**Title of the dataset:**

Whole genome sequencing of knockout mutants in *Streptococcus suis* strain P1/7 constructed using pSStarget, a novel plasmid-based CRISPR/Cas9 genome editing system

**Authors:**

Alex Gussak

Maria Laura Ferrando

Mels Schrama

Peter van Baarlen

Jerry Mark Wells

Host-Microbe Interactomics (HMI) Group, Department of Animal Sciences, Wageningen University and Research, Wageningen, the Netherlands

**Keywords**:

CRISPR, Cas9, genetic engineering, genomic manipulation, infectious disease, transformation, zoonotic agent, Streptococcus suis, whole genome sequencing, Illumina, NGS

**Related publication:**

A manuscript will be submitted for publication.

**Description:**

This dataset contains sequencing reads of the wild type strain *S. suis* P1/7 and two knockout mutants constructed using a novel CRISPR-Cas9 genome editing system for S. suis. Our data shows that our laboratory stock of strain P1/7 has 2 SNPs and a 5 bp deletion relative to the reference sequence (AM946016.1) published in GenBank. Moreover, we confirm that the two knockout strains lack the deleted *cpsEF* and s*ly* genes and do not have other mutations relative to the sequence of our laboratory stock P1/7.

**This dataset contains the following 6 files:**

SsuisP17\_1\_trimmed.fastq.gz

SsuisP17\_2\_trimmed.fastq.gz

SsuisP17\_dCPS\_1\_trimmed.fastq.gz

SsuisP17\_dCPS\_2\_trimmed.fastq.gz

SsuisP17\_dSLY\_1\_trimmed.fastq.gz

SsuisP17\_dSLY\_2\_trimmed.fastq.gz

**Explanation of variables:**

All files contain trimmed Illumina reads in the FASTQ (.fq) format in zipped (.gz) format.

SsuisP17 refers to the wild type strain *Streptococcus suis* P1/7. The appendix \_dCPS and \_dSLY refers to a knockout strain with the *cpsE/cpsF* and *sly* genes deleted, respectively

The number of the sample name, '\_1' and '\_2' refers to each member of a paired-end read pair.

**Materials and methods:**

All strains were grown on Todd-Hewitt (Oxoid) agar plates supplemented with 0.2% Bacto™ yeast extract (BD Biosciences) (THY) at 37 °C with 5% CO2. All growth was collected from the agar plate using a sterile loop and resuspended in a bead tube containing cryoperservative (Microbank™, Pro-Lab Diagnostics UK, United Kingdom) following MicrobesNG strain submission procedures. The bead tubes containing the bacteria were sent to MicrobesNG for DNA extraction and sequencing.

Five to forty microlitres of the bacterial suspension were lysed with 120 µL of TE buffer containing lysozyme (final concentration 0.1 mg/mL) and RNase A (ITW Reagents, Barcelona, Spain) (final concentration 0.1 mg/mL), incubated for 25 min at 37°C. Proteinase K (VWR Chemicals, Ohio, USA) (final concentration 0.1mg/mL) and SDS (Sigma-Aldrich, Missouri, USA) (final concentration 0.5% v/v) were added and incubated for 5 min at 65°C. Genomic DNA was purified using an equal volume of SPRI beads and resuspended in EB buffer (Qiagen, Germany). DNA concentration was quantified with the Quant-iT dsDNA HS kit (ThermoFisher Scientific) assay in an Eppendorf AF2200 plate reader (Eppendorf UK Ltd, United Kingdom).

Genomic DNA libraries were prepared using the Nextera XT Library Prep Kit (Illumina, San Diego, USA) following the manufacturer’s protocol with the following modifications: input DNA was increased 2-fold, and PCR elongation time was increased to 45 s. DNA quantification and library preparation were carried out on a Hamilton Microlab STAR automated liquid handling system (Hamilton Bonaduz AG, Switzerland). Pooled libraries were quantified using the Kapa Biosystems Library Quantification Kit for Illumina. Libraries were sequenced using Illumina sequencers (HiSeq/NovaSeq) using a 250bp paired end protocol. Reads were adapter trimmed using Trimmomatic 0.30 with a sliding window quality cutoff of Q15 [1]. De novo assembly was performed on samples using SPAdes version 3.7 [2], and contigs were annotated using Prokka 1.11 [3].

References

1. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114–2120. http://doi.org/10.1093/bioinformatics/btu170

2. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., ... Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology, 19(5), 455–477. <http://doi.org/10.1089/cmb.2012.0021>

3. Seemann T (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics. 30(14):2068-9

**License**

This dataset is published under the CC BY-SA (Attribution ShareAlike) license.