Comparative Community Genomics of the Gut Microbiota

M. Pilar Francino[1], Jenna Morgan[1], Parag Vaishampayan[1], Jeff Froula[1], Yogesh Shouche[2] and Howard Ochman[3]

1: Evolutionary Genomics Department, DOE Joint Genome Institute, Walnut Creek, CA
2: National Centre for Cell Science, University of Pune, India
3: Departments of Biochemistry & Molecular Biophysics, and Ecology & Evolutionary Biology, University of Arizona, Tucson, Arizona

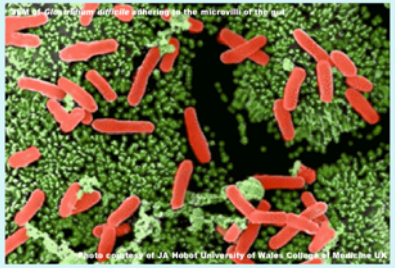
DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

Introduction

The global aim of this project is to increase our understanding of the diverse microbial community that inhabits the gastrointestinal (GI) tract. These microbes play essential roles in health, including a significant contribution to the digestive process, promotion of gut maturation, and integrity and modulation of the immune system (Berg, 1996).

Bacteria living in complex natural communities, such as the GI microbiota, produce antimicrobial compounds involved in intra- and interspecific competition. The GI microbial community is likely to be a major reservoir of antibiotic resistance genes; GI bacteria are frequently challenged by antibacterial compounds produced by GI community members, by incoming bacteria, and by medically prescribed antibiotics.

Moreover, GI microbes interact with pathogenic agents in several complex ways. On one hand, resident bacteria exert a protective barrier effect against enteropathogens (Hudault *et al.*, 2001); but on the other, they have the potential to enrich the arsenal of incoming pathogens through horizontal transmission of genes involved in host-microbe interaction or antibiotic resistance. Elucidating the composition and coding capabilities of the GI microbiota is therefore crucial for a comprehensive analysis of infectious disease.



Clostridium difficile is a spore forming bacteria which can be part of the normal intestinal flora in as many as 50% of children under age two, and less frequently in individuals over two years of age. *C. difficile* is the major cause of pseudomembranous colitis and antibiotic associated diarrhea.

Characterization of fosmid metagenomic libraries

Estimation of Biodiversity

From the GI bacterial DNA preparations, we have produced 2 fosmid libraries, from adult and infant (1 month old), each containing 50,000 to 70,000 clones (40 kb insert size). We are currently in the process of characterizing the biodiversity of these libraries. Our initial approach will include the production of 16s rRNA PCR libraries from pooled fosmids. In addition, we have also prepared 16s PCR libraries from the DNA agarose plugs from which the fosmid libraries were prepared in order to evaluate any potential biases in fosmid library construction (Figs. 1 and 3).

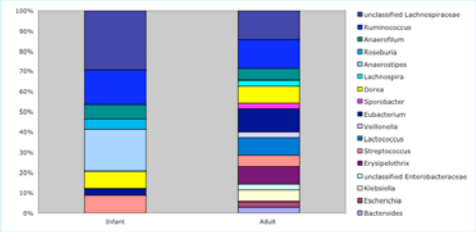


Fig. 1 96 clones were sequenced from each library. These clones were processed using the 16S pipeline created for the JGI by Ed Kirton, which assembles paired ends, BLASTs high-quality contigs to confirm that they are in fact 16S rRNA, and checks for chimeric clones using the Bellerophon software (Huber *et al.*, 2004). From the infant library, 72 clones were passed by the pipeline, and 44 adult clones were passed. The diversity recovered (at the generic level), determined by the "Classify" tool at greengenes.lbl.gov is depicted here.

Genetic Screening

With the aim of phylogenetically typing a large number of fosmid clones, we have initiated screenings of the fosmid libraries with 16s rRNA primers as well as a set of 11 primer sets directed towards universally conserved bacterial proteins (Santos and Ochman, 2004). Screening our libraries for 16S rRNA genes will render our analysis directly comparable to previous studies that have characterized the GI microbiota by sequencing PCR-amplified and cloned copies of this gene.

We will be further characterizing these libraries by screening for different types of genes involved in microbe-host interactions. We will target representatives of the following functions: capsular polysaccharide biosynthesis, porins, fimbrial adhesins, secreted proteases, cytotoxins and type III secretion systems. We have overcome our first major technical hurdle by optimizing our PCR-based screening strategy to avoid amplification of the host *E. coli* genes (Fig. 2).

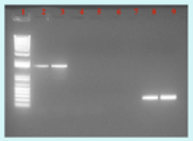


Fig. 2 Lane 1: 1kb plus ladder.
Lane 2: PCR product generated using universal 16S primers. Fosmid clones from 5 plates (384 wells) were pooled and then prepped using Qiagen's Qiafilter 96 plates, followed by an overnight incubation with Epicentre's Plasmid-Safe™ ATP-Dependent DNase. Two clones in this pool are known to contain the 16S sequence.
Lane 3: Same as Lane 2, but the pool of fosmid clones was incubated with Induction Solution before being prepped.
Lanes 4 and 5: No PCR product is generated when the template is a prepped fosmid clone from a Stalk-Eyed Fly (Diopsid) library that is know to NOT contain any 16S sequence.
Lanes 6 and 7: No PCR product is generated using primers designed for the Diopsid clone with the pooled bacterial fosmid clones.Lanes 8 and 9: PCR product generated using the diopsid primers with the diopsid fosmid used as template.

16S rRNA tree

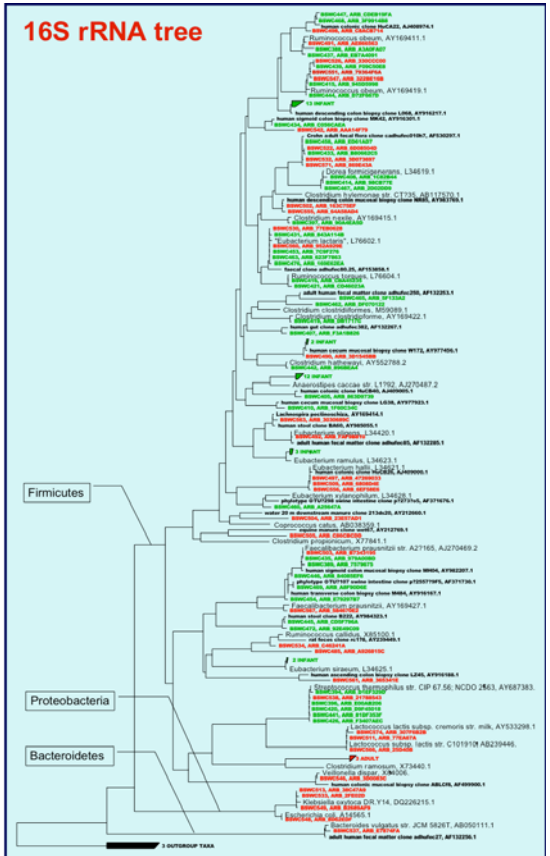


Fig. 3 A maximum likelihood tree reconstructed from an alignment of 16S sequences in our PCR library in addition to selected sequences from the 84218 aligned 16S rDNA records >1250nt maintained at http://greengenes.lbl.gov. The 72 infant sequences are shown in green, and the 44 adult sequences are shown in red. Despite the

Preparation of high molecular weight GI bacterial DNA from fecal samples

Human fecal samples from two healthy anonymous volunteers (mother and infant) were collected at the University of Arizona in collaboration with the laboratory of Howard Ochman. High molecular weight DNA for large insert library construction was prepared from these samples using several filtration and centrifugation steps to eliminate particulate matter and larger eukaryotic cells. Inspection of cells under the microscope revealed a large variety of bacterial types and no traces of eukaryotic cells. Moreover, PCR and Southern hybridization with primers and probes directed to human genes did not produce positive results, indicating that our preparations were free of human DNA contamination.

Future Work

Genes cloned in fosmid vectors can be heterologously expressed in *E. coli*, allowing for further characterization of their function (Blodgett *et al.*, 2005). Therefore, genes conferring important capabilities in the GI ecosystem, such as antibiotic production and resistance, can be phenotypically detected. We will screen our fosmid libraries for antibiotic production phenotypes by overlaying agar plates containing arrayed clones with top agar seeded with exponentially growing bacterial cultures and looking for growth inhibition halos. We will also screen for resistance to several classes of antibiotics, with different modes of action, by replicating our libraries on plates containing appropriate concentrations of each antibiotic.

We also have plans to study succession and development of the gut microbiota of the new born infant as a collaborative effort with a research institution in India. To clarify the initial acquisition and subsequent colonization of bacteria in an infant within the first few months after birth, phylogenetic analysis will be performed using 16S rRNA sequences amplified from DNA isolated from the feces at various stages of succession. In addition, fosmid libraries will be constructed from these DNA samples, and comparative analyses will be performed between the American and Indian microbial succession data.

References

Berg RD (1996) The indigenous gastrointestinal microflora. Trends Microbiol 4, 430-435.
Blodgett JA, Zhang JK, Metcalf WW (2005) Molecular cloning, sequence analysis, and heterologous expression of the phosphinothricin tripeptide biosynthetic gene cluster from Streptomyces viridochromogenes DSM 40736. Antimicrob Agents Chemother. 49(1), 230
Huber T, Faulkner F, and Hugenholtz, P (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics 20(14), 2317-2319.
Hudault S, Guignot J & Servin AL (2001) Escherichia coli strains colonising the gastrointestinal tract protect germfree mice against Salmonella typhimurium infection. Gut 49, 47-55.
Santos SR & Ochman H (2004) Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins. Environmental Microbiology 6, 754-759.

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under contract No. W-7405-ENG-36.