DNA RESEARCH **15**, 39–47, (2008) doi:10.1093/dnares/dsm030

Complete Genome Sequence of *Finegoldia magna*, an Anaerobic Opportunistic Pathogen

Takatsugu GOTO[1,*], Atsushi YAMASHITA[2], Hideki HIRAKAWA[3], Minenosuke MATSUTANI[4], Kozo TODO[1,5], Kenshiro OHSHIMA[2,6], Hidehiro TOH[2,7], Kazuaki MIYAMOTO[1], Satoru KUHARA[4], Masahira HATTORI[2,6,8], Tohru SHIMIZU[9], and Shigeru AKIMOTO[1]

Department of Microbiology, Wakayama Medical University, 811-1 Kimiidera, Wakayama, Wakayama 641-0012, Japan [1]; Laboratory of Genomic Information, Kitasato Institute for Life Sciences, Kitasato University, 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan[2]; Graduate School of Systems Life Sciences, Kyushu University, Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan[3]; Graduate School of Genetic Resource Technology, Kyushu University, Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan[4]; Department of Oral and Maxillofacial Surgery, Wakayama Medical University, 811-1 Kimiidera, Wakayama, Wakayama 641-0012, Japan[5]; Department of Computational Biology, Graduate School of Frontier Sciences, University of Tokyo, 5-1-5 Kashinoha, Kashiwa, Chiba 277-8561, Japan[6]; Computational and Experimental Systems Biology Group, RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan[7]; Genome Core Technology Facility, RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan[8] and Department of Bacteriology, Graduate School of Medical Science, Kanazawa University, 13-1 Takaramachi, Kanazawa, Ishikawa 920-8640, Japan[9]

(Received 23 October 2007; accepted on 5 December 2007; published online 7 February 2008)

Abstract

Finegoldia magna **(formerly *Peptostreptococcus magnus*), a member of the Gram-positive anaerobic cocci (GPAC), is a commensal bacterium colonizing human skin and mucous membranes. Moreover, it is also recognized as an opportunistic pathogen responsible for various infectious diseases. Here, we report the complete genome sequence of *F. magna* ATCC 29328. The genome consists of a 1 797 577 bp circular chromosome and an 189 163 bp plasmid (pPEP1). The metabolic maps constructed based on the genome information confirmed that most *F. magna* strains cannot ferment most sugars, except fructose, and have various aminopeptidase activities. Three homologs of albumin-binding protein, a known virulence factor useful for antiphagocytosis, are encoded on the chromosome, and one albumin-binding protein homolog is encoded on the plasmid. A unique feature of the genome is that *F. magna* encodes many sortase genes, of which substrates may be involved in bacterial pathogenesis, such as antiphagocytosis and adherence to the host cell. The plasmid pPEP1 encodes seven sortase and seven substrate genes, whereas the chromosome encodes four sortase and 19 substrate genes. These plasmid-encoded sortases may play important roles in the pathogenesis of *F. magna* by enriching the variety of cell wall anchored surface proteins.**

Key words: whole genome sequence; Gram-positive anaerobic cocci; *Peptostreptococcus magnus*; albumin-binding protein; sortase

1. Introduction

Gram-positive anaerobic cocci (GPAC) are a major part of the normal human flora colonizing skin and mucous membranes of the mouth and gastrointestinal tracts. GPAC account for 20–40% of all anaerobes

Edited by Satoshi Tabata
* To whom correspondence should be addressed. Tel. +81 73-441-0640. Fax. +81 73-448-1026. E-mail: t-goto@wakayama-med.ac.jp

© The Author 2008. Kazusa DNA Research Institute.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org

recovered from human clinical specimens.[1–3] In GPAC, *Finegoldia magna* (formerly *Peptostreptococcus magnus*) has the highest pathogenicity and is isolated most frequently from various infection sites, including soft tissue, bone and joint, and diabetic foot infections.[1] Approximately 70% of *F. magna* strains recovered from human clinical materials coexist with other bacterial species such as group D streptococci, *Staphylococcus*, *Bacteroides*, and *Fusobacterium*.[4] *Finegoldia magna* has well-known cell surface proteins such as albumin-binding protein and protein L, which may contribute to antiphagocytosis by binding the cell surface to human serum albumin and immunoglobulin light chain, respectively.[5–7] Despite this clinical importance of *F. magna*, other virulence factors and overall genomic information of *F. magna* have not been extensively studied. Therefore, we have so far constructed the physical map[8] and the genetic map[9,10] of *F. magna* strain ATCC 29328, originally isolated from an abdominal wound.

Here, we report the complete genome sequence of *F. magna* ATCC 29328 to characterize its genomic structure in detail and its nature as an opportunistic pathogen. This is the first complete genome sequence among GPAC, and we believe our data will be of great use for genetic studies of other GPAC as well as *F. magna*.

2. Materials and methods

2.1. Genome sequencing

The nucleotide sequence of the *Finegoldia magna* ATCC 29328 genome was determined by a whole-genome shotgun strategy. Briefly, *F. magna* ATCC 29328 was obtained from the American Type Culture Collection and was grown anaerobically at 37°C in liquid using Gifu Anaerobic Medium (GAM) (Nissui, Japan) for 39 h. The genomic DNA was isolated according to the conventional method using 0.4 mg/mL achromopeptidase (Wako, Japan), 1.3% sodium dodecyl sulfate and 2 mg Proteinase K (Roche, Germany). The isolated DNA was fragmented using the HydroShear process (GeneMachines Inc., USA). The fragmented DNA was fractionated by agarose gel electrophoresis to isolate the short (1–2 kb) and the long (4–5 kb) fractions of fragments. The 1–2 kb and 4-5 kb DNA were ligated into a *Hinc*II-digested pTS1 plasmid vector (Nippon genetech, Japan). The short-insert and the long-insert plasmids were introduced by electroporation into *Escherichia coli* DH5α and SURE2 (Stratagene, USA), respectively. A total of 38 776 reads (providing a 10-fold coverage) from both ends of inserts was sequenced using dye-terminator chemistry with MegaBACE1000, MegaBACE4000 (GE Healthcare, USA) and ABI 3037xl automated sequencers (ABI, USA). Sequence assembly was carried out using the Phred/Phrap/Consed package.[11] Gaps were closed by PCR direct sequencing with oligonucleotide primers designed to anneal to each end of the neighboring contigs and by primer walking with bridge clones. Sequences of four rRNA operons were independently determined by genomic PCR and with BAC library.[9,10] Sequencing of one long repeated region on the chromosome was performed by nested-deletion methods using Kilo-sequence deletion kit (Takara, Japan). Finally, sequences of one complete circular chromosome and one complete plasmid were obtained. Overall accuracy of the finished sequence was estimated to have an error rate of <1 per 10 000 bases (Phrap score ≥ 40).

2.2. Bioinformatics

Protein coding regions (ORFs >150 bp) were identified using GenomeGambler 1.51,[12] CRITICA,[13] GeneHacker,[14] and Glimmer 2.0[15] programs. In the initial step of ORF prediction, 205 ORFs showing high similarity to the housekeeping proteins in the Clusters of Orthologous Groups of Proteins (COGs) database[16] were selected and used to create the training data set for CRITICA, GeneHacker, and Glimmer programs. Individual predicted ORFs were reviewed manually for the presence of start codons (ATG, TTG, and GTG) and ribosome-binding sequences. Protein functional annotation was based on homology searches against NCBI's non-redundant protein database by BLASTP.[17] Transfer RNA genes were predicted by tRNAscan-SE.[18] Functional classification of ORFs was made by homology search against COGs using BLASTP. The metabolic pathway was constructed based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database[19] and was complemented manually according to pathways reported in several anaerobes.[20–23] Putative localization of predicted proteins was evaluated by the combination of PSORT[24] and SignalP program.[25] The characteristic motifs in all protein-coding sequence of ATCC 29328 genome were detected by the Pfam analyses (http://www.sanger.ac.uk/) and COGs database. Sortase homologs and their substrates from the sequences of 15 chromosomes and 30 plasmids derived from 15 Gram-positive bacteria species (include *F. magna* ATCC 29328), whose both chromosomes and plasmids were sequenced to date, were identified *in silico* according to methods described by Comfort and Clubb.[26] The genomic sequences and the protein tables of bacteria were obtained through the NCBI (National Center of Biological Information, http://www.ncbi.nlm.nih.gov/). The complete genome sequence and annotation data of *F. magna* ATCC 29328 is available from the DDBJ/EMBL/GenBank

database (accession nos. AP008971 and AP008972) and our web site (http://w4.grt.kyushu-u.ac.jp/Fmagna/).

3. Results and discussion

3.1. General genome features

The genome of *F. magna* ATCC 29328 consists of a circular chromosome (1 797 577 bp) and a plasmid pPEP1 (189 163 bp), with an average G + C content of 32.3 and 29.7%, respectively (Fig. 1 and Table 1). The chromosome size was in good agreement with the size (1886 ± 62 kb) estimated in the physical map which we constructed previously.[8] The recognition sites of *Apa*I, *Pme*I, *Sgr*AI, and I-*Ceu*I in our physical map[8] and the locations of ORFs in our genetic map[9,10] were in nearly agreement with those in the complete genome sequence, except missing of an *Apa*I-fragment (94 bp) in the chromosomal physical map and a inversion of *Pme*I fragments in the plasmid physical map. We could not find any regions showing significantly higher G + C content on the chromosome, except for four rRNA operon regions (Fig. 1). In the chromosome, we identified 1631 ORFs (FMG_0001–FMG_1631), covering 90.8% of the chromosome. Of these, 1177 (72.1%) could be assigned a putative function, 329 (20.2%) are conserved but the function unknown, and 125 (7.7%) are hypothetical genes unique to *F. magna* (Table 1). The replication origin of the chromosome was deduced on the basis of the transition point in GC-skew analysis and the vicinal presence of *dnaA* box.[27] A strong bias in gene orientation was observed, as 83.8% of ORFs on the chromosome are transcribed in the same orientation as DNA replication. Similarly, high bias in the gene orientation was observed in the genomes of *Clostridium perfringens* strain 13[20] and *Clostridium tetani* E88,[21] respectively, and appears to be a common feature of low-GC Gram-positive bacteria.[20,28,29] Plasmid pPEP1 encodes 182 ORFs (FMG_P0001–FMG_P0182). Of these, 45 (24.7%) could be assigned a putative function, 29 (15.9%) are conserved but the function unknown, and 108 (59.4%) are hypothetical genes (Table 1). The replication origin of the plasmid could not be deduced by GC-skew analysis. Of the total ORFs on the plasmid, 87.4% were encoded on one strand

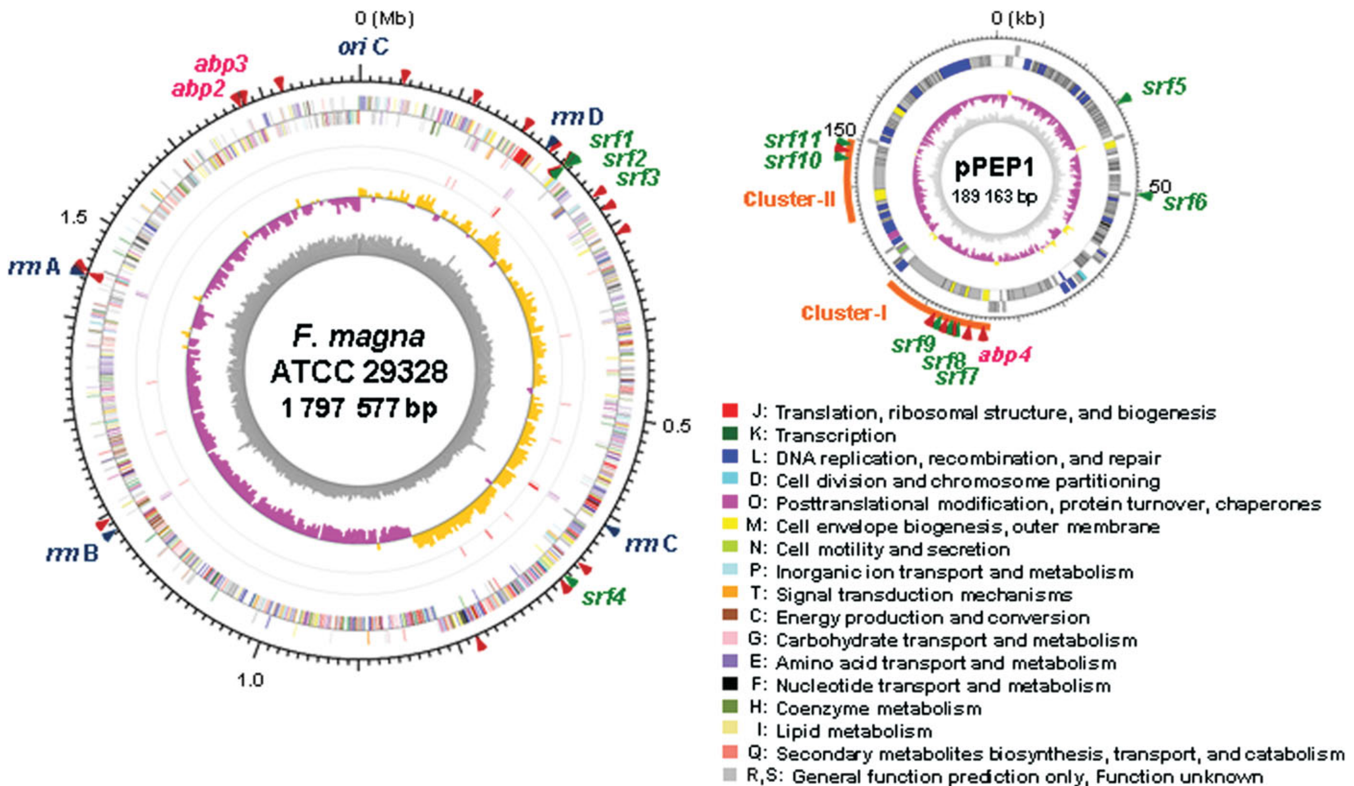

Figure 1. Circular maps of the chromosome and the pPEP1 plasmid of *F. magna* ATCC 29328. Each circle of represents (from inside out): G + C content, GC skew (G + C/G − C), tRNA genes, rRNA operons (all consisted of 16S–23S–5S rRNA genes), predicted genes transcribed into the counterclockwise direction, and those into clockwise direction. Green- and red-arrow heads show sortase homologs (*srf1* to *srf11*) and their substrates, respectively, identified *in silico* from the ATCC 29328 genome. The identified substrates contain three albumin-binding protein homologs (*abp2*, *abp3*, and *abp4*). The pPEP1 has two clusters (cluster-I and -II), containing several surface proteins identified by PSORT analyses[24] in one region. All of the predicted genes are colored according to the Clusters of Orthologous Groups of Proteins functional classification.[16]

Table 1. Genome features of *F. magna* ATCC 29328

	Chromosome	Plasmid pPEP1
Length of sequence, bp	1,797,577	189,163
G + C content, %	32.3	29.7
Total number of ORFs	1,631	182
No. with assigned functions	1,177	45
No. of conserved hypotheticals*	329	29
No. of hypothetical proteins†	125	108
Average ORF length, bp	1,005	888
Coding density, %	90.8	85.4
rRNA operon	4	0
tRNA gene	48	0

*Conserved hypothetical protein, sequence similarity to a translation of another ORF; however, there is currently no known function.

†Hypothetical protein, no significant similarity to another protein.

DNA (Fig. 1). The *F. magna* genome encodes only a few mobile genetic elements. The density of transposase genes per 1 Mb of chromosome in low-GC Gram-positive bacteria ranges from high (10.5 per Mb) in *Streptococcus pyogenes* MGAS8283 to low (3.63 per Mb) in *C. perfringens* strain 13. *F. magna* ATCC 29328 has only one transposase gene (FMG_0928) (0.56 per Mb) on the chromosome. In the plasmid, 14 putative transposase genes were encoded, but four out of 14 were truncated. *F. magna* has nine putative phage-related ORFs on the chromosome; however, these genes do not constitute a prophage region. Assuming that *F. magna* is closely related to clostridia[10], *F. magna* has a stable chromosome, like clostridia, and recent events of gene acquisition by horizontal gene transfer are less likely.[20–22]

3.2. Sugar and amino acid metabolisms

Most *F. magna* have a variety of aminopeptidase activities[30]; however, they have very limited saccharolytic activities.[1,31] It is known that most *F. magna* strains produce weak acid from the fermentation of fructose, and only a few strains produce weak acid from glucose, and all strains tested were unable to produce acid from maltose, mannose, and sucrose.[1,31] We found that *F. magna* ATCC 29328 has a complete glycolysis pathway only for fructose (Supplementary Fig. S1). Fructose could be taken in by a phosphotransferase system (PTS) for fructose and fed into glycolysis followed by production of acetate as a major end product. However, PTS genes for glucose, maltose, mannose, glucitol, cellobiose, and lactose were absent in the *F. magna* genome. PTS for mannitol, galactitol, and sucrose were incomplete.

The tricarboxylic acid (TCA) cycle of *F. magna* ATCC 29328 was incomplete; only the step from pyruvate to fumarate was present (Supplementary Fig. S1). Similarly, incompleteness of the TCA cycle in clostridia such as *C. tetani, C. acetobutylicum*, and *C. perfringens* also has been reported.[20–22] *F. magna* ATCC 29328 has only gene sets for the production of acetate (and ethanol); although clostridia have complete gene sets for production of butyrate, acetate, lactate, and ethanol.[20–22]

We also constructed a metabolic map of amino acid degradation to investigate which amino acids are used for energy acquisition in *F. magna* (Supplementary Fig. S1). The metabolic map showed that *F. magna* ATCC 29328 converts at least five amino acids (glycine, serine, threonine, aspartate, and asparagine) into pyruvate, and produces acetate and ATP. The metabolic map also showed that three (glycine, serine, and threonine) out of the five amino acids lead to NH_3 production. As to amino acid biosynthesis pathways, *F. magna* ATCC 29328 has the complete gene sets required for biosynthesis of only five amino acids (glycine, serine, threonine, aspartate, and asparagine), suggesting that many of amino acids must be taken in from the environment (Supplementary Fig. S1). *F. magna* ATCC 29328 has many amino acid/oligopeptide transporters, enabling it to utilize amino acids as a major energy source by taking amino acids from the environment. As reported previously, *F. magna* has more aminopeptidase activities than other GPAC species.[30] Bacterial aminopeptidases are generally regarded as a tissue-degrading enzyme and bacteria can utilize the degradation products as a nutrient source. In *F. magna* ATCC 29328 genome, four carboxypeptidase genes and six aminopeptidase genes including a Xaa-Pro aminopeptidase gene were found. These peptidase genes may contribute to bacterial growth and could account for higher pathogenicity of *F. magna* in GPAC.

Finegoldia magna ATCC 29328 has at least 249 transport-related genes, accounting for 15.3% of the total ORFs. Of them, 49 ATP-binding cassette transporter systems (ABC transporters) could be identified on the basis of the similarities with ABC transporters from *E. coli*, *Bacillus subtilis*, and *Mycobacterium tuberculosis*.[32–34] These ABC transporters were divided into 14 importers (e.g., amino acid/oligopeptide importers) and 35 exporters (e.g., putative multidrug exporters) (Supplementary Fig. S1). We also found other types of transporters such as electron, ion, a fructose-specific PTS, amino acid/oligopeptide, and multidrug-efflux transporters. *Finegoldia magna* has two vacuolar (V)-type ATPases (one is similar to an ATPase of *C. tetani* and another is similar to that of *Fusobacterium nucleatum*), but it has no F_0F_1-type ATPase.

3.3. *Regulation and stress responses*

Transcriptional and translational regulators, and stress responses of *F. magna* are poorly understood. *Finegoldia magna* ATCC 29328 has an RNA polymerase core enzyme (composed of RpoA, RpoB, RpoC, and RpoZ subunits), a major sigma factor, and two general stress sigma factors, but it has no an ECF-type sigma factor,[35,36] which has been found in several Gram-positive bacteria. *Finegoldia magna* encodes 12 two-component regulator systems consisting of 13 sensor histidine kinases and 12 response regulators. An autoinducer-2 production protein LuxS related to a quorum-sensing system was encoded in the chromosome. The *F. magna* genome encodes no superoxide dismutase gene, but encodes genes for superoxide reductase, NADH oxidase, and putative NADH dehydrogenase. These genes may play a major role in enabling the organism to survive in intermediate aerobic conditions, such as mucosa and skin.

3.4. *Virulence factors*

Putative virulence factors were screened from all ORFs of *F. magna* ATCC 29328 genome by BLAST, SignalP, PSORT, and Pfam analyses. We found some putative virulence-related proteins with signal sequences and characteristic motifs (Supplementary Table S1). Interestingly, the *F. magna* ATCC 29328 genome encodes the following four albumin-binding protein homologs: three (FMG_1501, FMG_1517, and FMG_1523) in the chromosome and one (FMGP_0118) in the plasmid. These four proteins have regions similar to the albumin-binding modules (GA module)[6,37,38] in PAB and urPAB proteins derived from *F. magna* ALB8 and ALB1B strains, respectively (Supplementary Fig. S2). These GA modules were also found in protein G from group C and G streptococci and in the protein L-like molecule from *F. magna* strain 3316.[39,40]

Other virulence factors, except albumin-binding protein and protein L, have not been well studied in *F. magna*. In the ATCC 29328 genome, we found ten collagen adhesion protein homologs (Supplementary Table S1). These proteins possess several domains similar to the B-type domain (Pfam accession PF05738) in a collagen adhesion protein (Cna) derived from *Staphylococcus aureus*.[41] Furthermore, the ATCC 29328 genome encodes 20 *N*-acetylmuramoyl-L-alanine amidase homologs (Cwp66 homologs) (Supplementary Table S1). These homologs have several putative cell wall binding repeat 2 (Pfam accession PF04122) motifs. The PF04122 motif is identified in autolysin and Cwp66, an adhesin of *Clostridium difficile*.[42] The Cwp66 has been suggested as a multifunctional protein[42]; the C-terminal domain of Cwp66 gave partial adherence ability of *C. difficile* to host cells, the N-terminal domain has homology to some autolysins and could contribute to autolysis. We are interested in whether Cwp66 homologs of *F. magna* can act as adhesins to host cells, since the Cwp66 of *C. difficile* has partial adherence ability and are regarded as a virulence factor. Moreover, the ATCC 29328 genome encodes putative virulence factors such as a cell wall-associated serine proteinase precursor (FMG_0035) and a putative biofilm-associated surface protein (FMG_1333) (Supplementary Table S1).

Taken together, the pathogenic *F. magna* ATCC 29328 has several antiphagocytotic proteins such as albumin-binding proteins, and multiple putative host cell adhesins such as collagen adhesion proteins and Cwp66 homologs. Although the functions of these adhesins are not yet understood, future studies should disclose the mechanisms of virulence and the significance of *F. magna* as a member of the normal human flora.

An extracellular transpeptidase, sortase, catalyzes cell wall anchoring of surface proteins by cleaving the threonine and glycine residues of conserved LPXTG-like motif.[43–46] Sortases (SrtA, SrtB) from *Staphylococcus aureus* were identified first and have been well studied recently.[47,48] Since some pathogens deficient in sortases exhibit reduced virulence, much interest has been focussed on sortase as a target for new antimicrobial agents.[43–46] Recently, Comfort and Clubb identified sortases and their substrates (surface proteins with the LPXTG-like motif) *in silico* from 241 bacterial chromosomes (96 species) sequenced completely, and they predicted some functional sortase–substrate linkages.[26] By the same method with Comfort, we identified sortases and their substrates from the ATCC 29328 genome and several Gram-positive bacterial genomes.

The ATCC 29328 chromosome encoded four sortase homologs and 19 putative substrates (Figs. 1 and 2A and Table 2; Supplementary Tables S2 and S3). These 19 putative substrates include two albumin-binding protein homologs (FMG_1517 and FMG_1523) and putative biofilm-associated surface protein (FMG_1333), which may act as virulence factor (Table 2; Supplementary Tables S1 and S3). Most interestingly, seven sortase homologs and seven substrates were identified from the plasmid pPEP1 (Figs. 1 and 2B and Table 2; Supplementary Tables S2 and S3). These seven substrates include an albumin-binding protein homolog (FMG_P0118) and four putative collagen adhesion proteins (FMG_P0123, FMG_P0125, FMG_P0146, and FMG_P0149) (Table 2; Supplementary Tables S1 and S3).

By the same method with Comfort, we predicted functional sortase–substrate linkages for all sortases

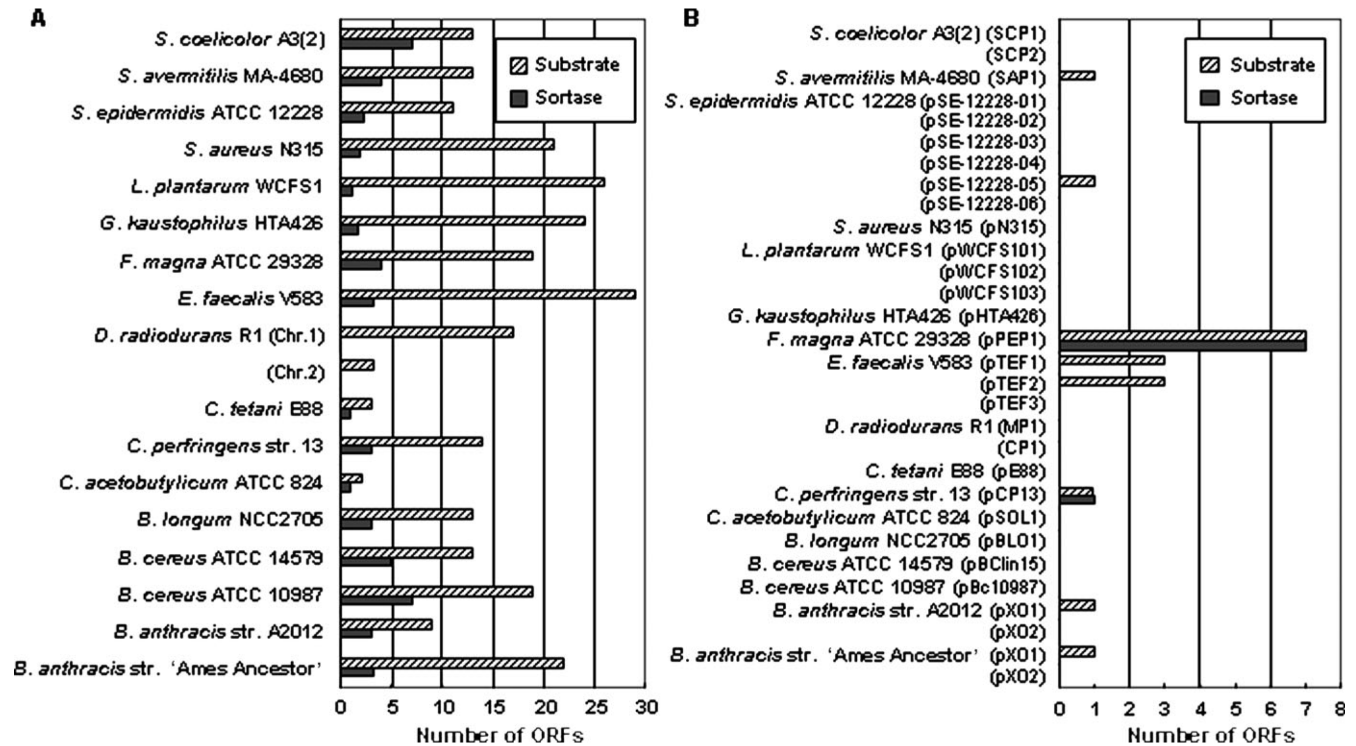

Figure 2. Sortase homologs and substrates identified *in silico* from Gram-positive bacterial genomes. (**A**) Total number of ORFs in chromosome (15 species). (**B**) Total number of ORFs in plasmid (15 species, 30 plasmids). We identified ORFs from 15 Gram-positive bacterial species, whose both chromosomes and plasmids were sequenced to date. Abbreviations of bacterial species and detailed ORF data are showed in Supplementary Tables S2 and S3.

and substrates identified from the tested genomes. From the ATCC 29328 genome, we predicted some linkage among two sortases (FMG_0189 and FMG_0190) and two substrates (FMG_0187 and FMG_0188), because these four ORFs belong to a common family (Family 3 defined by Comfort) and are located on one cluster in the chromosome (Table 2; Supplementary Tables S2 and S3). The remaining sortases and substrates of *F. magna* were assigned to an unclassified family (defined by Comfort), and their sortase–substrate linkages could not be predicted. Assuming that an unclassified substrate and its sortase in a common genomic cluster may have some functional linkage,[26] eight out of the 24 unclassified substrates can be linked to their specific sortases (Table 2; Supplementary Tables S2 and S3). These predicted sortase–substrate linkages on the genome should be further elucidated experimentally. For example, if we can identify a sortase that specifically recognize the LPKAG motif in albumin-binding proteins, it will lead to the development of sortase inhibitors as new antimicrobial agents.[43]

To our knowledge, the pPEP1 plasmid may have the largest number of sortase genes among Gram-positive bacterial plasmids analysed so far, because we searched sortases and their substrates in 29 plasmids of 14 Gram-positive bacterial species, except for *F. magna*, and found that no plasmids have a sortase gene, except for *C. perfringens* str.13 plasmid (pCP13) with a sortase gene (PCP56) (Fig. 2B; Supplementary Tables S2 and S3). We could not find a remarkable feature about the number of sortases and substrates among Gram-positive bacterial chromosomes (Fig. 2A). The presence of as many as seven sortase genes in a plasmid may be unique to *F. magna*, since there are no sortase genes even on the large plasmids of the following Gram-positive species: *Bacillus anthracis* pXO1 (181 677 bp), *Bacillus cereus* pBc10987 (208 369 bp), and *C. acetobutylicum* pSOL1 (192 000 bp).

3.5. Conclusions

This is the first report of a whole genome sequence of a member of any known GPAC. The sequence information presented and analysed here will contribute to an understanding of the taxonomy and pathogenicity of GPAC. The reason why *F. magna* is isolated most frequently and has the highest pathogenicity in GPAC may be due to the fact that *F. magna* genome encodes abundant aminopeptidases, multiple transporters and some virulence factors (e.g., albumin-binding proteins, Cwp66 homologs, and collagen adhesion proteins), beside with the limited capabilities of *F. magna* for sugar utilization and amino acid biosynthesis. It is of particular interest that four

Table 2. The substrates of sortases identified in *F. magna* ATCC 29328 genome

ORF no.	Product name	Motif*	Subfamily†	Cluster‡
Chromosome				
FMG_0037	Putative collagen adhesion protein	LPESGS	—	No
FMG_0095	Conserved hypothetical protein	PAAASF	—	No
FMG_0147	Putative membrane protein	IALTAV	—	No
FMG_0186	Hypothetical protein	YPLTGA	—	Yes
FMG_0187	Cell wall surface anchor family protein	IPQTGG	3	Yes
FMG_0188	Putative cell wall surface anchor family protein	IPKTGD	3	Yes
FMG_0229	Putative surface protein precursor	LPKAGT	—	No
FMG_0238	Branched-chain amino acid transporter	LPLSSV	—	No
FMG_0261	Conserved hypothetical protein	YAASKS	—	No
FMG_0595	Phosphate ABC transporter permease protein	MAYASG	—	No
FMG_0616	Cell division protein homolog	IPPTGL	—	No
FMG_0717	Cell division protein RodA homolog	MPITGI	—	No
FMG_1103	Conserved hypothetical protein	VPFAGL	—	No
FMG_1333	Putative biofilm-associated surface protein	LPKAGI	—	No
FMG_1335	Hypothetical protein	LPKAGA	—	No
FMG_1352	Hypothetical protein	LPKAGY	—	No
FMG_1517	Albumin-binding protein homolog-2	LPKAGS	—	No
FMG_1523	Putative albumin-binding protein homolog-3	LPKAGI	—	No
FMG_1550	Conserved hypothetical protein	LPKAGS	—	No
Plasmid				
FMG_P0118	Albumin-binding protein homolog-4 like protein L	LPKAGS	—	Yes
FMG_P0119	Hypothetical protein	LPKAGV	—	Yes
FMG_P0121	Hypothetical protein	YAKTNI	—	Yes
FMG_P0123	Putative collagen adhesion protein	VAIAGG	—	Yes
FMG_P0125	Putative collagen adhesion protein	LPVTGL	—	Yes
FMG_P0146	Putative collagen adhesion protein	IPSTGT	—	Yes
FMG_P0149	Putative collagen adhesion protein	VPGTGT	—	Yes

*LPXTG-like motifs in substrates. These motifs were identified according to the methods of Comfort and Clubb.[26]

†Categorization into subfamilies according to sequence homology using PSI-BLAST profiles and a hidden Markov model.'3' and '—' show the subfamily-3 and the unclassified subfamily, respectively, which were defined by Comfort.[26]

‡'Yes' shows that a sortase is genomically adjacent to its substrate.

sortase genes were identified on the chromosome and that seven sortase genes were identified on the plasmid. Seven plasmid-encoded sortase genes unique to *F. magna* are of special interest, considering that plasmids are commonly considered to be of foreign origin. These mobile element-encoded sortases may play important roles in the pathogenesis of *F. magna* by enriching the variety of cell wall anchored surface proteins, which may enhance the interaction with host tissues and with other bacteria in mixed infection sites. Future studies, such as sortase–substrate linkage analyses and animal model experiments using sortase-negative mutant strains, will be required to clarify the roles of sortase on pathogenesis in *F. magna*.

Acknowledgements: We thank T. Hayashi, N. Ogasawara, Y. Ohnishi, H. Hayashi, and H. Yoshikawa for their advice, support, and encouragement. We thank M. Yoshida, K. Furuya, C. Yoshino, A. Nakazawa, Y. Yamashita, and N. Ito for technical assistance.

The sequences reported in this paper have been deposited in the DDBJ/EMBL/GenBank database (accession chromosome, AP008971; plasmid, AP008972).

Supplementary Data: Supplementary data are available online at www.dnaresearch.oxfordjournals.org.

Funding

This work was supported by the Research for the Future Program from the Japan Society for the

Promotion of Science (JSPS-RFTF00L01411), in part by the Grant of the 21st Century COE Program, Ministry of Education, Culture, Sports, Science and Technology, Japan to Kitasato University, in part by Grant-in-Aid for Young Scientists (B) from the Ministry of Education, Culture, Sports, Science and Technology, Japan (MEXT-15790226), and in part by 2005 Wakayama Medical Award for Young Researchers.

References

1. Murdoch, D. A. 1998, Gram-positive anaerobic cocci, *Clin. Microbiol. Rev.*, **11**, 81–120.
2. Murdoch, D. A. 1998, In: Collier, L., Balows, A., and Sussman, M. (eds.), *Topley & Wilson's Microbiology and Microbial Infections*, 9th Ed., London: Arnold, pp. 783–797.
3. Koneman, E. W., Allen, S. D., Janda, W. M., Schreckenberger, P. C., and Winn, Jr. W. C. 1997, *Color Atlas and Textbook of Diagnostic Microbiology*, 5th Ed., Philadelphia: Lippincott, pp. 709–784.
4. Bourgault, A. M., Rosenblatt, J. E. and Fitzgerald, R. H 1980, *Peptococcus magnus*: a significant human pathogen, *Ann. Intern. Med.*, **93**, 244–248.
5. de Château, M. and Björck, L. 1996, Identification of interdomain sequences promoting the intronless evolution of a bacterial protein family, *Proc. Natl. Acad. Sci.* USA, **93**, 8490–8495.
6. Lejon, S., Frick, I. M., Björck, L., Wikström, M. and Svensson, S. 2004, Crystal structure and biological implications of a bacterial albumin binding module in complex with human serum albumin, *J. Biol. Chem.*, **279**, 42924–42928.
7. Smith, D., D'Argy, R., Nilsson, M., et al. 2004, Whole-body autoradiography reveals that the *Peptostreptococcus magnus* immunoglobulin-binding domains of protein L preferentially target B lymphocytes in the spleen and lymph nodes *in vivo*, *Cell. Microbiol.*, **6**, 609–623.
8. Todo, K., Goto, T., Miyamoto, K. and Akimoto, S. 2002, Physical and genetic map of the *Finegoldia magna* (formerly *Peptostreptococcus magnus*) ATCC 29328 genome, *FEMS Microbiol. Lett.*, **210**, 33–37.
9. Todo, K., Goto, T., Honda, A., et al. 2004, Comparative analysis of the four rRNA operons in *Finegoldia magna* ATCC 29328, *Syst. Appl. Microbiol.*, **27**, 18–26.
10. Goto, T., Todo, K., Miyamoto, K. and Akimoto, S. 2003, Bacterial artificial chromosome library of *Finegoldia magna* ATCC 29328 for genetic mapping and comparative genomics, *Microbiol. Immunol.*, **47**, 1005–1016.
11. Gordon, D., Desmarais, C. and Green, P 2001, Automated finishing with autofinish, *Genome Res.*, **11**, 614–625.
12. Sakiyama, T., Takami, H., Ogasawara, N., et al. 2000, An automated system for genome analysis to support microbial whole-genome shotgun sequencing, *Biosci. Biotechnol. Biochem.*, **64**, 670–673.
13. Badger, J. H. and Olsen, G. J. 1999, CRITICA: Coding region identification tool invoking comparative analysis, *Mol. Biol. Evol.*, **16**, 512–524.
14. Yada, T., Nakao, M., Totoki, Y. and Nakai, K. 1999, Modeling and predicting transcriptional units of *Escherichia coli* genes using hidden Markov models, *Bioinformatics*, **15**, 987–993.
15. Frishman, D., Mironov, A., Mewes, H. W. and Gelfand, M. 1998, Combining diverse evidence for gene recognition in completely sequenced bacterial genomes, *Nucleic Acids Res.*, **26**, 2941–2947.
16. Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., et al. 2001, The COG database: new developments in phylogenetic classification of proteins from complete genomes, *Nucleic Acids Res.*, **29**, 22–28.
17. Altschul, S. F., Madden, T. L., Schäffer, A. A., et al. 1997, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.*, **25**, 3389–3402.
18. Lowe, T. M. and Eddy, S. R. 1997, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence, *Nucleic Acids Res.*, **25**, 955–964.
19. Kanehisa, M., Goto, S., Kawashima, S. and Nakaya, A. 2002, The KEGG databases at GenomeNet, *Nucleic Acids Res.*, **30**, 42–46.
20. Shimizu, T., Ohtani, K., Hirakawa, H., et al. 2002, Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater, *Proc. Natl. Acad. Sci. USA*, **99**, 996–1001.
21. Brüggemann, H., Bäumer, S., Fricke, W. F., et al. 2003, The genome sequence of *Clostridium tetani*, the causative agent of tetanus disease, *Proc. Natl. Acad. Sci. USA*, **100**, 1316–1321.
22. Nolling, J., Breton, G., Omelchenko, M. V., et al. 2001, Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*, *J. Bacteriol.*, **183**, 4823–4838.
23. Kapatral, V., Anderson, I., Ivanova, N., et al. 2002, Genome sequence and analysis of the oral bacterium *Fusobacterium nucleatum* strain ATCC 25586, *J. Bacteriol.*, **184**, 2005–2018.
24. Nakai, K. and Horton, P. 1999, PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization, *Trends. Biochem. Sci.*, **24**, 34–36.
25. Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. 1997, Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites, *Protein Eng.*, **10**, 1–6.
26. Comfort, D. and Clubb, R. T. 2004, A comparative genome analysis identifies distinct sorting pathways in Gram-positive bacteria, *Infect. Immun.*, **72**, 2710–2722.
27. Salazar, L., Fsihi, H., de Rossi, E., et al. 1996, Organization of the origins of replication of the chromosomes of *Mycobacterium smegmatis*, *Mycobacterium leprae* and *Mycobacterium tuberculosis* and isolation of a functional origin from *M. semgmatis*, *Mol. Microbiol.*, **20**, 283–293.
28. Fraser, C. M., Gocayne, J. D., White, O., et al. 1995, The minimal gene complement of *Mycoplasma genitalium*, *Science*, **270**, 397–403.
29. Karlin, S. 1999, Bacterial DNA strand compositional asymmetry, *Trends Microbiol.*, **7**, 305–308.
30. Ng, J., Ng, L. K., Mayrand, D. and Dillon, J. A. 1998, Aminopeptidase activities in *Peptostreptococcus* spp. are

statistically correlated to gelatin hydrolysis, *Can. J. Microbiol.*, **44**, 303–306.
31. Ezaki, T., Yamamoto, N., Ninomiya, K., Suzuki, S. and Yabuuchi, E. 1983, Transfer of *Peptococcus indolicus*, *Peptococcus asaccharolyticus*, *Peptococcus prevotii*, and *Peptococcus magnus* to the Genus *Peptostreptococcus* and proposal of *Peptostreptococcus tetradius* sp. nov, *Int. J. Syst. Bacteriol.*, **33**, 683–698.
32. Linton, K. J. and Higgins, C. F. 1998, The *Escherichia coli* ATP-binding cassette (ABC) proteins, *Mol. Microbiol.*, **28**, 5–13.
33. Quentin, Y., Fichant, G. and Denizot, F. 1999, Inventory, assembly and analysis of *Bacillus subtilis* ABC transport systems, *J. Mol. Biol.*, **287**, 467–484.
34. Braibant, M., Gilot, P. and Content, J. 2000, The ATP binding cassette (ABC) transport systems of *Mycobacterium tuberculosis*, *FEMS Microbiol. Rev.*, **24**, 449–467.
35. Lonetto, M. A., Brown, K. L., Rudd, K. E. and Buttner, M. J. 1994, Analysis of the *Streptomyces coelicolor sigE* gene reveals the existence of a subfamily of eubacterial σ factors involved in the regulation of extracytoplasmic functions, *Proc. Natl. Acad. Sci.* USA, **91**, 7573–7577.
36. Helmann, J. D. 2002, The extracytoplasmic function (ECF) sigma factors, *Adv. Microb. Physiol.*, **46**, 47–110.
37. Johansson, M. U., de Château, M., Wikström, M., Forsén, S., Drakenberg, T. and Björck, L. 1997, Solution structure of the albumin-binding GA module: a versatile bacterial protein domain, *J. Mol. Biol.*, **266**, 859–865.
38. Johansson, M. U., Frick, I. M., Nilsson, H., et al. 2002, Structure, specificity, and mode of interaction for bacterial albumin-binding modules, *J. Biol. Chem.*, **277**, 8114–8120.
39. Guss, B., Eliasson, M., Olsson, A., et al. 1986, Structure of the IgG-binding regions of streptococcal protein G, *EMBO J.*, **5**, 1567–1575.
40. Murphy, J. P., Duggleby, C. J., Atkinson, M. A., Trowern, A. R., Atkinson, T. and Goward, C. R. 1994, The functional units of a peptostreptococcal protein L, *Mol. Microbiol.*, **12**, 911–920.
41. Patti, J. M., Jonsson, H., Guss, B., et al. 1992, Molecular characterization and expression of a gene encoding a *Staphylococcus aureus* collagen adhesion, *J. Biol. Chem.*, **267**, 4766–4772.
42. Waligora, A. J., Hennequin, C., Mullany, P., Bourlioux, P., Collignon, A. and Karjalainen, T. 2001, Characterization of a cell surface protein of *Clostridium difficile* with adhesive properties, *Infect. Immun.*, **69**, 2144–2153.
43. Cossart, P. and Jonquières, R. 2000, Sortase, a universal target for therapeutic agents against Gram-positive bacteria?, *Proc. Natl. Acad. Sci. USA*, **97**, 5013–5015.
44. Mazmanian, S. K., Ton-That, H., and Schneewind, O. 2001, Sortase-catalyzed anchoring of surface proteins to the cell wall of *Staphylococcus aureus*, *Mol. Microbiol.*, **40**, 1049–1057.
45. Navarre, W. W. and Schneewind, O. 1999, Surface proteins of Gram-positive bacteria and mechanisms of their targeting to the cell wall envelope, *Microbiol. Mol. Biol. Rev.*, **63**, 174–229.
46. Novick, R. P. 2000, Sortase: the surface protein anchoring transpeptidase and the LPXTG motif, *Trends. Microbiol.*, **8**, 148–151.
47. Zong, Y., Bice, T. W., Ton-That, H., Schneewind, O. and Narayana, S. V. 2004, Crystal structures of *Staphylococcus aureus* sortase A and its substrate complex, *J. Biol. Chem.*, **279**, 31383–31389.
48. Mazmanian, S. K., Ton-That, H., Su, K. and Schneewind, O. 2002, An iron-regulated sortase anchors a class of surface protein during *Staphylococcus aureus* pathogenesis, *Proc. Natl. Acad. Sci.* USA, **99**, 2293–2298.