Genome Sequences of *Salmonella enterica* Serovar Typhimurium, Choleraesuis, Dublin, and Gallinarum Strains of Well-Defined Virulence in Food-Producing Animals

Emily J. Richardson, Bhakti Limaye, Harshal Inamdar, Avik Datta, K. Sunitha Manjari, Gillian D. Pullinger, Nicholas R. Thomson, Rajendra R. Joshi, Michael Watson, and Mark P. Stevens

The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, Scotland, United Kingdom; Centre for Development of Advanced Computing, University of Pune Campus, Pune 411007, India; Enteric Bacterial Pathogens Laboratory, Institute for Animal Health, Compton, Berkshire, RG20 7NN, United Kingdom; and The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, United Kingdom

Received 23 March 2011/Accepted 31 March 2011

*Salmonella enterica* is an animal and zoonotic pathogen of worldwide importance and may be classified into serovars differing in virulence and host range. We sequenced and annotated the genomes of serovar Typhimurium, Choleraesuis, Dublin, and Gallinarum strains of defined virulence in each of three food-producing animal hosts. This provides valuable measures of intraserovar diversity and opportunities to formally link genotypes to phenotypes in target animals.

*Salmonella enterica* causes salmonellosis in humans and other warm-blooded animals. Over 2,600 serovars have been classified according to the reactivity of antisera to somatic lipopolysaccharide and flagellar antigens and are broadly grouped on the basis of host range and disease presentation. The molecular basis of the differential virulence and tropism of serovars remains ill defined (20). An understanding of such processes is required to develop strategies for disease control and to predict the threat posed by isolates from animals.

The extent to which currently sequenced strains are typical of the wider serovar is open to question. We report the sequencing and annotation of four strains representing serovars that produce significant illness in food-producing animals: *S. Typhimurium* strain ST4/74 (11), originally isolated from a calf with salmonellosis in the United Kingdom (17) and the parent of the widely used mouse virulent *hisG* auxotroph SL1344 (10); *S. Choleraesuis* var. kunzendorf strain SCSA50, a field isolate from a case of swine typhoid in the United Kingdom (17); *S. Dublin* strain SD3246, a Vi-negative isolate from a calf with systemic salmonellosis in the United Kingdom (24); and *S. Gallinarum* SG9, first described to cause fowl typhoid in orally dosed chickens by Smith in 1955 (19). Crucially, the virulence of each strain has been reciprocally compared in calves, pigs, and chickens (3, 4, 6, 14, 15, 16, 17, 24, 25, 26, 27), fulfilling Koch’s postulates and enabling strain genotypes to be linked to phenotypes in target hosts.

**Sequencing and annotation.** 36 cycle paired-end sequencing was carried out on an Illumina GAIIx, yielding between 80 and 150X coverage. SOAPdenovo (13) was used to generate de novo contigs, and reads aligned to a reference using Novoalign (NovoB drastically, Selangor, Malaysia). *S. Typhimurium* 4/74 reads were assembled on the genome and large plasmid of strain SL1344 (http://www.sanger.ac.uk/Projects/Salmonella/). *S. Choleraesuis* SCSA50 reads were assembled on the genome of strain SC-B67 (7) and its virulence plasmid (28). *S. Dublin* SD3246 reads were assembled on the genome of strain CT_02021853 (accession no. CP001144). *S. Gallinarum* SG9 reads were assembled on the genome of strain 287/91 (22). The *de novo* and reference contigs were combined using MUMmer (12) and Gap4 (5).

Sequences were annotated using GenoPipe (http://genopipe.bioinfo-portal.eds.ac.uk/) and a combination of gene prediction software (1, 8, 18, 21). Manual curation followed to enhance the annotation, including pseudogene prediction and assignment of start sites. Genes with unsuitable names for submission were searched against SwissProt (23), and genes with a large degree of overlap were checked for domains (2, 9) and hits in SwissProt. If no domains or matches were found, the gene was removed from the annotation.

Intraserovar comparisons indicated that the complete *S. Typhimurium* 4/74 genome contained just eight single-nucleotide polymorphisms (SNPs) relative to SL1344, consistent with the shared history of the strains and high-quality sequencing and assembly. The *hisG* allele varied between the strains as expected (10).

**Nucleotide accession numbers.** Sequences were deposited in GenBank and assigned the following accession numbers: *S. Typhimurium* 4/74 (CP002487-CP002490), *S. Choleraesuis* SCSA50 (CM001062 to CM001063), *S. Dublin* SD3246 (CM001151 to CM001152), and *S. Gallinarum* SG9 (CM001153 to CM001154).

We gratefully acknowledge the support of the European Commission (EADGENE network of excellence, contract number FOODCT-2004-506416), the Biotechnology & Biological Research Council (core...
strategic grants to The Roslin Institute and The Institute for Animal Health and India Partnering Award (IPA1825) and the Department of Information Technology, Govt. of India.

REFERENCES