

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

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

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, $\approx 20\%$ 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

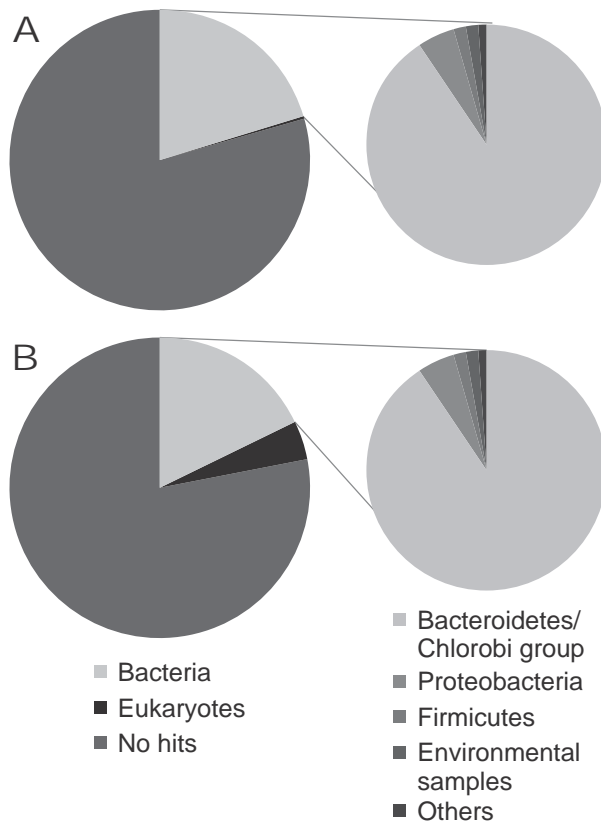


Figure. Comparison of the organisms from which the best matches for the sequences were derived from a BLASTN (http://blast.ncbi.nlm.nih.gov) search with an expect-value cutoff of 10^{-5} . A) DNA from nondiarrheic fecal sample collected 3 months after patient had recovered. B) DNA from diarrheic fecal 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‡
<i>Bacteroides vulgatus</i>	5,944 (50.5)	4,743 (56.5)
<i>Homo sapiens</i>	2,955 (25.1)	84 (1.0)
<i>Parabacteroides distasonis</i>	818 (6.9)	1,283 (15.3)
<i>B. thetaiotaomicron</i>	767 (6.5)	1,046 (12.5)
<i>B. fragilis</i>	759 (6.4)	842 (10.0)
Uncultured bacterium	195 (1.7)	227 (2.7)
<i>Campylobacter jejuni</i>	156 (1.3)	0
<i>B. ovatus</i>	48 (0.4)	63 (0.8)
Uncultured <i>Bacteroides</i> spp.	20 (0.2)	19 (0.2)
<i>B. uniformis</i>	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

- Service RF. Gene sequencing: the race for the \$1000 genome. *Science*. 2006;311:1544–6. DOI: 10.1126/science.311.5767.1544
- 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.
- 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.
- 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.
- Penner JL. The genus *Campylobacter*: a decade of progress. *Clin Microbiol Rev*. 1988;1:157–72.
- Young KT, Davis LM, DiRita VJ. *Campylobacter jejuni*: molecular biology and pathogenesis. *Nat Rev Microbiol*. 2007;5:665–79. DOI: 10.1038/nrmicro1718
- 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.
- Altekruse SF, Stern NJ, Fields PI, Swerdlow DL. *Campylobacter jejuni*—an emerging foodborne pathogen. *Emerg Infect Dis*. 1999;5:28–35.
- Morens DM, Folkers GK, Fauci AS. The challenge of emerging and re-emerging infectious diseases. *Nature*. 2004;430:242–9. DOI: 10.1038/nature02759
- 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
- 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
- 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
- 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
- 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
- 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

EMERGING INFECTIOUS DISEASES *online*

www.cdc.gov/eid

To receive tables of contents of new issues send an email to listserv@cdc.gov with `subscribe eid-toc` in the body of your message.