAD1/JAB1 superfamily	3424	1	0	0	0	0	1	Strain-specific
1859	COG1653	[G] COG1653 ABC-type sugar transport system, periplasmic component	3423	1	0	0	0	0	1	Strain-specific
1860	COG4694	[S] COG4694 Uncharacterized protein conserved in bacteria	3422	1	0	0	0	0	1	Strain-specific
1861	COG2165	[NU] COG2165 Type II secretory pathway, pseudopilin PulG	3421	1	0	0	0	0	1	Strain-specific
1862			3420	1	0	0	0	0	1	Strain-specific
1863	COG0847	[L] COG0847 DNA polymerase III, epsilon subunit and related 3'-5' exonucleases	3419	1	0	0	0	0	1	Strain-specific
1864	COG1917	[S] COG1917 Uncharacterized conserved protein, contains double-stranded beta-he	3418	1	0	0	0	0	1	Strain-specific
1865	COG0326	[O] COG0326 Molecular chaperone, HSP90 family	3417	1	0	0	0	0	1	Strain-specific
1866			3416	1	0	0	0	0	1	Strain-specific
1867	COG3550	[R] COG3550 Uncharacterized protein related to capsule biosynthesis enzymes	3415	1	0	0	0	0	1	Strain-specific
1868	COG0625	[O] COG0625 Glutathione S-transferase	3414	1	0	0	0	0	1	Strain-specific
1869			3413	1	0	0	0	0	1	Strain-specific
1870	COG0582	[L] COG0582 Integrase	3412	1	0	0	0	0	1	Strain-specific
1871			3411	1	0	0	0	0	1	Strain-specific
1872	COG0610	[V] COG0610 Type I site-specific restriction-modification system, R (restriction) subu	3410	1	0	0	0	0	1	Strain-specific
1873	COG0740	[OU] COG0740 Protease subunit of ATP-dependent Clp proteases	3409	1	0	0	0	0	1	Strain-specific
1874	COG0841	[V] COG0841 Cation/multidrug efflux pump	3408	1	0	0	0	0	1	Strain-specific
1875			3406	1	0	0	0	0	1	Strain-specific
1876			3405	1	0	0	0	0	1	Strain-specific
1877	COG1038	[C] COG1038 Pyruvate carboxylase	3404	1	0	0	0	0	1	Strain-specific
1878	COG3153	[R] COG3153 Predicted acetyltransferase	3403	1	0	0	0	0	1	Strain-specific
1879			3402	1	0	0	0	0	1	Strain-specific
1880	COG0433	[R] COG0433 Predicted ATPase	3401	1	0	0	0	0	1	Strain-specific
1881			3400	1	0	0	0	0	1	Strain-specific
1882	COG5001	[T] COG5001 Predicted signal transduction protein containing a membrane domain	3399	1	0	0	0	0	1	Strain-specific
1883	COG4974	[L] COG4974 Site-specific recombinase XerD	3398	1	0	0	0	0	1	Strain-specific
1884			3397	1	0	0	0	0	1	Strain-specific
1885			3396	1	0	0	0	0	1	Strain-specific
1886	COG3344	[L] COG3344 Retron-type reverse transcriptase	3363	1	0	0	0	0	1	Strain-specific
1887			3359	1	0	0	0	0	1	Strain-specific
1888	COG3227	[E] COG3227 Zinc metalloprotease (elastase)	3357	1	0	0	0	0	1	Strain-specific
1889	COG0438	[M] COG0438 Glycosyltransferase	3355	1	0	0	0	0	1	Strain-specific
1890	COG1004	[M] COG1004 Predicted UDP-glucose 6-dehydrogenase	3353	1	0	0	0	0	1	Strain-specific
1891	COG3306	[M] COG3306 Glycosyltransferase involved in LPS biosynthesis	3351	1	0	0	0	0	1	Strain-specific
1892	COG3501	[S] COG3501 Uncharacterized protein conserved in bacteria	3349	1	0	0	0	0	1	Strain-specific
1893			3348	1	0	0	0	0	1	Strain-specific
1894			3347	1	0	0	0	0	1	Strain-specific
1895	COG3668	[R] COG3668 Plasmid stabilization system protein	3345	1	0	0	0	0	1	Strain-specific
1896	COG2865	[K] COG2865 Predicted transcriptional regulator containing an HTH domain and an	3344	1	0	0	0	0	1	Strain-specific
1897	COG5606	[S] COG5606 Uncharacterized conserved small protein	3342	1	0	0	0	0	1	Strain-specific
1898			3341	1	0	0	0	0	1	Strain-specific
1899	COG4974	[L] COG4974 Site-specific recombinase XerD	3340	1	0	0	0	0	1	Strain-specific
1900			3339	1	0	0	0	0	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
1901	COG4176	[E] COG4176 ABC-type proline/glycine betaine transport system, permease compor	3338	1	0	0	0	0	1	Strain-specific
1902	COG0204	[I] COG0204 1-acyl-sn-glycerol-3-phosphate acyltransferase	3335	1	0	0	0	0	1	Strain-specific
1903			3334	1	0	0	0	0	1	Strain-specific
1904			3333	1	0	0	0	0	1	Strain-specific
1905			3332	1	0	0	0	0	1	Strain-specific
1906	COG3549	[R] COG3549 Plasmid maintenance system killer protein	3331	1	0	0	0	0	1	Strain-specific
1907	COG0607	[P] COG0607 Rhodanese-related sulfurtransferase	3329	1	0	0	0	0	1	Strain-specific
1908			3328	1	0	0	0	0	1	Strain-specific
1909	COG4771	[P] COG4771 Outer membrane receptor for ferrienterochelin and colicins	3327	1	0	0	0	0	1	Strain-specific
1910			3326	1	0	0	0	0	1	Strain-specific
1911			3325	1	0	0	0	0	1	Strain-specific
1912	COG0162	[J] COG0162 Tyrosyl-tRNA synthetase	3324	1	0	0	0	0	1	Strain-specific
1913	COG5002	[T] COG5002 Signal transduction histidine kinase	3323	1	0	0	0	0	1	Strain-specific
1914			3322	1	0	0	0	0	1	Strain-specific
1915	COG0583	[K] COG0583 Transcriptional regulator	3321	1	0	0	0	0	1	Strain-specific
1916			3320	1	0	0	0	0	1	Strain-specific
1917	COG1004	[M] COG1004 Predicted UDP-glucose 6-dehydrogenase	3319	1	0	0	0	0	1	Strain-specific
1918			3314	1	0	0	0	0	1	Strain-specific
1919			3313	1	0	0	0	0	1	Strain-specific
1920			3312	1	0	0	0	0	1	Strain-specific
1921			3310	1	0	0	0	0	1	Strain-specific
1922	COG3385	[L] COG3385 FOG: Transposase and inactivated derivatives	3309	1	0	0	0	0	1	Strain-specific
1923	COG3153	[R] COG3153 Predicted acetyltransferase	3181	1	0	0	0	0	1	Strain-specific
1924			3176	1	0	0	0	0	1	Strain-specific
1925	COG0732	[V] COG0732 Restriction endonuclease S subunits	3175	1	0	0	0	0	1	Strain-specific
1926			3174	1	0	0	0	0	1	Strain-specific
1927	COG1129	[G] COG1129 ABC-type sugar transport system, ATPase component	3172	1	0	0	0	0	1	Strain-specific
1928			3170	1	0	0	0	0	1	Strain-specific
1929	COG3950	[R] COG3950 Predicted ATP-binding protein involved in virulence	3168	1	0	0	0	0	1	Strain-specific
1930	COG1961	[L] COG1961 Site-specific recombinases, DNA invertase Pin homologs	3167	1	0	0	0	0	1	Strain-specific
1931	COG1209	[M] COG1209 dTDP-glucose pyrophosphorylase	3166	1	0	0	0	0	1	Strain-specific
1932			3165	1	0	0	0	0	1	Strain-specific
1933	COG3547	[L] COG3547 Transposase and inactivated derivatives	3164	1	0	0	0	0	1	Strain-specific
1934	COG3051	[C] COG3051 Citrate lyase, alpha subunit	3163	1	0	0	0	0	1	Strain-specific
1935	COG1263	[G] COG1263 Phosphotransferase system IIC components, glucose/maltose/N-acety	3162	1	0	0	0	0	1	Strain-specific
1936	COG1112	[L] COG1112 Superfamily I DNA and RNA helicases and helicase subunits	3161	1	0	0	0	0	1	Strain-specific
1937			3160	1	0	0	0	0	1	Strain-specific
1938	COG0338	[L] COG0338 Site-specific DNA methylase	3159	1	0	0	0	0	1	Strain-specific
1939			3157	1	0	0	0	0	1	Strain-specific
1940	COG2246	[S] COG2246 Predicted membrane protein	3156	1	0	0	0	0	1	Strain-specific
1941			3155	1	0	0	0	0	1	Strain-specific
1942			3153	1	0	0	0	0	1	Strain-specific
1943			3150	1	0	0	0	0	1	Strain-specific
1944			3149	1	0	0	0	0	1	Strain-specific
1945	COG3501	[S] COG3501 Uncharacterized protein conserved in bacteria	3147	1	0	0	0	0	1	Strain-specific
1946			3146	1	0	0	0	0	1	Strain-specific
1947			3145	1	0	0	0	0	1	Strain-specific
1948			3144	1	0	0	0	0	1	Strain-specific
1949	COG1204	[R] COG1204 Superfamily II helicase	3142	1	0	0	0	0	1	Strain-specific
1950			3141	1	0	0	0	0	1	Strain-specific
1951			3139	1	0	0	0	0	1	Strain-specific
1952			3138	1	0	0	0	0	1	Strain-specific
1953	COG3899	[R] COG3899 Predicted ATPase	3137	1	0	0	0	0	1	Strain-specific
1954	COG1264	[G] COG1264 Phosphotransferase system IIB components	3136	1	0	0	0	0	1	Strain-specific
1955			3135	1	0	0	0	0	1	Strain-specific
1956	COG4096	[V] COG4096 Type I site-specific restriction-modification system, R (restriction) subu	3134	1	0	0	0	0	1	Strain-specific
1957			3133	1	0	0	0	0	1	Strain-specific
1958	COG0223	[J] COG0223 Methionyl-tRNA formyltransferase	3132	1	0	0	0	0	1	Strain-specific
1959	COG3593	[L] COG3593 Predicted ATP-dependent endonuclease of the OLD family	3131	1	0	0	0	0	1	Strain-specific
1960			3129	1	0	0	0	0	1	Strain-specific
1961			3127	1	0	0	0	0	1	Strain-specific
1962	COG0657	[I] COG0657 Esterase/lipase	3125	1	0	0	0	0	1	Strain-specific
1963			3124	1	0	0	0	0	1	Strain-specific
1964			3123	1	0	0	0	0	1	Strain-specific
1965	COG0610	[V] COG0610 Type I site-specific restriction-modification system, R (restriction) subu	3122	1	0	0	0	0	1	Strain-specific
1966	COG2175	[Q] COG2175 Probable taurine catabolism dioxygenase	3121	1	0	0	0	0	1	Strain-specific
1967			2997	1	0	0	0	0	1	Strain-specific
1968			2993	1	0	0	0	0	1	Strain-specific
1969			2992	1	0	0	0	0	1	Strain-specific
1970	COG4319	[S] COG4319 Ketosteroid isomerase homolog	2991	1	0	0	0	0	1	Strain-specific
1971	COG0507	[L] COG0507 ATP-dependent exoDNAse (exonuclease V), alpha subunit - helicase su	2989	1	0	0	0	0	1	Strain-specific
1972	COG1974	[KT] COG1974 SOS-response transcriptional repressors (RecA-mediated autopeptid	2988	1	0	0	0	0	1	Strain-specific
1973	COG2026	[JD] COG2026 Cytotoxic translational repressor of toxin-antitoxin stability system	2987	1	0	0	0	0	1	Strain-specific
1974			2985	1	0	0	0	0	1	Strain-specific
1975	COG0732	[V] COG0732 Restriction endonuclease S subunits	2984	1	0	0	0	0	1	Strain-specific
1976			2983	1	0	0	0	0	1	Strain-specific
1977			2982	1	0	0	0	0	1	Strain-specific
1978			2980	1	0	0	0	0	1	Strain-specific
1979	COG1104	[E] COG1104 Cysteine sulfinate desulfinase/cysteine desulfurase and related enzym	2979	1	0	0	0	0	1	Strain-specific
1980	COG0545	[O] COG0545 FKBP-type peptidyl-prolyl cis-trans isomerases 1	2978	1	0	0	0	0	1	Strain-specific
1981	COG0695	[O] COG0695 Glutaredoxin and related proteins	2977	1	0	0	0	0	1	Strain-specific
1982	COG3675	[I] COG3675 Predicted lipase	2976	1	0	0	0	0	1	Strain-specific
1983	COG1396	[K] COG1396 Predicted transcriptional regulators	2975	1	0	0	0	0	1	Strain-specific
1984			2974	1	0	0	0	0	1	Strain-specific
1985			2973	1	0	0	0	0	1	Strain-specific
1986	COG4146	[R] COG4146 Predicted symporter	2971	1	0	0	0	0	1	Strain-specific
1987	COG3206	[M] COG3206 Uncharacterized protein involved in exopolysaccharide biosynthesis	2969	1	0	0	0	0	1	Strain-specific
1988			2968	1	0	0	0	0	1	Strain-specific
1989	COG2303	[E] COG2303 Choline dehydrogenase and related flavoproteins	2967	1	0	0	0	0	1	Strain-specific
1990	COG2453	[T] COG2453 Predicted protein-tyrosine phosphatase	2966	1	0	0	0	0	1	Strain-specific
1991			2964	1	0	0	0	0	1	Strain-specific
1992			2963	1	0	0	0	0	1	Strain-specific
1993			2962	1	0	0	0	0	1	Strain-specific
1994			2960	1	0	0	0	0	1	Strain-specific
1995			2959	1	0	0	0	0	1	Strain-specific
1996	COG0451	[MG] COG0451 Nucleoside-diphosphate-sugar epimerases	2958	1	0	0	0	0	1	Strain-specific
1997	COG1069	[C] COG1069 Ribulose kinase	2957	1	0	0	0	0	1	Strain-specific
1998			2956	1	0	0	0	0	1	Strain-specific
1999	COG0514	[L] COG0514 Superfamily II DNA helicase	2955	1	0	0	0	0	1	Strain-specific
2000	COG0583	[K] COG0583 Transcriptional regulator	2954	1	0	0	0	0	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
2001			2953	1	0	0	0	0	1	Strain-specific
2002	COG0732	[V] COG0732 Restriction endonuclease S subunits	2952	1	0	0	0	0	1	Strain-specific
2003			2951	1	0	0	0	0	1	Strain-specific
2004	COG3238	[S] COG3238 Uncharacterized protein conserved in bacteria	2950	1	0	0	0	0	1	Strain-specific
2005			2948	1	0	0	0	0	1	Strain-specific
2006			2946	1	0	0	0	0	1	Strain-specific
2007			2945	1	0	0	0	0	1	Strain-specific
2008			2944	1	0	0	0	0	1	Strain-specific
2009	COG0583	[K] COG0583 Transcriptional regulator	2943	1	0	0	0	0	1	Strain-specific
2010			2942	1	0	0	0	0	1	Strain-specific
2011	COG4128	[R] COG4128 Zonula occludens toxin	2806	1	0	0	0	0	1	Strain-specific
2012	COG0583	[K] COG0583 Transcriptional regulator	2805	1	0	0	0	0	1	Strain-specific
2013			2804	1	0	0	0	0	1	Strain-specific
2014	COG4096	[V] COG4096 Type I site-specific restriction-modification system, R (restriction) subu	2803	1	0	0	0	0	1	Strain-specific
2015	COG4710	[R] COG4710 Predicted DNA-binding protein with an HTH domain	2802	1	0	0	0	0	1	Strain-specific
2016	COG0827	[L] COG0827 Adenine-specific DNA methylase	2800	1	0	0	0	0	1	Strain-specific
2017	COG0110	[R] COG0110 Acetyltransferase (isoleucine patch superfamily)	2799	1	0	0	0	0	1	Strain-specific
2018	COG4580	[G] COG4580 Maltoporin (phage lambda and maltose receptor)	2798	1	0	0	0	0	1	Strain-specific
2019			2797	1	0	0	0	0	1	Strain-specific
2020	COG2161	[D] COG2161 Antitoxin of toxin-antitoxin stability system	2796	1	0	0	0	0	1	Strain-specific
2021			2795	1	0	0	0	0	1	Strain-specific
2022	COG0534	[V] COG0534 Na+-driven multidrug efflux pump	2794	1	0	0	0	0	1	Strain-specific
2023	COG5391	[UR] COG5391 Phox homology (PX) domain protein	2793	1	0	0	0	0	1	Strain-specific
2024			2792	1	0	0	0	0	1	Strain-specific
2025			2791	1	0	0	0	0	1	Strain-specific
2026	COG3436	[L] COG3436 Transposase and inactivated derivatives	2790	1	0	0	0	0	1	Strain-specific
2027	COG2946	[L] COG2946 Putative phage replication protein RstA	2787	1	0	0	0	0	1	Strain-specific
2028	COG1403	[V] COG1403 Restriction endonuclease	2786	1	0	0	0	0	1	Strain-specific
2029	COG4974	[L] COG4974 Site-specific recombinase XerD	2782	1	0	0	0	0	1	Strain-specific
2030			2778	1	0	0	0	0	1	Strain-specific
2031	COG2267	[I] COG2267 Lysophospholipase	2777	1	0	0	0	0	1	Strain-specific
2032			2776	1	0	0	0	0	1	Strain-specific
2033			2775	1	0	0	0	0	1	Strain-specific
2034	COG3153	[R] COG3153 Predicted acetyltransferase	2774	1	0	0	0	0	1	Strain-specific
2035			2773	1	0	0	0	0	1	Strain-specific
2036			2772	1	0	0	0	0	1	Strain-specific
2037			2771	1	0	0	0	0	1	Strain-specific
2038	COG5290	[K] COG5290 IkappaB kinase complex, IKAP component	2769	1	0	0	0	0	1	Strain-specific
2039			2768	1	0	0	0	0	1	Strain-specific
2040			2767	1	0	0	0	0	1	Strain-specific
2041			2766	1	0	0	0	0	1	Strain-specific
2042			2764	1	0	0	0	0	1	Strain-specific
2043	COG2148	[M] COG2148 Sugar transferases involved in lipopolysaccharide synthesis	2763	1	0	0	0	0	1	Strain-specific
2044	COG4316	[S] COG4316 Uncharacterized protein conserved in bacteria	2762	1	0	0	0	0	1	Strain-specific
2045	COG4974	[L] COG4974 Site-specific recombinase XerD	2761	1	0	0	0	0	1	Strain-specific
2046	COG3436	[L] COG3436 Transposase and inactivated derivatives	2760	1	0	0	0	0	1	Strain-specific
2047			2759	1	0	0	0	0	1	Strain-specific
2048	COG2842	[R] COG2842 Uncharacterized ATPase, putative transposase	2756	1	0	0	0	0	1	Strain-specific
2049			2755	1	0	0	0	0	1	Strain-specific
2050			2753	1	0	0	0	0	1	Strain-specific
2051	COG4974	[L] COG4974 Site-specific recombinase XerD	2750	1	0	0	0	0	1	Strain-specific
2052			2749	1	0	0	0	0	1	Strain-specific
2053	COG3962	[E] COG3962 Acetolactate synthase	2621	1	0	0	0	0	1	Strain-specific
2054			2619	1	0	0	0	0	1	Strain-specific
2055	COG0732	[V] COG0732 Restriction endonuclease S subunits	2618	1	0	0	0	0	1	Strain-specific
2056	COG1450	[NU] COG1450 Type II secretory pathway, component PulD	2614	1	0	0	0	0	1	Strain-specific
2057	COG0834	[ET] COG0834 ABC-type amino acid transport/signal transduction systems, periplas	2613	1	0	0	0	0	1	Strain-specific
2058	COG2871	[C] COG2871 Na+-transporting NADH:ubiquinone oxidoreductase, subunit NqrF	2611	1	0	0	0	0	1	Strain-specific
2059			2610	1	0	0	0	0	1	Strain-specific
2060			2608	1	0	0	0	0	1	Strain-specific
2061	COG3914	[O] COG3914 Predicted O-linked N-acetylglucosamine transferase, SPINDLY family	2606	1	0	0	0	0	1	Strain-specific
2062			2605	1	0	0	0	0	1	Strain-specific
2063	COG5281	[S] COG5281 Phage-related minor tail protein	2603	1	0	0	0	0	1	Strain-specific
2064	COG0399	[M] COG0399 Predicted pyridoxal phosphate-dependent enzyme apparently involv	2602	1	0	0	0	0	1	Strain-specific
2065	COG3668	[R] COG3668 Plasmid stabilization system protein	2599	1	0	0	0	0	1	Strain-specific
2066	COG5433	[L] COG5433 Transposase	2598	1	0	0	0	0	1	Strain-specific
2067			2597	1	0	0	0	0	1	Strain-specific
2068	COG3051	[C] COG3051 Citrate lyase, alpha subunit	2594	1	0	0	0	0	1	Strain-specific
2069	COG1664	[M] COG1664 Integral membrane protein CcmA involved in cell shape determinatic	2593	1	0	0	0	0	1	Strain-specific
2070			2592	1	0	0	0	0	1	Strain-specific
2071	COG4637	[R] COG4637 Predicted ATPase	2591	1	0	0	0	0	1	Strain-specific
2072			2589	1	0	0	0	0	1	Strain-specific
2073	COG3422	[S] COG3422 Uncharacterized conserved protein	2587	1	0	0	0	0	1	Strain-specific
2074			2585	1	0	0	0	0	1	Strain-specific
2075			2581	1	0	0	0	0	1	Strain-specific
2076	COG0558	[I] COG0558 Phosphatidylglycerophosphate synthase	2580	1	0	0	0	0	1	Strain-specific
2077	COG0346	[E] COG0346 Lactoylglutathione lyase and related lyases	2579	1	0	0	0	0	1	Strain-specific
2078			2578	1	0	0	0	0	1	Strain-specific
2079	COG0162	[J] COG0162 Tyrosyl-tRNA synthetase	2577	1	0	0	0	0	1	Strain-specific
2080			2576	1	0	0	0	0	1	Strain-specific
2081	COG2801	[L] COG2801 Transposase and inactivated derivatives	2575	1	0	0	0	0	1	Strain-specific
2082			2574	1	0	0	0	0	1	Strain-specific
2083			2573	1	0	0	0	0	1	Strain-specific
2084			2572	1	0	0	0	0	1	Strain-specific
2085			2571	1	0	0	0	0	1	Strain-specific
2086	COG1069	[C] COG1069 Ribulose kinase	2569	1	0	0	0	0	1	Strain-specific
2087			2567	1	0	0	0	0	1	Strain-specific
2088			2566	1	0	0	0	0	1	Strain-specific
2089	COG4974	[L] COG4974 Site-specific recombinase XerD	2565	1	0	0	0	0	1	Strain-specific
2090			2564	1	0	0	0	0	1	Strain-specific
2091	COG0641	[R] COG0641 Arylsulfatase regulator (Fe-S oxidoreductase)	2563	1	0	0	0	0	1	Strain-specific
2092			2561	1	0	0	0	0	1	Strain-specific
2093			2559	1	0	0	0	0	1	Strain-specific
2094			2557	1	0	0	0	0	1	Strain-specific
2095	COG4096	[V] COG4096 Type I site-specific restriction-modification system, R (restriction) subu	2556	1	0	0	0	0	1	Strain-specific
2096			2555	1	0	0	0	0	1	Strain-specific
2097			2554	1	0	0	0	0	1	Strain-specific
2098			2409	1	0	0	0	0	1	Strain-specific
2099	COG1328	[F] COG1328 Oxygen-sensitive ribonucleoside-triphosphate reductase	2406	1	0	0	0	0	1	Strain-specific
2100			2403	1	0	0	0	0	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
2101	COG3077	[L] COG3077 DNA-damage-inducible protein J	2398	1	0	0	0	0	1	Strain-specific
2102			2397	1	0	0	0	0	1	Strain-specific
2103			2240	1	0	0	0	0	1	Strain-specific
2104	COG1263	[G] COG1263 Phosphotransferase system IIC components, glucose/maltose/N-acety	2239	1	0	0	0	0	1	Strain-specific
2105	COG3981	[R] COG3981 Predicted acetyltransferase	2237	1	0	0	0	0	1	Strain-specific
2106			2236	1	0	0	0	0	1	Strain-specific
2107			2231	1	0	0	0	0	1	Strain-specific
2108			2227	1	0	0	0	0	1	Strain-specific
2109			2224	1	0	0	0	0	1	Strain-specific
2110			2221	1	0	0	0	0	1	Strain-specific
2111	COG3328	[L] COG3328 Transposase and inactivated derivatives	2220	1	0	0	0	0	1	Strain-specific
2112	COG1714	[S] COG1714 Predicted membrane protein/domain	2219	1	0	0	0	0	1	Strain-specific
2113	COG3950	[R] COG3950 Predicted ATP-binding protein involved in virulence	2216	1	0	0	0	0	1	Strain-specific
2114			2213	1	0	0	0	0	1	Strain-specific
2115	COG3093	[R] COG3093 Plasmid maintenance system antidote protein	2211	1	0	0	0	0	1	Strain-specific
2116			2208	1	0	0	0	0	1	Strain-specific
2117			2207	1	0	0	0	0	1	Strain-specific
2118			2205	1	0	0	0	0	1	Strain-specific
2119			2202	1	0	0	0	0	1	Strain-specific
2120	COG4646	[KL] COG4646 DNA methylase	2200	1	0	0	0	0	1	Strain-specific
2121			2191	1	0	0	0	0	1	Strain-specific
2122			2190	1	0	0	0	0	1	Strain-specific
2123	COG1704	[S] COG1704 Uncharacterized conserved protein	2187	1	0	0	0	0	1	Strain-specific
2124	COG0117	[H] COG0117 Pyrimidine deaminase	2186	1	0	0	0	0	1	Strain-specific
2125	COG4623	[M] COG4623 Predicted soluble lytic transglycosylase fused to an ABC-type amino a	2184	1	0	0	0	0	1	Strain-specific
2126			2183	1	0	0	0	0	1	Strain-specific
2127			2178	1	0	0	0	0	1	Strain-specific
2128	COG4580	[G] COG4580 Maltoporin (phage lambda and maltose receptor)	2177	1	0	0	0	0	1	Strain-specific
2129	COG0451	[MG] COG0451 Nucleoside-diphosphate-sugar epimerases	2175	1	0	0	0	0	1	Strain-specific
2130			2172	1	0	0	0	0	1	Strain-specific
2131	COG0790	[R] COG0790 FOG: TPR repeat, SEL1 subfamily	2171	1	0	0	0	0	1	Strain-specific
2132			2170	1	0	0	0	0	1	Strain-specific
2133	COG2932	[K] COG2932 Predicted transcriptional regulator	2169	1	0	0	0	0	1	Strain-specific
2134			2168	1	0	0	0	0	1	Strain-specific
2135	COG1479	[S] COG1479 Uncharacterized conserved protein	2167	1	0	0	0	0	1	Strain-specific
2136	COG4916	[S] COG4916 Uncharacterized protein containing a TIR (Toll-Interleukin 1-resistance	2166	1	0	0	0	0	1	Strain-specific
2137	COG4096	[V] COG4096 Type I site-specific restriction-modification system, R (restriction) subu	2165	1	0	0	0	0	1	Strain-specific
2138			2164	1	0	0	0	0	1	Strain-specific
2139	COG5642	[S] COG5642 Uncharacterized conserved protein	2160	1	0	0	0	0	1	Strain-specific
2140			2159	1	0	0	0	0	1	Strain-specific
2141	COG3501	[S] COG3501 Uncharacterized protein conserved in bacteria	2156	1	0	0	0	0	1	Strain-specific
2142			2155	1	0	0	0	0	1	Strain-specific
2143	COG0697	[GER] COG0697 Permeases of the drug/metabolite transporter (DMT) superfamily	2153	1	0	0	0	0	1	Strain-specific
2144	COG1388	[M] COG1388 FOG: LysM repeat	1863	1	0	0	0	0	1	Strain-specific
2145			1861	1	0	0	0	0	1	Strain-specific
2146	COG2207	[K] COG2207 AraC-type DNA-binding domain-containing proteins	1858	1	0	0	0	0	1	Strain-specific
2147			1857	1	0	0	0	0	1	Strain-specific
2148	COG4096	[V] COG4096 Type I site-specific restriction-modification system, R (restriction) subu	1852	1	0	0	0	0	1	Strain-specific
2149	COG1116	[P] COG1116 ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase con	1849	1	0	0	0	0	1	Strain-specific
2150	COG0625	[O] COG0625 Glutathione S-transferase	1846	1	0	0	0	0	1	Strain-specific
2151	COG5433	[L] COG5433 Transposase	1845	1	0	0	0	0	1	Strain-specific
2152	COG3950	[R] COG3950 Predicted ATP-binding protein involved in virulence	1843	1	0	0	0	0	1	Strain-specific
2153			1842	1	0	0	0	0	1	Strain-specific
2154			1840	1	0	0	0	0	1	Strain-specific
2155			1836	1	0	0	0	0	1	Strain-specific
2156	COG0381	[M] COG0381 UDP-N-acetylglucosamine 2-epimerase	1835	1	0	0	0	0	1	Strain-specific
2157	COG3203	[M] COG3203 Outer membrane protein (porin)	1830	1	0	0	0	0	1	Strain-specific
2158			1829	1	0	0	0	0	1	Strain-specific
2159			1826	1	0	0	0	0	1	Strain-specific
2160			1822	1	0	0	0	0	1	Strain-specific
2161	COG0399	[M] COG0399 Predicted pyridoxal phosphate-dependent enzyme apparently involv	1813	1	0	0	0	0	1	Strain-specific
2162			1811	1	0	0	0	0	1	Strain-specific
2163	COG5293	[S] COG5293 Uncharacterized protein conserved in bacteria	1810	1	0	0	0	0	1	Strain-specific
2164			1806	1	0	0	0	0	1	Strain-specific
2165			1805	1	0	0	0	0	1	Strain-specific
2166	COG4682	[S] COG4682 Predicted membrane protein	1803	1	0	0	0	0	1	Strain-specific
2167			1801	1	0	0	0	0	1	Strain-specific
2168			1800	1	0	0	0	0	1	Strain-specific
2169	COG4974	[L] COG4974 Site-specific recombinase XerD	1796	1	0	0	0	0	1	Strain-specific
2170			1795	1	0	0	0	0	1	Strain-specific
2171	COG0286	[V] COG0286 Type I restriction-modification system methyltransferase subunit	1794	1	0	0	0	0	1	Strain-specific
2172			1793	1	0	0	0	0	1	Strain-specific
2173			1792	1	0	0	0	0	1	Strain-specific
2174	COG2067	[I] COG2067 Long-chain fatty acid transport protein	1791	1	0	0	0	0	1	Strain-specific
2175	COG2067	[I] COG2067 Long-chain fatty acid transport protein	1789	1	0	0	0	0	1	Strain-specific
2176	COG0790	[R] COG0790 FOG: TPR repeat, SEL1 subfamily	1788	1	0	0	0	0	1	Strain-specific
2177	COG1476	[K] COG1476 Predicted transcriptional regulators	1787	1	0	0	0	0	1	Strain-specific
2178			1786	1	0	0	0	0	1	Strain-specific
2179	COG3547	[L] COG3547 Transposase and inactivated derivatives	1785	1	0	0	0	0	1	Strain-specific
2180			1784	1	0	0	0	0	1	Strain-specific
2181			1783	1	0	0	0	0	1	Strain-specific
2182			1780	1	0	0	0	0	1	Strain-specific
2183	COG5642	[S] COG5642 Uncharacterized conserved protein	1777	1	0	0	0	0	1	Strain-specific
2184			1776	1	0	0	0	0	1	Strain-specific
2185			1775	1	0	0	0	0	1	Strain-specific
2186	COG2333	[R] COG2333 Predicted hydrolase (metallo-beta-lactamase superfamily)	1774	1	0	0	0	0	1	Strain-specific
2187			1771	1	0	0	0	0	1	Strain-specific
2188	COG2826	[L] COG2826 Transposase and inactivated derivatives, IS30 family	1770	1	0	0	0	0	1	Strain-specific
2189	COG1205	[R] COG1205 Distinct helicase family with a unique C-terminal domain including a r	1769	1	0	0	0	0	1	Strain-specific
2190	COG0431	[R] COG0431 Predicted flavoprotein	1766	1	0	0	0	0	1	Strain-specific
2191	COG0790	[R] COG0790 FOG: TPR repeat, SEL1 subfamily	1466	1	0	0	0	0	1	Strain-specific
2192	COG1247	[M] COG1247 Sortase and related acyltransferases	1462	1	0	0	0	0	1	Strain-specific
2193	COG3110	[S] COG3110 Uncharacterized protein conserved in bacteria	1457	1	0	0	0	0	1	Strain-specific
2194	COG3765	[M] COG3765 Chain length determinant protein	1450	1	0	0	0	0	1	Strain-specific
2195	COG0732	[V] COG0732 Restriction endonuclease S subunits	1447	1	0	0	0	0	1	Strain-specific
2196			1446	1	0	0	0	0	1	Strain-specific
2197	COG1271	[C] COG1271 Cytochrome bd-type quinol oxidase, subunit 1	1444	1	0	0	0	0	1	Strain-specific
2198	COG5614	[R] COG5614 Bacteriophage head-tail adaptor	1442	1	0	0	0	0	1	Strain-specific
2199	COG1134	[GM] COG1134 ABC-type polysaccharide/polyol phosphate transport system, ATPas	1441	1	0	0	0	0	1	Strain-specific
2200			1438	1	0	0	0	0	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
2201	COG4974	[L] COG4974 Site-specific recombinase XerD	1437	1	0	0	0	0	1	Strain-specific
2202			1433	1	0	0	0	0	1	Strain-specific
2203			1431	1	0	0	0	0	1	Strain-specific
2204			1427	1	0	0	0	0	1	Strain-specific
2205	COG3550	[R] COG3550 Uncharacterized protein related to capsule biosynthesis enzymes	1426	1	0	0	0	0	1	Strain-specific
2206			1425	1	0	0	0	0	1	Strain-specific
2207	COG1450	[NU] COG1450 Type II secretory pathway, component PulD	1421	1	0	0	0	0	1	Strain-specific
2208	COG4096	[V] COG4096 Type I site-specific restriction-modification system, R (restriction) subu	1412	1	0	0	0	0	1	Strain-specific
2209	COG1767	[H] COG1767 Triphosphoribosyl-dephospho-CoA synthetase	1411	1	0	0	0	0	1	Strain-specific
2210	COG5302	[R] COG5302 Post-segregation antitoxin (ccd killing mechanism protein) encoded by	1408	1	0	0	0	0	1	Strain-specific
2211			1405	1	0	0	0	0	1	Strain-specific
2212	COG0433	[R] COG0433 Predicted ATPase	1401	1	0	0	0	0	1	Strain-specific
2213			1400	1	0	0	0	0	1	Strain-specific
2214			1329	1	0	0	0	0	1	Strain-specific
2215	COG0657	[I] COG0657 Esterase/lipase	1328	1	0	0	0	0	1	Strain-specific
2216			1327	1	0	0	0	0	1	Strain-specific
2217	COG3505	[U] COG3505 Type IV secretory pathway, VirD4 components	1326	1	0	0	0	0	1	Strain-specific
2218	COG3883	[S] COG3883 Uncharacterized protein conserved in bacteria	1325	1	0	0	0	0	1	Strain-specific
2219	COG0394	[T] COG0394 Protein-tyrosine-phosphatase	1324	1	0	0	0	0	1	Strain-specific
2220			1323	1	0	0	0	0	1	Strain-specific
2221	COG1209	[M] COG1209 dTDP-glucose pyrophosphorylase	1322	1	0	0	0	0	1	Strain-specific
2222	COG4653	[R] COG4653 Predicted phage phi-C31 gp36 major capsid-like protein	1320	1	0	0	0	0	1	Strain-specific
2223	COG3765	[M] COG3765 Chain length determinant protein	1319	1	0	0	0	0	1	Strain-specific
2224	COG4974	[L] COG4974 Site-specific recombinase XerD	1318	1	0	0	0	0	1	Strain-specific
2225	COG4469	[R] COG4469 Competence protein	1317	1	0	0	0	0	1	Strain-specific
2226	COG4974	[L] COG4974 Site-specific recombinase XerD	1314	1	0	0	0	0	1	Strain-specific
2227	COG1662	[L] COG1662 Transposase and inactivated derivatives, IS1 family	1312	1	0	0	0	0	1	Strain-specific
2228			1311	1	0	0	0	0	1	Strain-specific
2229			1310	1	0	0	0	0	1	Strain-specific
2230	COG3613	[F] COG3613 Nucleoside 2-deoxyribosyltransferase	1309	1	0	0	0	0	1	Strain-specific
2231			1308	1	0	0	0	0	1	Strain-specific
2232	COG2165	[NU] COG2165 Type II secretory pathway, pseudopilin PulG	1307	1	0	0	0	0	1	Strain-specific
2233	COG1401	[V] COG1401 GTPase subunit of restriction endonuclease	1306	1	0	0	0	0	1	Strain-specific
2234	COG1883	[C] COG1883 Na+-transporting methylmalonyl-CoA/oxaloacetate decarboxylase, be	1305	1	0	0	0	0	1	Strain-specific
2235			1304	1	0	0	0	0	1	Strain-specific
2236	COG4974	[L] COG4974 Site-specific recombinase XerD	1303	1	0	0	0	0	1	Strain-specific
2237	COG3049	[M] COG3049 Penicillin V acylase and related amidases	1302	1	0	0	0	0	1	Strain-specific
2238			1300	1	0	0	0	0	1	Strain-specific
2239			1299	1	0	0	0	0	1	Strain-specific
2240			1298	1	0	0	0	0	1	Strain-specific
2241			1297	1	0	0	0	0	1	Strain-specific
2242			1296	1	0	0	0	0	1	Strain-specific
2243	COG1209	[M] COG1209 dTDP-glucose pyrophosphorylase	1294	1	0	0	0	0	1	Strain-specific
2244			1293	1	0	0	0	0	1	Strain-specific
2245			1292	1	0	0	0	0	1	Strain-specific
2246			1291	1	0	0	0	0	1	Strain-specific
2247			1290	1	0	0	0	0	1	Strain-specific
2248	COG3706	[T] COG3706 Response regulator containing a CheY-like receiver domain and a GGD	1289	1	0	0	0	0	1	Strain-specific
2249			1288	1	0	0	0	0	1	Strain-specific
2250	COG1290	[C] COG1290 Cytochrome b subunit of the bc complex	1287	1	0	0	0	0	1	Strain-specific
2251	COG1767	[H] COG1767 Triphosphoribosyl-dephospho-CoA synthetase	1286	1	0	0	0	0	1	Strain-specific
2252			1285	1	0	0	0	0	1	Strain-specific
2253			1284	1	0	0	0	0	1	Strain-specific
2254	COG1566	[V] COG1566 Multidrug resistance efflux pump	1216	1	0	0	0	0	1	Strain-specific
2255			1214	1	0	0	0	0	1	Strain-specific
2256	COG0653	[U] COG0653 Preprotein translocase subunit SecA (ATPase, RNA helicase)	1213	1	0	0	0	0	1	Strain-specific
2257			1212	1	0	0	0	0	1	Strain-specific
2258			1211	1	0	0	0	0	1	Strain-specific
2259	COG0399	[M] COG0399 Predicted pyridoxal phosphate-dependent enzyme apparently involv	1210	1	0	0	0	0	1	Strain-specific
2260	COG0476	[H] COG0476 Dinucleotide-utilizing enzymes involved in molybdopterin and thiami	1209	1	0	0	0	0	1	Strain-specific
2261			1208	1	0	0	0	0	1	Strain-specific
2262	COG1335	[Q] COG1335 Amidases related to nicotinamidase	1207	1	0	0	0	0	1	Strain-specific
2263			1206	1	0	0	0	0	1	Strain-specific
2264			1205	1	0	0	0	0	1	Strain-specific
2265			1204	1	0	0	0	0	1	Strain-specific
2266	COG0399	[M] COG0399 Predicted pyridoxal phosphate-dependent enzyme apparently involv	1202	1	0	0	0	0	1	Strain-specific
2267	COG4974	[L] COG4974 Site-specific recombinase XerD	1201	1	0	0	0	0	1	Strain-specific
2268			1200	1	0	0	0	0	1	Strain-specific
2269			1199	1	0	0	0	0	1	Strain-specific
2270	COG4974	[L] COG4974 Site-specific recombinase XerD	1196	1	0	0	0	0	1	Strain-specific
2271	COG0372	[C] COG0372 Citrate synthase	1195	1	0	0	0	0	1	Strain-specific
2272			1194	1	0	0	0	0	1	Strain-specific
2273	COG1522	[K] COG1522 Transcriptional regulators	1192	1	0	0	0	0	1	Strain-specific
2274			1191	1	0	0	0	0	1	Strain-specific
2275	COG4748	[S] COG4748 Uncharacterized conserved protein	1190	1	0	0	0	0	1	Strain-specific
2276			1189	1	0	0	0	0	1	Strain-specific
2277	COG3307	[M] COG3307 Lipid A core - O-antigen ligase and related enzymes	1188	1	0	0	0	0	1	Strain-specific
2278	COG3415	[L] COG3415 Transposase and inactivated derivatives	1187	1	0	0	0	0	1	Strain-specific
2279			1186	1	0	0	0	0	1	Strain-specific
2280	COG0583	[K] COG0583 Transcriptional regulator	1185	1	0	0	0	0	1	Strain-specific
2281	COG4974	[L] COG4974 Site-specific recombinase XerD	1184	1	0	0	0	0	1	Strain-specific
2282			1183	1	0	0	0	0	1	Strain-specific
2283	COG0604	[CR] COG0604 NADPH:quinone reductase and related Zn-dependent oxidoreductas	1182	1	0	0	0	0	1	Strain-specific
2284			1181	1	0	0	0	0	1	Strain-specific
2285			1180	1	0	0	0	0	1	Strain-specific
2286			1179	1	0	0	0	0	1	Strain-specific
2287	COG0610	[V] COG0610 Type I site-specific restriction-modification system, R (restriction) subu	1178	1	0	0	0	0	1	Strain-specific
2288	COG0732	[V] COG0732 Restriction endonuclease S subunits	1177	1	0	0	0	0	1	Strain-specific
2289			1176	1	0	0	0	0	1	Strain-specific
2290	COG1088	[M] COG1088 dTDP-D-glucose 4,6-dehydratase	1175	1	0	0	0	0	1	Strain-specific
2291			1174	1	0	0	0	0	1	Strain-specific
2292			1173	1	0	0	0	0	1	Strain-specific
2293			1172	1	0	0	0	0	1	Strain-specific
2294			1170	1	0	0	0	0	1	Strain-specific
2295			1169	1	0	0	0	0	1	Strain-specific
2296	COG0542	[O] COG0542 ATPases with chaperone activity, ATP-binding subunit	1168	1	0	0	0	0	1	Strain-specific
2297			1167	1	0	0	0	0	1	Strain-specific
2298	COG4969	[NU] COG4969 Tfp pilus assembly protein, major pilin PilA	1166	1	0	0	0	0	1	Strain-specific
2299	COG1922	[M] COG1922 Teichoic acid biosynthesis proteins	1096	1	0	0	0	0	1	Strain-specific
2300			1095	1	0	0	0	0	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
2301			1094	1	0	0	0	0	1	Strain-specific
2302	COG2378	[K] COG2378 Predicted transcriptional regulator	1093	1	0	0	0	0	1	Strain-specific
2303			1092	1	0	0	0	0	1	Strain-specific
2304	COG3620	[K] COG3620 Predicted transcriptional regulator with C-terminal CBS domains	1091	1	0	0	0	0	1	Strain-specific
2305	COG0451	[MG] COG0451 Nucleoside-diphosphate-sugar epimerases	1090	1	0	0	0	0	1	Strain-specific
2306			1089	1	0	0	0	0	1	Strain-specific
2307	COG0207	[F] COG0207 Thymidylate synthase	1088	1	0	0	0	0	1	Strain-specific
2308			1087	1	0	0	0	0	1	Strain-specific
2309			1086	1	0	0	0	0	1	Strain-specific
2310	COG0657	[I] COG0657 Esterase/lipase	1085	1	0	0	0	0	1	Strain-specific
2311	COG0582	[L] COG0582 Integrase	1084	1	0	0	0	0	1	Strain-specific
2312	COG5338	[S] COG5338 Uncharacterized protein conserved in bacteria	1083	1	0	0	0	0	1	Strain-specific
2313	COG3501	[S] COG3501 Uncharacterized protein conserved in bacteria	1082	1	0	0	0	0	1	Strain-specific
2314	COG2274	[V] COG2274 ABC-type bacteriocin/lantibiotic exporters, contain an N-terminal dou	1081	1	0	0	0	0	1	Strain-specific
2315			1080	1	0	0	0	0	1	Strain-specific
2316	COG0384	[R] COG0384 Predicted epimerase, PhzC/PhzF homolog	1079	1	0	0	0	0	1	Strain-specific
2317			1077	1	0	0	0	0	1	Strain-specific
2318			1076	1	0	0	0	0	1	Strain-specific
2319	COG2160	[G] COG2160 L-arabinose isomerase	1075	1	0	0	0	0	1	Strain-specific
2320	COG4453	[S] COG4453 Uncharacterized protein conserved in bacteria	1074	1	0	0	0	0	1	Strain-specific
2321	COG3183	[V] COG3183 Predicted restriction endonuclease	1073	1	0	0	0	0	1	Strain-specific
2322	COG3791	[S] COG3791 Uncharacterized conserved protein	1072	1	0	0	0	0	1	Strain-specific
2323			1071	1	0	0	0	0	1	Strain-specific
2324	COG0125	[F] COG0125 Thymidylate kinase	1070	1	0	0	0	0	1	Strain-specific
2325			1069	1	0	0	0	0	1	Strain-specific
2326	COG3436	[L] COG3436 Transposase and inactivated derivatives	1068	1	0	0	0	0	1	Strain-specific
2327	COG0526	[OC] COG0526 Thiol-disulfide isomerase and thioredoxins	1067	1	0	0	0	0	1	Strain-specific
2328			1066	1	0	0	0	0	1	Strain-specific
2329			1065	1	0	0	0	0	1	Strain-specific
2330			1064	1	0	0	0	0	1	Strain-specific
2331	COG0210	[L] COG0210 Superfamily I DNA and RNA helicases	1063	1	0	0	0	0	1	Strain-specific
2332			1062	1	0	0	0	0	1	Strain-specific
2333	COG4695	[S] COG4695 Phage-related protein	1061	1	0	0	0	0	1	Strain-specific
2334	COG3926	[R] COG3926 Putative secretion activating protein	1060	1	0	0	0	0	1	Strain-specific
2335	COG3892	[S] COG3892 Uncharacterized protein conserved in bacteria	1058	1	0	0	0	0	1	Strain-specific
2336	COG0210	[L] COG0210 Superfamily I DNA and RNA helicases	1057	1	0	0	0	0	1	Strain-specific
2337	COG2963	[L] COG2963 Transposase and inactivated derivatives	1056	1	0	0	0	0	1	Strain-specific
2338	COG3950	[R] COG3950 Predicted ATP-binding protein involved in virulence	1055	1	0	0	0	0	1	Strain-specific
2339			1054	1	0	0	0	0	1	Strain-specific
2340	COG3593	[L] COG3593 Predicted ATP-dependent endonuclease of the OLD family	1052	1	0	0	0	0	1	Strain-specific
2341			1051	1	0	0	0	0	1	Strain-specific
2342			1050	1	0	0	0	0	1	Strain-specific
2343	COG1359	[S] COG1359 Uncharacterized conserved protein	1049	1	0	0	0	0	1	Strain-specific
2344	COG3593	[L] COG3593 Predicted ATP-dependent endonuclease of the OLD family	1048	1	0	0	0	0	1	Strain-specific
2345			1000	1	0	0	0	0	1	Strain-specific
2346	COG0583	[K] COG0583 Transcriptional regulator	999	1	0	0	0	0	1	Strain-specific
2347	COG0286	[V] COG0286 Type I restriction-modification system methyltransferase subunit	998	1	0	0	0	0	1	Strain-specific
2348			997	1	0	0	0	0	1	Strain-specific
2349			996	1	0	0	0	0	1	Strain-specific
2350			995	1	0	0	0	0	1	Strain-specific
2351	COG0790	[R] COG0790 FOG: TPR repeat, SEL1 subfamily	994	1	0	0	0	0	1	Strain-specific
2352			992	1	0	0	0	0	1	Strain-specific
2353	COG3829	[KT] COG3829 Transcriptional regulator containing PAS, AAA-type ATPase, and DNA	991	1	0	0	0	0	1	Strain-specific
2354	COG4096	[V] COG4096 Type I site-specific restriction-modification system, R (restriction) subu	990	1	0	0	0	0	1	Strain-specific
2355	COG4928	[R] COG4928 Predicted P-loop ATPase	989	1	0	0	0	0	1	Strain-specific
2356	COG4626	[R] COG4626 Phage terminase-like protein, large subunit	988	1	0	0	0	0	1	Strain-specific
2357			987	1	0	0	0	0	1	Strain-specific
2358			983	1	0	0	0	0	1	Strain-specific
2359			980	1	0	0	0	0	1	Strain-specific
2360			977	1	0	0	0	0	1	Strain-specific
2361			972	1	0	0	0	0	1	Strain-specific
2362			971	1	0	0	0	0	1	Strain-specific
2363	COG0454	[KR] COG0454 Histone acetyltransferase HPA2 and related acetyltransferases	970	1	0	0	0	0	1	Strain-specific
2364	COG0697	[GER] COG0697 Permeases of the drug/metabolite transporter (DMT) superfamily	967	1	0	0	0	0	1	Strain-specific
2365	COG2378	[K] COG2378 Predicted transcriptional regulator	617	1	0	0	0	0	1	Strain-specific
2366			615	1	0	0	0	0	1	Strain-specific
2367			610	1	0	0	0	0	1	Strain-specific
2368			600	1	0	0	0	0	1	Strain-specific
2369			599	1	0	0	0	0	1	Strain-specific
2370			598	1	0	0	0	0	1	Strain-specific
2371			596	1	0	0	0	0	1	Strain-specific
2372	COG0367	[E] COG0367 Asparagine synthase (glutamine-hydrolyzing)	591	1	0	0	0	0	1	Strain-specific
2373	COG3593	[L] COG3593 Predicted ATP-dependent endonuclease of the OLD family	582	1	0	0	0	0	1	Strain-specific
2374	COG1208	[MJ] COG1208 Nucleoside-diphosphate-sugar pyrophosphorylase involved in lipop	581	1	0	0	0	0	1	Strain-specific
2375			579	1	0	0	0	0	1	Strain-specific
2376			578	1	0	0	0	0	1	Strain-specific
2377			573	1	0	0	0	0	1	Strain-specific
2378	COG1519	[M] COG1519 3-deoxy-D-manno-octulosonic-acid transferase	565	1	0	0	0	0	1	Strain-specific
2379	COG2202	[T] COG2202 FOG: PAS/PAC domain	564	1	0	0	0	0	1	Strain-specific
2380	COG4974	[L] COG4974 Site-specific recombinase XerD	562	1	0	0	0	0	1	Strain-specific
2381	COG1686	[M] COG1686 D-alanyl-D-alanine carboxypeptidase	558	1	0	0	0	0	1	Strain-specific
2382			555	1	0	0	0	0	1	Strain-specific
2383			552	1	0	0	0	0	1	Strain-specific
2384			548	1	0	0	0	0	1	Strain-specific
2385	COG0745	[TK] COG0745 Response regulators consisting of a CheY-like receiver domain and a	547	1	0	0	0	0	1	Strain-specific
2386			546	1	0	0	0	0	1	Strain-specific
2387			544	1	0	0	0	0	1	Strain-specific
2388	COG4973	[L] COG4973 Site-specific recombinase XerC	543	1	0	0	0	0	1	Strain-specific
2389			542	1	0	0	0	0	1	Strain-specific
2390			539	1	0	0	0	0	1	Strain-specific
2391			538	1	0	0	0	0	1	Strain-specific
2392			537	1	0	0	0	0	1	Strain-specific
2393	COG2274	[V] COG2274 ABC-type bacteriocin/lantibiotic exporters, contain an N-terminal dou	536	1	0	0	0	0	1	Strain-specific
2394			535	1	0	0	0	0	1	Strain-specific
2395			531	1	0	0	0	0	1	Strain-specific
2396			529	1	0	0	0	0	1	Strain-specific
2397	COG3550	[R] COG3550 Uncharacterized protein related to capsule biosynthesis enzymes	527	1	0	0	0	0	1	Strain-specific
2398	COG1038	[C] COG1038 Pyruvate carboxylase	524	1	0	0	0	0	1	Strain-specific
2399	COG0515	[RTKL] COG0515 Serine/threonine protein kinase	523	1	0	0	0	0	1	Strain-specific
2400			522	1	0	0	0	0	1	Strain-specific

Table S4. (Continued)

	COG	COG term	ClutserID	strain 100719-1/3A	strain LMG 13583	strain VK1	strain 090707-1/1A	strain LMG 4437	sum	type
2401			521	1	0	0	0	0	1	Strain-specific
2402			519	1	0	0	0	0	1	Strain-specific
2403	COG1086	[MG] COG1086 Predicted nucleoside-diphosphate sugar epimerases	518	1	0	0	0	0	1	Strain-specific
2404	COG0207	[F] COG0207 Thymidylate synthase	517	1	0	0	0	0	1	Strain-specific
2405	COG1846	[K] COG1846 Transcriptional regulators	514	1	0	0	0	0	1	Strain-specific
2406	COG0656	[R] COG0656 Aldo/keto reductases, related to diketogulonate reductase	115	1	0	0	0	0	1	Strain-specific
2407	COG3609	[K] COG3609 Predicted transcriptional regulators containing the CopG/Arc/MetJ DN	113	1	0	0	0	0	1	Strain-specific
2408	COG1474	[LO] COG1474 Cdc6-related protein, AAA superfamily ATPase	101	1	0	0	0	0	1	Strain-specific
2409			92	1	0	0	0	0	1	Strain-specific
2410	COG0270	[L] COG0270 Site-specific DNA methylase	91	1	0	0	0	0	1	Strain-specific
2411	COG3547	[L] COG3547 Transposase and inactivated derivatives	90	1	0	0	0	0	1	Strain-specific
2412			89	1	0	0	0	0	1	Strain-specific
2413			88	1	0	0	0	0	1	Strain-specific
2414			85	1	0	0	0	0	1	Strain-specific
2415			82	1	0	0	0	0	1	Strain-specific
2416	COG3708	[S] COG3708 Uncharacterized protein conserved in bacteria	81	1	0	0	0	0	1	Strain-specific
2417			75	1	0	0	0	0	1	Strain-specific
2418	COG1205	[R] COG1205 Distinct helicase family with a unique C-terminal domain including a r	68	1	0	0	0	0	1	Strain-specific
2419			67	1	0	0	0	0	1	Strain-specific
2420	COG4974	[L] COG4974 Site-specific recombinase XerD	65	1	0	0	0	0	1	Strain-specific
2421			61	1	0	0	0	0	1	Strain-specific
2422	COG0028	[EH] COG0028 Thiamine pyrophosphate-requiring enzymes [acetolactate synthase,	60	1	0	0	0	0	1	Strain-specific
2423			52	1	0	0	0	0	1	Strain-specific
2424	COG0494	[LR] COG0494 NTP pyrophosphohydrolases including oxidative damage repair enzy	47	1	0	0	0	0	1	Strain-specific
2425			40	1	0	0	0	0	1	Strain-specific
2426			36	1	0	0	0	0	1	Strain-specific
2427	COG2946	[L] COG2946 Putative phage replication protein RstA	35	1	0	0	0	0	1	Strain-specific
2428			31	1	0	0	0	0	1	Strain-specific
2429			30	1	0	0	0	0	1	Strain-specific
2430			29	1	0	0	0	0	1	Strain-specific
2431			27	1	0	0	0	0	1	Strain-specific
2432	COG0328	[L] COG0328 Ribonuclease HI	26	1	0	0	0	0	1	Strain-specific
2433			25	1	0	0	0	0	1	Strain-specific
2434	COG0790	[R] COG0790 FOG: TPR repeat, SEL1 subfamily	24	1	0	0	0	0	1	Strain-specific
2435			22	1	0	0	0	0	1	Strain-specific
2436	COG0419	[L] COG0419 ATPase involved in DNA repair	21	1	0	0	0	0	1	Strain-specific
2437	COG0507	[L] COG0507 ATP-dependent exoDNAse (exonuclease V), alpha subunit - helicase su	19	1	0	0	0	0	1	Strain-specific
2438			17	1	0	0	0	0	1	Strain-specific
2439	COG0515	[RTKL] COG0515 Serine/threonine protein kinase	16	1	0	0	0	0	1	Strain-specific
2440	COG1883	[C] COG1883 Na+-transporting methylmalonyl-CoA/oxaloacetate decarboxylase, be	14	1	0	0	0	0	1	Strain-specific
2441			13	1	0	0	0	0	1	Strain-specific
2442	COG0451	[MG] COG0451 Nucleoside-diphosphate-sugar epimerases	7	1	0	0	0	0	1	Strain-specific
			6	1	0	0	0	0	1	Strain-specific
			3	1	0	0	0	0	1	Strain-specific