JOURNAL OF BACTERIOLOGY, June 2011, p. 2896–2897 Vol. 193, No. 11
0021-9193/11/$12.00 doi:10.1128/JB.00301-11

Complete Genome Sequence of the Pathogenic Bacterium *Riemerella anatipestifer* Strain RA-GD▽

Jianfeng Yuan,[1]† Wanfei Liu,[2,3]† Mingfei Sun,[1] Shuhui Song,[2] Jianping Cai,[1]* and Songnian Hu[2]*

Institute of Veterinary Medicine, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, People's Republic of China[1]; Key Laboratory of Genome Science and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100029, People's Republic of China[2]; and Graduate University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China[3]

Received 3 March 2011/Accepted 16 March 2011

***Riemerella anatipestifer* is a well-described pathogen of waterfowl and other avian species which can cause a great loss to the poultry industry. Here we obtained the complete genome sequence of *R. anatipestifer* strain RA-GD, which was isolated from an infected duck in Guangzhou, China, and was cultivated in our laboratory.**

Riemerella anatipestifer is a Gram-negative, nonsporulating, pleomorphic rod bacterium and belongs to the family *Flavobacteriaceae*. *R. anatipestifer* was first reported by Riemer (in 1904) as a pathogen of geese. Since then it has been isolated from many avian species, including ducks, turkeys, chickens, and wild waterfowl, and can cause septicemic and exudative diseases in susceptible hosts. To date, 21 serotypes have been reported for *R. anatipestifer*, but there is no cross-immunoprotection among them (15). In most duck farms of China, morbidity due to *R. anatipestifer* infection is more than 50% and the lethality can reach 80%, even though chemotherapeutic methods have been used. Although *R. anatipestifer* of serotype 1 is the predominant prevalent agent, little is known about its pathogenesis and immune-response-triggering mechanisms (18). In order to develop more-effective preventive methods for the diseases, we sequenced the complete genome of *R. anatipestifer* strain RA-GD of serotype 1.

The whole genome was sequenced using Roche's 454 sequencing. Then the raw reads were assembled into 125 contigs with Newbler, and gaps were closed by PCR amplification and ABI3730XL sequencing. After that, genome finishing was carried out using the Phred/Phrap/Consed software package (2, 3, 5). The prediction of protein-coding sequences was generated using Glimmer 3.0 and Genemark (1, 14). The tRNAs and rRNAs were detected using tRNAscan-SE and RNAmmer, respectively (12, 13). Genome annotation was performed by searching against NCBI nonredundant proteins, clusters of orthologous groups (COGs), Interproscan, and the Kyoto Encyclopedia of Genes and Genomes (KEGG) (9–11, 16, 17, 19).

R. anatipestifer strain RA-GD has one circular chromosome of 2,166,384 bp with a 35.01% G+C content and no plasmid. The genome has 2,091 coding sequences (1,985 protein-coding genes, 54 pseudogenes, 40 tRNA genes, 9 rRNA genes, and 3 small RNAs), with an average length of 947 bases, constituting 91.42% of the genome. In addition, there are two prophage sequences, 17 insertion sequences (IS), two definite CRISPRs, and one possible CRISPR (4, 6–8).

Among all proteins, we found that 28 proteins are related to absorption and use of Fe; this suggests that Fe maybe very important to growth and pathogenesis of RA-GD. Moreover, we found 464 signal peptide proteins (22.19%), and they may contribute to host infection. Furthermore, we didn't find any secretion system like those previously found in Gram-negative bacteria, but the Sec pathway and the two-arginine (Tat) pathway were identified. More importantly, we identified 143 potential virulence factors, such as those related to adherence, invasion, motility, toxin, stress, antiphagocytosis, and serum resistance. Among these potential virulence factors, we found four proteins that have signal peptides, and they belong to three clusters of signal peptide proteins. One cluster includes an OmpA/MotB domain-containing protein (RIA_0060) and a hypothetical protein with a signal peptide (RIA_0059); meanwhile, one D-alanyl-D-alanine carboxypeptidase/endopeptidase (RIA_0058) is located in a complementary chain nearby. We thought that this cluster may be involved in the bacterium-host interaction. The complete genome sequence of RA-GD will provide essential information on understanding *R. anatipestifer* and give us new insights into control of *Riemerella anatipestifer* infection.

Nucleotide sequence accession number. The complete nucleotide sequence of *R. anatipestifer* strain RA-GD was deposited in GenBank under accession number CP002562.

* Corresponding author. Mailing address for Songnian Hu: Beijing Institute of Genomics, Chinese Academy of Sciences, No. 7 Beitucheng West Road, Chaoyang District, Beijing 100029, People's Republic of China. Phone: 86-10-82995362. Fax: 86-10-82995401. E-mail: husn@big.ac.cn. Mailing address for Jianping Cai: Institute of Veterinary Medicine, Guangdong Academy of Agricultural Sciences, Wushan, Tianhe District, Guangzhou 510640, People's Republic of China. Phone: 86-20-85292237. Fax: 86-20-85291691. E-mail: jianpcai@gmail.com.

† These authors contributed equally to this work.

▽ Published ahead of print on 25 March 2011.

This research was supported by the Natural Science Foundation of Guangdong Province (under grant no. 9251064001000010 to Jianing Cai) and the Key Science & Technology Brainstorm Project of Guangzhou (under grant no. 2008GZ-C211 to Jianing Cai).

REFERENCES

1. **Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg.** 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. **27:**4636–4641.
2. **Ewing, B., and P. Green.** 1998. Base-calling of automated sequencer traces usingPhred. II. Error probabilities. Genome Res. **8:**186.
3. **Ewing, B., L. D. Hillier, M. C. Wendl, and P. Green.** 1998. Base-calling of

automated sequencer traces using Phred. I. Accuracy assessment. Genome Res **8:**175.

4. **Fouts, D. E.** 2006. Phage_Finder: automated identification and classification of prophage regions in complete bacterial genome sequences. Nucleic Acids Res. **34:**5839–5851.
5. **Gordon, D., C. Abajian, and P. Green.** 1998. Consed: a graphical tool for sequence finishing. Genome Res. **8:**195–202.
6. **Grissa, I., G. Vergnaud, and C. Pourcel.** 2008. CRISPRcompar: a website to compare clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. **36:**W145–W148.
7. **Grissa, I., G. Vergnaud, and C. Pourcel.** 2007. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics **8:**172.
8. **Grissa, I., G. Vergnaud, and C. Pourcel.** 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. **35:**W52–W57.
9. **Kanehisa, M., and S. Goto.** 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. **28:**27–30.
10. **Kanehisa, M., S. Goto, M. Furumichi, M. Tanabe, and M. Hirakawa.** 2010. KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res. **38:**W355–W360.
11. **Kanehisa, M., et al.** 2006. From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res. **34:**W354–W357.
12. **Lagesen, K., et al.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35:**3100–3108.
13. **Lowe, T. M., and S. R. Eddy.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25:**955–964.
14. **Salzberg, S. L., A. L. Delcher, S. Kasif, and O. White.** 1998. Microbial gene identification using interpolated Markov models. Nucleic Acids Res. **26:**544–548.
15. **Sandhu, T. S.** 2008. Riemerella anatipestifer infection, p. 758–764. *In* Y. M. Saif (ed.), Diseases of poultry, 12th ed. Blackwell Publishing Ltd., Oxford, United Kingdom.
16. **Tatusov, R. L., et al.** 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics **4:**41.
17. **Tatusov, R. L., E. V. Koonin, and D. J. Lipman.** 1997. A genomic perspective on protein families. Science **278:**631–637.
18. **Wu, C. Y., et al.** 2009. Serotypes and drug resistances of Riemerella anatipestifer isolates from Guangdong districts. Anim. Husbandry Vet. Med. **41**(5)**:**22–25. (In Chinese.)
19. **Zdobnov, E. M., and R. Apweiler.** 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. Bioinformatics **17:** 847–848.