17. V. L. Schramm, W. Shi, *Curr. Opin. Struct. Biol.* **11**, 657 (2001).
18. Y. F. Li *et al.*, *J. Biol. Chem.* **273**, 15035 (1998).
19. I. S. Ridder, H. J. Rozeboom, K. H. Kalk, B. W. Dijkstra, *J. Biol. Chem.* **274**, 30672 (1999).
20. J. Allard, P. Grochulski, J. Sygusch, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3679 (2001).
21. H. Kack, K. J. Gibson, Y. Lindqvist, G. Schneider, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5495 (1998).
22. A. Heine *et al.*, *Science* **294**, 369 (2001).
23. I. Schlichting *et al.*, *Science* **287**, 1615 (2000).
24. M. J. Brubaker, D. H. Dyer, B. Stoddard, D. E. Koshland Jr., *Biochemistry* **35**, 2854 (1996).
25. R. D. Allison, D. L. Purich, *Methods Enzymol.* **354**, 455 (2002).
26. N. Qian, G. A. Stanley, B. Hahn-Hagerdal, P. Radstrom, *J. Bacteriol.* **176**, 5304 (1994).
27. R. E. Hanes Jr., V. M. Lynch, E. V. Anslyn, K. N. Dalby, *Org. Lett.* **4**, 201 (2001).
28. J. B. Dai, Y. Liu, W. J. Ray Jr., M. Konno, *J. Biol. Chem.* **267**, 6322 (1992).
29. N. Qian, G. A. Stanley, A. Bunte, P. Radstrom, *Microbiology* **143 (Pt 3)**, 855 (1997).
30. L. E. Naught, P. A. Tipton, *Arch. Biochem. Biophys.* **396**, 111 (2001).
31. C. Regni, P. A. Tipton, L. J. Beamer, *Structure (Cambridge)* **10**, 269 (2002).
32. S. D. Lahiri, G. Zhang, D. Dunaway-Mariano, K. N. Allen, *Biochemistry* **41**, 8351 (2002).
33. S. Muller *et al.*, *J. Mol. Biol.* **315**, 141 (2002).
34. In earlier work, β-PGM was isolated from the cell in its phosphorylated form and crystallized in the presence of its Mg(II) cofactor for structure determination (*32*). Overall, β-PGM is a monomer that is composed of two distinct domains: a helical cap domain and an α/β core domain. Whereas the conformations of individual domains of the phosphorylated enzyme-Mg(II) complex and the enzyme-Mg(II)-β-G1,6-P complexes are nearly indistinguishable, the positions of the cap domains relative to the core domains are noticeably different. A hinge-like motion between the two domains, acting as rigid bodies, produces the conversion of one enzyme conformer to the other, enveloping the β-G1,6-P intermediate.
35. Materials and methods are available as supporting material on *Science* Online.
36. G. M. Sheldrick, T. R. Schneider, *Methods Enzymol.* **277B**, 319 (1997).
37. D. E. C. C. Corbridge, *The Structural Chemistry of Phosphorus* (Elsevier, Amsterdam, 1974).
38. R. R. Holmes, *Pentacoordinated Phosphorous Vol. I* (ACS Monograph 175, Washington, DC, 1980), vol. 1.
39. L. Pauling, in *The Nature of the Chemical Bond* (Cornell Univ. Press, Ithaca, NY, 1960) pp. 255–260.
40. A. S. Mildvan, *Proteins* **29**, 401 (1997).
41. G. Kemp, S. Trippett, *Tetrahedron Lett.* **48**, 4381 (1976).
42. F. H. Westheimer, *Acc. Chem. Res.* **1**, 70 (1968).
43. F. Ramirez, J. F. Marecek, *J. Am. Chem. Soc.* **101**, 1460 (1979).
44. J. R. Knowles, *Annu. Rev. Biochem.* **49**, 877 (1980).
45. Hydrogen bonds are made between the C(6)phosphate and the N118 side chain NH2, S116 side chain OH, K117 amide NH, the two R49 guanidinium NHs, Wat 231, Wat 48, Wat5, Wat129; hexose C(4)O and Wat129 and V47 amide NH; hexose C(3)O and Wat68; C(2)O and Wats68 and 29; and hexose C(5)O and S116 amide NH.
46. For this analysis, the model of the ligand β-G1,6-P was refined in a crystallography and nuclear magnetic resonance system with geometric constraints for normal tetrahedral hybridization at phosphorus during the initial stages of refinement of the 1.2 Å structure. The model of the ligand was docked into the active site by overlaying the hexose ring, C(6)phosphate, and C(1)phosphate phosphorus portion of the molecule with that of the pentavalent intermediate.
47. T. C. Bruice, F. C. Lightstone, *Acc. Chem. Res.* **32**, 127 (1999).
48. A. L. Morris, M. W. MacArthur, E. G. Hutchinson, J. M. Thornton, *Proteins* **12**, 345 (1992).
49. Supported by a grant from NIH (grant GM16099) (to K.N.A and D.D-.M). Use of the Advanced Photon Source

Role of Mobile DNA in the Evolution of Vancomycin-Resistant *Enterococcus faecalis*

I. T. Paulsen,[1,2]* L. Banerjei,[1] G. S. A. Myers,[1] K. E. Nelson,[1] R. Seshadri,[1] T. D. Read,[1] D. E. Fouts,[1] J. A. Eisen,[1,2] S. R. Gill,[1] J. F. Heidelberg,[1] H. Tettelin,[1] R. J. Dodson,[1] L. Umayam,[1] L. Brinkac,[1] M. Beanan,[1] S. Daugherty,[1] R. T. DeBoy,[1] S. Durkin,[1] J. Kolonay,[1] R. Madupu,[1] W. Nelson,[1] J. Vamathevan,[1] B. Tran,[1] J. Upton,[1] T. Hansen,[1] J. Shetty,[1] H. Khouri,[1] T. Utterback,[1] D. Radune,[1] K. A. Ketchum,[1]† B. A. Dougherty,[1]‡ C. M. Fraser[1,3]

The complete genome sequence of *Enterococcus faecalis* V583, a vancomycin-resistant clinical isolate, revealed that more than a quarter of the genome consists of probable mobile or foreign DNA. One of the predicted mobile elements is a previously unknown *vanB* vancomycin-resistance conjugative transposon. Three plasmids were identified, including two pheromone-sensing conjugative plasmids, one encoding a previously undescribed pheromone inhibitor. The apparent propensity for the incorporation of mobile elements probably contributed to the rapid acquisition and dissemination of drug resistance in the enterococci.

The Gram-positive bacterium *Enterococcus faecalis* is a natural inhabitant of the mammalian gastrointestinal tract and is commonly found in soil, sewage, water, and food, frequently through fecal contamination (*1*). *E. faecalis* can withstand oxidative stress, desiccation, and extremes of temperature and pH, and it has high endogenous resistance to salinity, bile acids, detergents, and antimicrobials (*1*).

E. faecalis is an opportunistic pathogen that is a major cause of urinary tract infections, bacteremia, and infective endocarditis (*2*). The intrinsic resistance of *E. faecalis* to many antibiotics and its acquisition of resistance to other antimicrobial agents, particularly vancomycin, which is used to treat serious infections by drug-resistant Gram-positive pathogens, has led to the emergence of *E. faecalis* as a nosocomial pathogen that is refractory to most therapeutic options (*3*). Recent reports of the long-predicted emergence of vancomycin-resistant *Staphylococcus aureus* clinical isolates from transfer of enterococcal genes is a serious health care concern (*4*). Here we report the complete genome sequence of *E. faecalis* strain V583 (*5*), the first vancomycin-resistant clinical isolate reported in the United States (*6*). The genome sequence provides insight into the pathogenesis and biology of *E. faecalis*, the role of mobile elements in genome evolution, and the transfer of vancomycin resistance.

A total of 3337 predicted protein-encoding open reading frames (ORFs) were identified on the chromosome and three plasmids of *E. faecalis* V583 (Table 1; fig. S1) (*7*). Over a quarter of the *E. faecalis* V583 genome consists of mobile and/or exogenously acquired DNA, including seven probable integrated phage regions, 38 insertion elements (IS), multiple conjugative and composite

[1]The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA. [2]Johns Hopkins University, Charles and 34th Streets, Baltimore, MD 21218, USA. [3]The George Washington University School of Medicine, Departments of Pharmacology and Microbiology and Tropical Medicine, 2300 Eye Street NW, Washington, DC 20037, USA.

*To whom correspondence should be addressed. E-mail: ipaulsen@tigr.org
†Present address: Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA.
‡Present address: Bristol-Myers Squibb PRI, 5 Research Parkway, Wallingford, CT 06492, USA.

transposons, a putative pathogenicity island, and integrated plasmid genes. To our knowledge, this represents one of the highest proportions of mobile elements observed in a bacterial genome. The plethora of mobile elements probably contributed to the accumulation of virulence and drug resistance factors by *E. faecalis*.

Vancomycin resistance in *E. faecalis* V583 appears to be encoded within a previously unknown mobile element (EF2282-EF2334) with some similarities to the probable *E. faecalis vanB* vancomycin-resistance conjugative transposon Tn*1549* (*8*). The vancomycin-resistance genes (EF1955-EF1963) encode vancomycin resistance via synthesis of modified peptidoglycan precursors terminating in D-lactate (*9*), and they are essentially identical to the *vanB* genes from Tn*1549*. The remainder of the element is very divergent from Tn*1549* with multiple insertions, deletions, and rearrangements (Fig. 1); relatively low sequence similarity between conserved genes; and a different recombinational system (EF2283). Even though *E. faecalis* V583 is the earliest known vancomycin-resistant clinical isolate from the United States, the conjugative transposon-like features and atypical trinucleotide of this element indicate it was likely obtained as a cassette by lateral gene transfer. It is also flanked by Tn*916*-like genes (Fig. 1), which may have played a role in the acquisition of this element.

Highly similar Tn*916*-like genes are also found in association with a locus (EF1869-EF1863) encoding homologs of the *Streptococcus pneumoniae* VncRS two-component signal transduction system and Vex secretion proteins (Fig. 1). The *vncRS* locus has been associated with vancomycin tolerance in *S. pneumoniae* via mutation of *vncS* (*10*), although recent evidence has cast doubt on this association (*11*). In *E. faecalis* V583, the *vncS* gene (EF1866) is disrupted by a nonfunctional ISL3 family insertion sequence. The possible role of this element in vancomycin tolerance in *E. faecalis* is unclear, but it is flanked by copies of IS*256* and may also have been laterally acquired.

Thirty-eight IS elements were identified (table S1), with three types predominating: ISEf1, IS256, and IS1216. There are two clusters of IS elements on the chromosome (fig. S1). One is associated with a pathogenicity island and integrated plasmid genes. The second cluster includes several types of IS elements that flank a region of atypical trinucleotide composition (EF1860-EF1858) that may have been acquired by lateral gene transfer [Supporting Online Material (SOM) Text] and encodes three of the four steps of pantothenate biosynthesis.

A large pathogenicity island has previously been identified in *E. faecalis* V583 (*12*) (EF0479-EF0628), including genes for aggregation substance, cytolysin, and other possible virulence or adaptation genes. Trinucleotide composition analysis indicated that most of this island has highly atypical composition, except for a region containing integrated plasmid genes from a pTEF1-like plasmid (fig. S1). The presence of multiple IS elements and the integrated plasmid genes hints at a complex evolutionary history for this element. It is flanked at one end by an integrase gene, possibly responsible for integration of this element.

Table 1. General features of the *E. faecalis* genome. No., number; rRNA, ribosomal RNA; tRNA, transfer RNA.

|	Chromosome	pTEF1	pTEF2	pTEF3
Size (base pairs)	3218031	66320	57660	17963
G+C content (%)	37.5	34.4	33.9	33.3
Protein-coding genes				
No. similar to known proteins	1760	37	22	10
No. with unknown function	221	2	1	1
No. of conserved hypotheticals	495	18	10	3
No. with no database match	706	17	29	5
Total	3182	74	62	19
Average ORF size (base pairs)	889	743	816	721
Coding (%)	87.9	83.0	87.8	76.4
rRNA genes	12	0	0	0
tRNA genes	68	0	0	0
Structural RNA	2	0	0	0
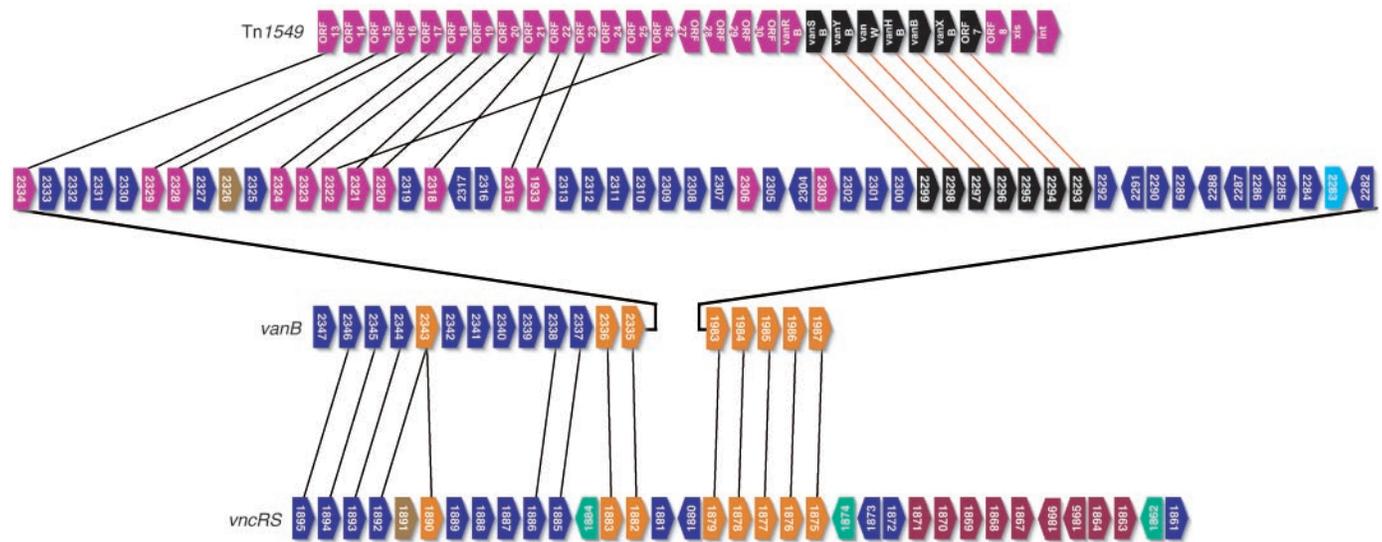
Fig. 1. Linear representation of the *E. faecalis* V583 *vanB* vancomycin-resistance gene region and its relationship with the vancomycin-resistance transposon Tn*1549* (*8*) and the *E. faecalis* V583 *vncRS/vex* locus. Genes are shown as arrowheads (not to scale) colored by predicted function or transposable element: black, vanB vancomycin-resistance genes; magenta, Tn*1549*-like genes; orange, Tn*916*-like genes; green, transposases; red, *vncRS/vex* locus; brown, group II intron; light blue, transposon resolvase. Black lines connect best matches (BLAST P-value < 1 × 10^{-5}), and red lines connect best matches with greater than 99% identity. Genes are labeled by name or by the appropriate ORF or locus numbers.

There are seven regions derived from probable integrated phage (fig. S1). These putative prophage are most closely related to phage from other low-GC Gram-positive bacteria. The integrated phage regions encode multiple homologs of *Streptococcus mitis* PblA and PblB, which have been implicated in binding human platelets, an interaction important in the pathogenesis of infective endocarditis (*13*). A ferrochetalase gene (EF1989) encoded within one of the phage regions may allow *E. faecalis* to utilize coproporphyrinogen III for heme synthesis (SOM Text).

A variety of diverse plasmids have been previously described in *E. faecalis*, particularly conjugative plasmids that encode a mating response to sex pheromone peptides secreted by plasmid-free recipient strains (*14*). Three plasmids are present in *E. faecalis* V583 (Table 1; fig. S2): pTEF1 and pTEF2 are structurally similar to the archetypal pheromone responsive plasmids pAD1 (*15*) and pCF10 (*14*), respectively, and pTEF3 belongs to the family of pAMβ1 broad host range plasmids.

The sex pheromone inhibitor (iAD1) and surface aggregation substance (Asa1) encoded by pTEF1 are identical to those of pAD1, and both plasmids share extensive regions of sequence similarity (fig. S2). There is a 31-kb inversion in pTEF1 relative to pAD1 that probably affects regulation of conjugation in *E. faecalis* V583 (SOM Text). The unique regions in pTEF1 include a Tn*4001*-like transposon encoding aminoglycoside resistance and another IS-flanked element carrying erythromycin resistance and multidrug resistance genes.

pTEF2 and the sex pheromone plasmid pCF10 share regions of similarity, including identical copies of the conjugation genes *prgA-prgB-prgC*, but pTEF2 lacks the pCF10 pheromone inhibitor *prgQ* gene, encoding a previously undescribed predicted pheromone inhibitor gene (EFB0005.1) in the equivalent position. pTEF3 is a nonconjugative plasmid but has acquired a pTEF2-like *prgZ* pheromone receptor, whose gene is adjacent to multiple IS elements. The occurrence of a novel pheromone inhibitor on pTEF2 suggests the possible occurrence of a broad diversity of different *E. faecalis* pheromones and pheromone inhibitors in nature. Five sex pheromones encoded within lipoprotein signal peptides have been identified in the genome of *E. faecalis* V583 (*16*). An additional 76 predicted lipoproteins were identified (table S2), some of which could represent previously unknown pheromone precursors.

The chromosome of *E. faecalis* contains at least three integrated plasmid remnants. Two of these regions encode homologs of aggregation substance (EF0485, EF0149) that may be important for virulence (fig.

S3). Plasmid-encoded aggregation substance is a surface protein that enhances conjugative transfer and has been implicated in adhesion to colonic mucosal fibronectin and in translocation across intestinal epithelium (*17*). One of the integrated plasmid regions (EF0506-EF0485) is located within the probable pathogenicity island (*12*); the other two regions include genes of plasmid, phage, and conjugative transposon origin (fig. S3; SOM Text) and encode lipoproteins (EF0164 and EF2512) whose signal peptides resemble those of known pheromone precursors. The presence of multiple integrated plasmid remnants and three resident plasmids in *E. faecalis* V583 emphasizes the importance of plasmids in genome plasticity of the enterococci.

Comparison of the predicted protein set of *E. faecalis* with those of other sequenced genomes confirmed the relationship of *E. faecalis* within the low-GC Gram-positive bacteria. Over 85% of *E. faecalis* ORFs with significant database matches had their best match to other sequenced low-GC Gram-positive organisms, and 519 *E. faecalis* ORFs were conserved in a set of 10 sequenced low-GC Gram-positive organisms (table S3) (*5*). The distribution of these conserved genes in the *E. faecalis* genome revealed a number of regions with a low abundance of conserved genes (fig. S1), largely corresponding to the identified phage, integrated plasmid, pathogenicity island, and conjugative transposon regions. This set of conserved genes is involved in essential processes such as transcription, translation, and protein synthesis but also includes a considerable number of proteins from large paralogous families such as the PTS and ABC transporter families.

There is essentially no large-scale gene synteny between the *E. faecalis* genome and that of any sequenced low-GC Gram-positive bacterium. The multiplicity of mobile or foreign elements such as phage and IS elements suggests the *E. faecalis* genome is highly malleable and may have undergone multiple rearrangement events, explaining the lack of gene synteny. Despite this lack of synteny, there is a strong transcriptional skew (fig. S1) as found in other sequenced low-GC Gram-positive bacteria. Genes encoded within the mobile elements also show a transcriptional skew; within the phage regions, for instance, 90% of the ORFs align with the direction of replication. Strong selective pressure for genes to be transcribed in the direction of replication appears to be a common feature of the low-GC Gram-positive bacteria.

Analysis of the transport and metabolic capabilities of *E. faecalis* V583 emphasizes the importance of fermentation of nonabsorbed sugars in the gastrointestinal tract in its life-style. *E. faecalis* has 35 probable PTS-

type sugar transporters, comparable to *Listeria* species and considerably more than any other sequenced organism, as well as ABC-type and other sugar uptake systems (table S4). Consistent with these transport capabilities, *E. faecalis* encodes pathways for the utilization of more than 15 different sugars (table S4). Energy production from these substrates occurs via glycolysis or the pentose phosphate pathways, with the trichloroacetic acid (TCA) cycle absent.

E. faecalis is one of the few bacteria that are substantial producers of extracellular O_2^- (*18*), and an array of oxidative stress resistance mechanisms are evident from its genome (table S5). *E. faecalis* is well endowed with cation homeostasis mechanisms (table S5), which probably contribute to its pH, salt, metal, and desiccation resistance, including 14 predicted metal ion P-type ATPases, more than any other currently sequenced bacterium. Despite its stress resistance capabilities, *E. faecalis* possesses a modest collection of regulatory genes (table S6), including only three alternate sigma factors.

E. faecalis is known to adhere to a variety of host cells or extracellular matrix components. Adhesins such as aggregation substance, MSCRAMM, hemagglutinin, and other virulence factors have been described for *Enterococcus* spp. (*1*). A comprehensive genome-wide exploration for signal sequences, lipoprotein motifs, and other potential host cell components or binding motifs yielded an additional 134 putative surface-exposed proteins that may be associated with early colonization stages or virulence (table S7).

Forty-seven candidates with potential choline- or integrin-binding motifs were identified that may play a role in adherence and internalization processes (table S7). A family of four probable adhesin lipoproteins was identified, including a characterized endocarditis-specific antigen (EF2076). A paralogous family of 21 LPxTG-motif cell wall surface anchor proteins is present, as are three corresponding sortases (EF3056, EF1094, and EF2524). Several of the genes for LPxTG proteins have an atypical nucleotide composition, suggesting a foreign origin. Translocation across the intestinal epithelium (*19*) may be facilitated by homologs of *L. monocytogenes* internalin A (EF2686), an *S. aureus* exfoliative toxin (EF0645), and probable secreted hyaluronidases (EF3023 and EF0818) and a protease (EF1818).

Antigenic or phase variation of surface structures is a common immune evasion tactic amongst pathogens and has been implicated in activating conjugation function in *E. faecalis* (*20*). Out of 134 putative surface-exposed proteins, 65 were found to contain stretches of homopolymeric sequence or iterative nucleotide motifs located within the predicted ORF or promoter region, which

may enable phase variation via a slippage type mechanism (table S7).

An unprecedented amount of the *E. faecalis* V583 genome consists of intact or partial mobile elements. Many of these regions have complex mosaic structures comprised of different elements, suggesting they are "hotspots" or "graveyards" for mobile element insertion. This apparent propensity for the incorporation of mobile elements probably contributed to the rapid acquisition and dissemination of drug resistance in the enterococci and suggests that they act as a reservoir for the further dissemination of drug resistance traits such as vancomycin resistance via mobile elements and/or conjugative plasmids. The complete genome sequence of *E. faecalis* V583 has enabled the identification of numerous predicted virulence factors and surface-exposed proteins that may facilitate the development of therapeutic approaches to combat this important nosocomial pathogen.

References and Notes
1. I. Klare, G. Werner, W. Witte, *Contrib. Microbiol.* **8**, 108 (2001).
2. S. A. Tailor, E. M. Bailey, M. J. Rybak, *Ann. Pharmacother.* **27**, 1231 (1993).
3. M. M. Huycke, D. F. Sahm, M. S. Gilmore, *Emerg. Infect. Dis.* **4**, 239 (1998).
4. CDC, *Morb. Mortal. Wkly. Rep.* **51**, 902 (2002).
5. Materials and methods are available as supporting material on *Science* Online.
6. D. F. Sahm et al., *Antimicrob. Agents Chemother.* **33**, 1588 (1989).
7. GenBank accession numbers of the chromosome and plasmids are as follows: AE016830 (chromosome), AE016833 (pTEF1), AE016831 (pTEF2), AE016832 (pTEF3).
8. F. Garnier, S. Taourit, P. Glaser, P. Courvalin, M. Galimand, *Microbiology* **146 (no. 6)**, 1481 (2000).
9. M. Arthur et al., *Antimicrob. Agents Chemother.* **36**, 867 (1992).
10. R. Novak, B. Henriques, E. Charpentier, S. Normark, E. Tuomanen, *Nature* **399**, 590 (1999).
11. G. T. Robertson et al., *J. Bacteriol.* **184**, 6987 (2002).
12. N. Shankar, A. S. Baghdayan, M. S. Gilmore, *Nature* **417**, 746 (2002).
13. B. A. Bensing, I. R. Siboo, P. M. Sullam, *Infect. Immun.* **69**, 6186 (2001).
14. G. M. Dunny, B. A. Leonard, *Annu. Rev. Microbiol.* **51**, 527 (1997).
15. M. V. Francia et al., *Plasmid* **46**, 117 (2001).
16. D. B. Clewell, F. Y. An, S. E. Flannagan, M. Antiporta, G. M. Dunny, *Mol. Microbiol.* **35**, 246 (2000).
17. S. D. Sussmuth et al., *Infect. Immun.* **68**, 4900 (2000).
18. M. M. Huycke, V. Abrams, D. R. Moore, *Carcinogenesis* **23**, 529 (2002).
19. C. L. Wells, R. P. Jechorek, S. L. Erlandsen, *J. Infect. Dis.* **162**, 82 (1990).
20. L. T. Pontius, D. B. Clewell, *Plasmid* **26**, 172 (1991).
21. This work was supported by the NIH, National Institute of Allergy and Infectious Disease Grant AI40963-02. We thank M. Heaney, S. Lo, M. Holmes, B. Lee, R. Karamchedu, and V. Sapiro for database and information technology support at TIGR, and we thank the TIGR faculty and sequencing core for expert advice and assistance. We thank M. Gilmore for generously providing *E. faecalis* V583. We thank D. Clewell for providing the complete sequence of pAD1 for comparative analysis.

A Genomic View of the Human–*Bacteroides thetaiotaomicron* Symbiosis

Jian Xu, Magnus K. Bjursell, Jason Himrod, Su Deng, Lynn K. Carmichael, Herbert C. Chiang, Lora V. Hooper, Jeffrey I. Gordon*

The human gut is colonized with a vast community of indigenous microorganisms that help shape our biology. Here, we present the complete genome sequence of the Gram-negative anaerobe *Bacteroides thetaiotaomicron*, a dominant member of our normal distal intestinal microbiota. Its 4779-member proteome includes an elaborate apparatus for acquiring and hydrolyzing otherwise indigestible dietary polysaccharides and an associated environment-sensing system consisting of a large repertoire of extracytoplasmic function sigma factors and one- and two-component signal transduction systems. These and other expanded paralogous groups shed light on the molecular mechanisms underlying symbiotic host-bacterial relationships in our intestine.

A major theme of life on our planet is the complex and beneficial interactions that occur between eukaryotes and prokaryotes. Humans are no exception. As adults, we harbor diverse communities of microorganisms whose total number exceeds the sum of all of our somatic and germ cells (*1*). As yet, the ways in which these communities contribute to normal postnatal development and adult physiology are largely unexplored. The human gut contains the largest such collection of microbes [10^{11} organisms per ml proximal colonic contents (*1*)]. An estimated 2 to 4 million genes are embedded in the aggregate genome (microbiome) of an intestinal community of ~500 to 1000 bacterial species (*2*). The products of these genes provide metabolic capacities not encoded in our own genome (*3*).

The gut microbiota is a key regulator of the human immune system; it acts to induce tolerance to microbial epitopes and thus to reduce responses to commonly encountered foodstuffs and other environmental antigens (*4*). Functional genomic studies of germfree mice colonized with components of the human intestinal microbiota are revealing other functions affected by indigenous bacteria, including fortification of the mucosal barrier and angiogenesis (*5–7*).

These observations emphasize the need to understand more about the roles played by the microbiota in host biology, as well as the potential for control and modulation.

Here, we describe the complete 6.26-Mb genome sequence of the Gram-negative anaerobe, *Bacteroides thetaiotaomicron* (figs. S1 to S5 in supporting online material). This genetically manipulatable organism is a predominant member of the normal human (and murine) distal small intestinal and colonic microbiota (*8*) and has been used as a model for understanding the impact of constituents of the microbiota on gut gene expression (*5, 9*). The genome sequences of members of the Bacteroidetes phylum, which diverged early in the evolution of Bacteria (*10*), have not yet been reported.

The *B. thetaiotaomicron* type strain, VPI-5482 (ATCC 29148), was originally isolated from the feces of a healthy adult human. Of the 4779 predicted proteins in its proteome, 2782 (58%) were assigned putative functions on the basis of homology to other known proteins. Of the predicted proteins, 848 (18%) have homology to proteins with no known function, whereas 1149 (24%) have no appreciable homology to entries in public databases. The most markedly expanded paralogous groups are involved in polysaccharide uptake and degradation (glycosylhydrolases, cell-surface carbohydrate-binding proteins); capsular polysaccharide biosynthesis (e.g., glycosyltransferases); environmental sensing and signal transduction [one- and two-component systems; extracytoplasmic function (ECF)-type sigma factors]; and DNA mobilization (transposases, conjugative transposons) (table S1). These expansions reveal strategies used by *B. thetaiotaomicron* to

Department of Molecular Biology and Pharmacology, Washington University School of Medicine, St. Louis, MO 63110, USA.

*To whom correspondence should be addressed. E-mail: jgordon@molecool.wustl.edu