Metagenomic Diagnosis of Bacterial Infections

Shota Nakamura, Norihiro Maeda, Ionut Mihai Miron, Myonsun Yoh, Kaori Izutsu, Chidoh Kataoka, Takeshi Honda, Teruo Yasunaga, Takaaki Nakaya, Jun Kawai, Yoshihide Hayashizaki, Toshihiro Horii, and Tetsuya Iida

To test the ability of high-throughput DNA sequencing to detect bacterial pathogens, we used it on DNA from a patient's feces during and after diarrheal illness. Sequences showing best matches for *Campylobacter jejuni* were detected only in the illness sample. Various bacteria may be detectable with this metagenomic approach.

Infectious diseases are caused by various pathogens, including as-yet unidentified microorganisms. Because procedures for detecting and identifying pathogens vary according to the target microorganism, clinical examinations require a variety of media, reagents, and culture methods. In addition, conventional examination protocols usually require much labor, time, and skill, thus forming an obstacle to a prompt diagnosis.

Newly developed, "next-generation" DNA sequencers can determine >100 megabases of DNA sequences per run (*1*). These new technologies eliminate the bacterial cloning step used in traditional Sanger sequencing; instead, they amplify single isolated DNA molecules and analyze them with massively parallel processing. To develop a new system to promptly detect and identify various infectious pathogens, we tapped into the potential of these novel sequencers. We directly detected the causative pathogenic microbe in a clinical human sample (diarrheic feces) by means of unbiased high-throughput DNA sequencing.

The Study

A 34-year-old man had become ill after eating dinner out with his family. After 3 days, severe diarrhea, stomach ache, and shivering developed in the only 3 persons (the patient plus 2 family members) who had eaten undercooked chicken that night. Four days after onset of clinical signs, feces were collected from the patient and stored in a freezer at –80°C. At a clinical laboratory in Osaka, Japan, conventional culture methods were used to examine the sample for possible bacterial enteropathogens (*2*), and specific reverse transcriptase–PCR was used to test for norovirus (*3*); however, no candidate pathogens were detected.

We therefore analyzed this fecal sample for possible pathogens by means of high-throughput DNA sequencing. DNA was extracted from the diarrhea sample (hereafter referred to as the illness DNA sample) with a QIAamp DNA Stool Mini Kit (QIAGEN, Valencia, CA, USA). After the man had completely recovered 3 months later, another fecal sample was collected (hereafter referred to as the recovery DNA sample) and maintained at –80°C until DNA extraction. Both DNA samples were subjected to unbiased high-throughput DNA sequencing with a GS20 sequencer (454 Life Sciences, Branford, CT, USA) (*4*).

Sequencing produced 96,941 effective sequences for the illness DNA sample and 106,327 for the recovery sample. The average length of the sequences was 102.1 bp. The DNA sequences obtained were searched with the BLASTN program for the National Center for Biotechnology Information nucleotide sequence database (http://blast.ncbi.nlm.nih.gov). The BLASTN output was then analyzed by using a classification system consisting of the Center's taxonomy database and its searching system. This system, devised with the aid of Perl language (www.perl.com) and the MySQL database (www.mysql.com), facilitates the identification of scientific names and statistical analysis. The Figure shows the organisms from which the sequences in the database were derived that showed the best matches for the sequences queried (expect [E]-value $<10^{-5}$). For both DNA samples, ≈20% of the total sequences showed the best matches for the currently reported bacterial DNA sequences. The Table shows the frequency distributions of species from which close matches for the sequences were derived (E-value $<10^{-40}$). The most frequently detected bacterial species in both samples belonged to the phylum Bacteroidetes, the normal flora of the human intestine. No major differences were found in the frequency of the species between the illness and recovery DNA samples.

A striking difference between the 2 samples, however, was that 156 sequences of the illness DNA sample showed best matches for the sequences derived from *Campylobacter jejuni*, but no sequences of the recovery DNA sample showed any such significant matches. The *C. jejuni* sequences from the illness DNA sample included many housekeeping genes, such as the genes for the ribosomal RNAs and DNA polymerases (online Appendix Table, available from www.cdc.gov/EID/content/14/11/1784-appT.htm); thus, they strongly suggested the presence of *C. jejuni* in the illness fecal sample.

Because *C. jejuni* is a bacterium that causes acute gastroenteritis and is normally not present in the intestines of

Author affiliations: Osaka University, Suita, Japan (S. Nakamura, I.M. Miron, M. Yoh, K. Izutsu, C. Kataoka, T. Honda, T. Yasunaga, T. Nakaya, T. Horii, T. Iida); and RIKEN Yokohama Institute, Yokohama, Japan (N. Maeda, J. Kawai, Y. Hayashizaki)

DOI: 10.3201/eid1411.080589

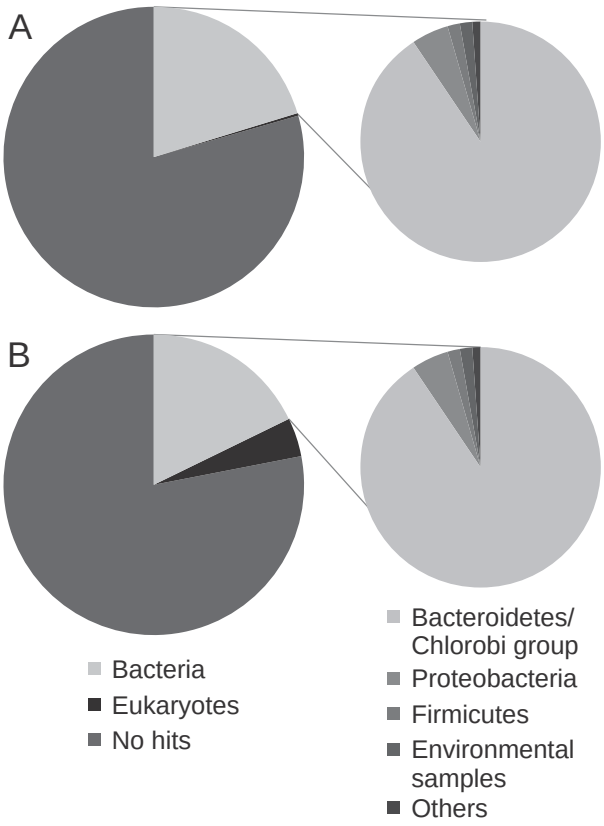

Figure. Comparison of the organisms from which the best matches for the sequences were derived from a BLASTN (http://blast.ncbi.nlm.nih.gov) search with an expect-value cutoff of 10^{-5}. A) DNA from nondiarrheic fecal sample collected 3 months after patient had recovered. B) DNA from diarrheic fecal sample collected while patient was ill.

healthy persons (*5*,*6*), these results prompted us to reexamine the illness fecal sample for *C. jejuni*. For the illness sample but not the recovery DNA sample, *Campylobacter*-specific PCR (*7*) produced a typical banding pattern that is unique to *C. jejuni* (data not shown). The recovery rate of *Campylobacter* spp. from patient specimens substantially decreases when the specimens are frozen before isolation (*8*). To obtain higher recovery of *Campylobacter* spp. and thus validate the presence of *C. jejuni* in the illness sample, we performed cultures with enrichment and selective media again on the frozen illness fecal sample (*5*). *C. jejuni*–like bacteria with corkscrew motility grew on selective agar plates. Biochemical identification using the API Campy kit (API-bioMérieux, Marcy L'Etoile, France) demonstrated that the organism was *C. jejuni*, thus proving its presence in the illness fecal sample.

Conclusions

We directly detected a bacterial pathogen in a patient sample by using high-throughput DNA sequencing. This finding implies that basically any kind of bacterial pathogen may be detectable with a common procedure. The method is directly applicable not only to fecal samples but also to other types of clinical samples; it could detect and identify bacterial pathogens that are usually difficult to ascertain with conventional examination procedures. Because this novel approach can be expected to have major potential for detection of pathogens in various infectious diseases, it warrants further investigation.

The approach reported here also enabled us to directly analyze the ratio of pathogenic to commensal bacteria in the human intestine. Assessment of the relative population of intestinal bacteria would enable us to investigate the dynamics of bacterial pathogens in human intestines, in relation to associated intestinal microbial flora, during infectious disease processes.

Many causative agents of emerging infectious diseases are of animal origin, and many are previously identified microbes (*9*,*10*). Because a vast amount of genome information about various microorganisms is continually being accumulated in databases, the approach we used will become increasingly useful. Recent metagenomic studies have identified unknown virus pathogens (*11*–*13*). Using the present approach to analyze various clinical cases, especially of outbreaks of infectious diseases with as-yet unidentified causative agents, may lead to the discovery of novel bacteria that are currently not known to be pathogenic to humans.

The current cost for high-throughput sequencing may limit the use of this method to specialized purposes, such as the hunt for novel pathogens for research or detection of bioterrorism (*14*). However, because the progress of DNA sequencing technology has been rapid (*1*), the cost, time, and labor for sequencing have been greatly reduced, and this trend will likely continue for the foreseeable future

Table. Frequency distributions of species in fecal samples taken from patient during illness and after recovery, as determined by BLASTN*

Organism	No. (%)	
	Illness†	Recovery‡
Bacteroides vulgatus	5,944 (50.5)	4,743 (56.5)
Homo sapiens	2,955 (25.1)	84 (1.0)
Parabacteroides distasonis	818 (6.9)	1,283 (15.3)
B. thetaiotaomicron	767 (6.5)	1,046 (12.5)
B. fragilis	759 (6.4)	842 (10.0)
Uncultured bacterium	195 (1.7)	227 (2.7)
Campylobacter jejuni	156 (1.3)	0
B. ovatus	48 (0.4)	63 (0.8)
Uncultured *Bacteroides* spp.	20 (0.2)	19 (0.2)
B. uniformis	14 (0.1)	8 (0.1)

*BLASTN available from http://blast.ncbi.nlm.nih.gov. Expect-value cutoff 10^{-40}.
†Diarrheic fecal sample collected while patient was ill. Total sequences 96,941; total (100%) BLAST matches 11,777.
‡Nondiarrheic fecal sample collected 3 mo after patient had recovered. Total sequences 106,327; total (100%) BLAST matches 8,397.

(*15*). Therefore, high-throughput DNA sequencing may soon be adopted as the main method for examining microorganisms in major clinical laboratories. The data presented here represent an example of this major innovation in the field of clinical examination for causative agents of infectious diseases.

Acknowledgments

We are grateful to Y. Nagai and Y. Okamoto for their help coordinating this study, to R. Dryselius and Y. Nishimune for their helpful suggestions, to M. Tagami and H. Sano for technical support, and to N.M.Q. Palacpac for valuable comments on the text.

This study was supported by the Program of Founding Research Centers for Emerging and Reemerging Infectious Diseases, by Grants-in-Aid for Scientific Research, and by a Research Grant for the RIKEN Genome Exploration Research Project (to Y. H.) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

This study was approved by the ethical review committees of the Research Institute for Microbial Diseases, Osaka University, and RIKEN. The sequencing data reported here are available in the Short Read Archive database at the National Center for Biotechnology Information under accession no. SRA001127.

Dr Nakamura is a researcher in the Section of Bioinformatics, Thailand-Japan Research Collaboration Center on Emerging and Reemerging Infections, Research Institute for Microbial Diseases, Osaka University. His research interests have included crystallographic analysis for biomacromolecules, which he currently applies to his work in bioinformatics.

References

1. Service RF. Gene sequencing: the race for the $1000 genome. Science. 2006;311:1544–6. DOI: 10.1126/science.311.5767.1544
2. Saidi SM, Iijima Y, Sang WK, Mwangudza AK, Oundo JO, Taga K, et al. Epidemiological study on infectious diarrheal diseases in children in a coastal rural area of Kenya. Microbiol Immunol. 1997;41:773–8.
3. Sakon N, Yamazaki K, Yoda T, Tsukamoto T, Kase T, Taniguchi K, et al. Norovirus storm in Osaka, Japan, last winter (2006/2007). Jpn J Infect Dis. 2007;60:409–10.
4. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437:376–80.
5. Penner JL. The genus *Campylobacter*: a decade of progress. Clin Microbiol Rev. 1988;1:157–72.
6. Young KT, Davis LM, DiRita VJ. *Campylobacter jejuni*: molecular biology and pathogenesis. Nat Rev Microbiol. 2007;5:665–79. DOI: 10.1038/nrmicro1718
7. Fermér C, Engvall EO. Specific PCR identification and differentiation of the thermophilic campylobacters, *Campylobacter jejuni, C. coli, C. lari*, and *C. upsaliensis.* J Clin Microbiol. 1999;37:3370–3.
8. Altekruse SF, Stern NJ, Fields PI, Swerdlow DL. *Campylobacter jejuni*—an emerging foodborne pathogen. Emerg Infect Dis. 1999;5:28–35.
9. Morens DM, Folkers GK, Fauci AS. The challenge of emerging and re-emerging infectious diseases. Nature. 2004;430:242–9. DOI: 10.1038/nature02759
10. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature. 2008;451:990–3. DOI: 10.1038/nature06536
11. Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD, Moran NA, et al. A metagenomic survey of microbes in honey bee colony collapse disorder. Science. 2007;318:283–7. DOI: 10.1126/science.1146498
12. Palacios G, Druce J, Du L, Tran T, Birch C, Briese T, et al. A new arenavirus in a cluster of fatal transplant-associated diseases. N Engl J Med. 2008;358:991–8. DOI: 10.1056/NEJMoa073785
13. Finkbeiner SR, Allred AF, Tarr PI, Klein EJ, Kirkwood CD, Wang D. Metagenomic analysis of human diarrhea: viral detection and discovery. PLoS Pathog. 2008;4:e1000011. DOI: 10.1371/journal.ppat.1000011
14. Lim DV, Simpson JM, Kearns EA, Kramer MF. Current and developing technologies for monitoring agents of bioterrorism and biowarfare. Clin Microbiol Rev. 2005;18:583–607. DOI: 10.1128/CMR.18.4.583-607.2005
15. von Bubnoff A. Next-generation sequencing: the race is on. Cell. 2008;132:721–3. DOI: 10.1016/j.cell.2008.02.028

Address for correspondence: Tetsuya Iida, International Research Center for Infectious Diseases, Research Institute for Microbial Diseases, Osaka University, 3-1 Yamadaoka, Suita, Osaka 565-0871, Japan; email: iida@biken.osaka-u.ac.jp

Appendix Table. BLASTN matches for *Campylobacter jejuni* sequences*

Accession no.	Start	End	E-value†	BLAST hit
CP000538	84217	84073	6.00E-75	Flagellar motor switch protein (*fliY*)
CP000538	820953	821083	1.00E-66	Ribosomal protein S15 (*rpsO*)
AF130466	7941	7806	2.00E-65	Putative glycosyltransferase
AF130466	7941	7806	2.00E-65	Putative glycosyltransferase
CP000538	896576	896450	3.00E-64	Peptidyl-arginine deiminase family protein
CP000538	248199	248324	1.00E-63	Biotin sulfoxide reductase, molybdopterin containing oxidoreductase
CP000538	222134	222259	1.00E-63	Conserved hypothetical protein
AF167344	22157	22033	4.00E-63	Heavy metal translocating P-type ATPase
CP000538	1498737	1498860	2.00E-62	Translation initiation factor IF-1 (*infA*)
CP000538	513409	513287	6.00E-62	Conserved hypothetical protein
CP000538	864240	864119	2.00E-61	GTP-binding protein YchF
CP000538	185407	185288	4.00E-60	Transporter
CP000538	113295	113413	1.00E-59	Hypothetical protein
CP000538	461714	461596	1.00E-59	Putative periplasmic protein
CP000538	556839	556714	2.00E-59	MOSC domain protein
AL111168	1233367	1233484	6.00E-59	3-oxoacyl-(acyl-carrier-protein) synthase
CP000538	774644	774761	6.00E-59	tRNA pseudouridine synthase A (*truA*)
CP000538	774644	774761	6.00E-59	tRNA pseudouridine synthase A (*truA*)
CP000768	196036	196153	6.00E-59	Putative periplasmic protein
DQ493919	2207	2083	6.00E-59	Heat-inducible transcription repressor (*hrcA*)
CP000025	212978	213094	2.00E-58	Hypothetical protein
CP000538	1310351	1310467	2.00E-58	Catalase (*katA*)
CP000768	196036	196152	2.00E-58	Putative periplasmic protein

CP000538	1521206	1521082	3.00E-58	Hemin ABC transporter
CP000538	339049	338934	9.00E-58	RND efflux system, inner membrane transporter (*cmeB*)
CP000538	809625	809502	9.00E-58	Site-specific recombinase, phage integrase family
CP000538	461702	461588	3.00E-57	Putative periplasmic protein
CP000538	1522395	1522509	3.00E-57	Putative radical SAM domain protein
CP000538	451930	452044	3.00E-57	Translation elongation factor G (*fusA*)
CP000538	774644	774758	3.00E-57	tRNA pseudouridine synthase A (*truA*)
CP000025	211254	211367	1.00E-56	Tail tape measure protein, TP901 family
CP000538	303524	303637	1.00E-56	3-oxoacyl-(acyl-carrier-protein) synthase III (*fabH*)
CP000538	1471284	1471397	1.00E-56	NADH-quinone oxidoreductase, N subunit (*nuoN*)
CP000538	789384	789271	1.00E-56	Putative lipoprotein
AY102622	19934	19822	5.00E-56	Motility accessory factor
AY102622	19934	19822	5.00E-56	Motility accessory factor
CP000538	1371661	1371549	5.00E-56	Peptide chain release factor 2 (*prfB*)
DQ174139	1132	1020	5.00E-56	16S ribosomal RNA (*rrsA*)
CP000538	39117	39006	2.00E-55	Intergenic sequence
CP000538	236407	236518	2.00E-55	Major facilitator superfamily protein
AL111168	905046	905156	8.00E-55	Putative periplasmic protein
AL111168	905046	905156	8.00E-55	Putative periplasmic protein
BX545860	5780	5890	8.00E-55	Putative methyltransferase
CP000538	346235	346125	8.00E-55	Conserved hypothetical protein
CP000538	346235	346125	8.00E-55	Conserved hypothetical protein
CP000768	196036	196153	8.00E-55	Putative periplasmic protein
AL111168	1123274	1123154	3.00E-54	Phosphate transporter family protein
CP000025	211681	211572	3.00E-54	Tail tape measure protein, TP901 family
CP000538	1126394	1126503	3.00E-54	Putative dihydroorotase
CP000538	461697	461588	3.00E-54	Putative periplasmic protein
CP000538	185853	185962	3.00E-54	Putative transporter
AY102622	29478	29354	1.00E-53	Motility accessory factor

CP000538	727769	727649	1.00E-53	Conserved hypothetical protein
CP000538	1265191	1265083	1.00E-53	Hypothetical protein
AL111168	756034	755923	5.00E-53	Lipid export ABC transport protein (*msbA*)
CP000538	267947	268054	5.00E-53	Chemotaxis protein CheA (*cheA*)
CP000538	339049	338934	5.00E-53	RND efflux system, inner membrane transporter (*cmeB*)
CP000538	339049	338934	5.00E-53	RND efflux system, inner membrane transporter (*cmeB*)
AF387299	7	127	2.00E-52	Putative integral membrane protein
AL111168	14747	14853	2.00E-52	Glutamate synthase small subunit (*gltD*)
AL111168	145443	145330	2.00E-52	Putative ABC transporter
CP000538	1498249	1498127	2.00E-52	Conserved domain protein
CP000538	531483	531377	2.00E-52	GatB/Yqey family protein
CP000538	185157	185263	2.00E-52	Transporter
CP000538	841649	841754	7.00E-52	Amino acid ABC tansporter
CP000538	419915	419806	7.00E-52	DNA polymerase III subunit epsilon
CP000538	667737	667842	7.00E-52	DNA polymerase III, alpha subunit (*dnaE*)
CP000538	933041	932936	7.00E-52	Porphobilinogen synthase (*hemB*)
CP000768	1832415	1832520	7.00E-52	3-isopropylmalate dehydratase, large subunit (*leuC*)
CP000549	41769	41881	8.00E-52	Tetracycline resistance gene (*tetO*)
CP000538	1263606	1263735	9.00E-52	1-deoxy-D-xylulose 5-phosphate reductoisomerase (*dxr*)
CP000538	1190525	1190405	9.00E-52	Low molecular weight phosphotyrosine protein phosphatase family
CP000538	1204980	1205109	9.00E-52	Tyrosyl-tRNA synthetase (*tyrS*)
CP000538	675765	675661	3.00E-51	Molybdenum cofactor biosynthesis protein (*mog*)
CP000538	84217	84089	4.00E-51	Flagellar motor switch protein (*fliY*)
AY330117	206	309	1.00E-50	DNA gyrase subunit B (*gyrB*)
CP000538	123002	123105	1.00E-50	ATP synthase F1, alpha subunit (*atpA*)
CP000538	346235	346125	1.00E-50	Conserved hypothetical protein
CP000538	346235	346125	1.00E-50	Conserved hypothetical protein
CP000538	346235	346125	1.00E-50	Conserved hypothetical protein
CP000538	7914	7811	1.00E-50	Na+/H+ antiporter family protein

CP000538	774644	774758	1.00E-50	tRNA pseudouridine synthase A (*truA*)
CP000538	44292	44394	4.00E-50	23S ribosomal RNA AL111168 (*rrlA*)
CP000538	385697	385595	4.00E-50	Putative GMC oxidoreductase subunit
CP000538	385697	385595	4.00E-50	Putative GMC oxidoreductase subunit
CP000025	183185	183055	2.00E-49	Conserved hypothetical protein
CP000538	126825	126716	2.00E-49	MotA/TolQ/ExbB proton channel family protein
CP000538	802224	802325	2.00E-49	Signal peptidase I (*lepB*)
CP000538	554078	553978	6.00E-49	Chemotaxis protein MotB, putative
CP000538	554078	553978	6.00E-49	Putative chemotaxis protein MotB
CP000538	1413419	1413519	6.00E-49	Sodium/proline permease (*putP*)
CP000538	410969	411069	6.00E-49	TenA/Thi-4 family protein
CP000538	24172	24065	7.00E-49	Methyl-accepting chemotaxis protein
CP000538	24172	24065	7.00E-49	Methyl-accepting chemotaxis protein
AL111168	1335881	1335995	8.00E-49	Putative triosephosphate isomerase (*tpiA*)
CP000538	1140999	1141123	8.00E-49	DedA family protein
CP000538	926620	926742	8.00E-49	Membrane protein
CP000538	579903	580018	9.00E-49	[NiFe] hydrogenase maturation protein (*hypF*)
AL111168	1337472	1337373	3.00E-48	Glyceraldehyde 3-phosphate dehydrogenase (*gapA*)
CP000538	530419	530539	3.00E-48	3,4-dihydroxy-2-butanone 4-phosphate synthase (*ribB*)
CP000538	530419	530539	3.00E-48	3,4-dihydroxy-2-butanone 4-phosphate synthase (*ribB*)
CP000538	355517	355616	3.00E-48	GTPase family protein
CP000538	24172	24065	3.00E-48	Methyl-accepting chemotaxis protein
CP000538	1264854	1264755	3.00E-48	Phosphatidate cytidylyltransferase (*cdsA*)
CP000538	1277793	1277892	3.00E-48	Polyphosphate kinase (*ppk*)
CP000538	1188103	1187981	3.00E-48	Putative isomerase
CP000538	1190500	1190401	4.00E-48	Low molecular weight phosphotyrosine protein phosphatase family
BX545860	7068	7185	1.00E-47	Putative methyltransferase
CP000025	1648716	1648818	1.00E-47	NADH dehydrogenase I chain D (*nuoD*)
CP000538	27981	28104	1.00E-47	RNA pseudouridine synthase family protein

CP000538	27981	28104	1.00E-47	RNA pseudouridine synthase family protein
CP000538	1014735	1014833	2.00E-47	Uroporphyrinogen-III synthase (*hemD*)
CP000538	448381	448284	4.00E-47	DNA-directed RNA polymerase, beta' subunit (*rpoC*)
CP000538	24172	24065	4.00E-47	Methyl-accepting chemotaxis protein
CP000768	1780736	1780635	4.00E-47	Conserved hypothetical protein
CP000768	1780736	1780635	4.00E-47	Conserved hypothetical protein
CP000768	1780736	1780635	4.00E-47	Conserved hypothetical protein
AF167344	1857	1761	1.00E-46	Putative UDP-glucose-4-epimerase
CP000538	922881	922977	1.00E-46	CjaA protein (*cjaA*)
CP000538	1114408	1114504	1.00E-46	Ubiquinol--cytochrome c reductase, cytochrome b subunit (*petB*)
AL111168	1335439	1335343	2.00E-46	Enoyl- [acyl-carrier-protein] reductase (*fabI*)
CP000538	257912	257804	2.00E-46	ATP-dependent Clp protease, ATP-binding subunit (*clpX*)
CP000538	1146400	1146304	2.00E-46	Peptidase, M23/M37 family
CP000538	402908	402798	2.00E-46	Phospho-N-acetylmuramoyl-pentapeptide–transferase (*mraY*)
CP000538	409993	410093	2.00E-46	Succinate dehydrogenase, C subunit (*sdhC*)
CP000538	84217	84121	3.00E-46	Flagellar motor switch protein (*fliY*)
CP000538	1528838	1528743	6.00E-46	L-serine ammonia-lyase (*sdaA*)
DQ174139	314	219	6.00E-46	16S ribosomal RNA (*rrsA*)
CP000538	1413419	1413522	7.00E-46	Sodium/proline permease (*putP*)
CP000768	176574	176681	7.00E-46	Iron ABC transporter
CP000025	211681	211572	3.00E-45	Tail tape measure protein, TP901 family
CP000538	770432	770331	3.00E-45	Phosphopantothenoylcysteine decarboxylase/phosphopantothenate--cysteine ligase (*coaBC*)
AL111168	831387	831480	9.00E-45	D-3-phosphoglycerate dehydrogenase (*serA*)
CP000538	4919	4822	9.00E-45	DNA gyrase, B subunit (*gyrB*)
CP000538	458602	458699	9.00E-45	Putative periplasmic protein
CP000538	822020	821927	1.00E-44	Cell division protein FtsK, putative
CP000538	1094669	1094778	1.00E-44	Copper-translocating P-type ATPase
CP000538	24172	24065	4.00E-44	Methyl-accepting chemotaxis protein
CP000538	530419	530539	5.00E-44	3,4-dihydroxy-2-butanone 4-phosphate synthase (*ribB*)

AY681275	1023	931	6.00E-44	Cytochrome c family protein
DQ140271	1883	1784	1.00E-43	UDP-GlcNAc/Glc 4-epimerase
CP000538	296556	296647	2.00E-43	1-deoxy-D-xylulose-5-phosphate synthase (*dxs*)
CP000538	554181	554272	2.00E-43	Chemotaxis protein MotB
AL111168	1133069	1132975	6.00E-43	5,10-methylenetetrahydrofolate reductase (*metF*)
CP000768	1780736	1780635	6.00E-43	Conserved hypothetical protein
CP000538	1028222	1028113	7.00E-43	Protein-export membrane protein SecF (*secF*)
CP000538	741897	741782	7.00E-43	Putative aminotransferase
CP000538	535230	535341	7.00E-43	S-adenosylmethionine:tRNA ribosyltransferase-isomerase(*queA*)
CP000538	535230	535341	7.00E-43	S-adenosylmethionine:tRNA ribosyltransferase-isomerase (*queA*)
CP000538	221318	221229	2.00E-42	Ribonucleoside-diphosphate reductase, beta subunit (*nrdB*)
CP000538	147831	147943	3.00E-42	Conserved hypothetical protein
CP000538	348795	348678	1.00E-41	ATPase, AAA family protein
CP000538	442651	442738	3.00E-41	DNA-directed RNA polymerase, beta subunit (*rpoB*)
CP000538	442651	442738	3.00E-41	DNA-directed RNA polymerase, beta subunit (*rpoB*)
AY681249	350	459	4.00E-41	Conserved hypothetical protein
DQ518908	8640	8529	4.00E-41	Putative UDP-N-acetylglucosamine 2-epimerase (*neuC1*)

*BLASTN available from http://blast.ncbi.nlm.nih.gov. MOSC, molybdenum cofactor sulfurase C; ABC, ATP binding cassette; SAM, S-adenosylmethinione; GMC, glucose–methanol–choline.

†Expect (E)–value cutoff 10^{-40}.