48-Plex Long-Read Sequencing to Generate Complete Assemblies and Determine Genome Structure in *Salmonella* Agona

Emma Waters[1], Winnie Lee[1,2], Amina Ismail Ahmed[3], Marie Chattaway[3], Gemma Langridge[1]

[1]Quadram Institute Bioscience, Norwich, UK. [2]University of Bristol, Bristol, UK. [3]United Kingdom Health Security Agency, London, UK.

Background

Short-read sequencing, alongside multiplexing, provides the resolution and high-throughput required to routinely identify SNPs in bacterial pathogens. Such small nucleotide-level variation can have huge effects, from changing antibiotic resistance to altering entire metabolic pathways.

Bacteria can also exhibit genomic variation, where large genome fragments shift position and/or orientation in the genome due to homologous recombination, for example around the long-repeat sequences of ribosomal operons (~5kb in length), producing different unique genome structures (GSs).[1] This variation cannot be identified by short-read sequencing as long-repeat sequences require reads of thousands of base pairs for resolution.

Using 48-plex, native-ligation, long-read sequencing (Oxford Nanopore Technologies), we can produce reads which span the repeat sequences of bacterial ribosomal operons to identify chromosomal DNA either side, assemble complete genomes and ultimately identify GSs. **This high-throughput, long-read sequencing method allows us to reliably identify and routinely investigate the impact of rearrangement in laboratory and clinical isolates of pathogens.**

Here we investigate GS relationships between strains of *Salmonella* Agona, associated with acute infection and chronic carriage. This non-typhoidal *Salmonella,* known for biofilm production and extreme persistence in food environments,[2] has been associated with 147 outbreaks across 5 EU countries between 2014 and 2016 alone.[3]

Methods

UKHSA identified 2,233 *S.* Agona isolates from infections in the UK (2004-2020). 1,155 had short-read sequencing data, which was interrogated for phylogenetic relationships using maximum likelihood trees (RAxML) where Snippy was used for core genome alignment and Gubbins was used to mask recombinant sequences.

208 isolates were selected for high-throughput, long-read sequencing to generate hybrid assemblies and investigate the role of GS in *S.* Agona infections (Fig. 1).

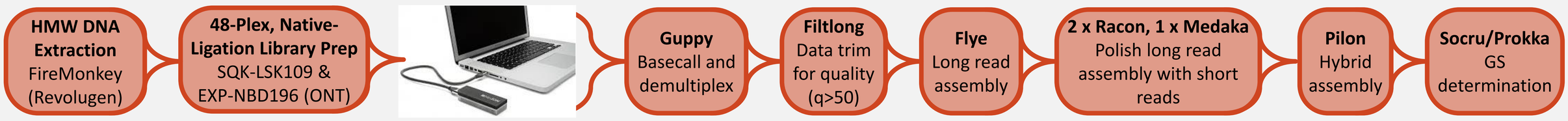

Figure 1. *Experimental and bioinformatic workflow to determine genomic variation, from high-molecular weight (HMW) DNA to long- and hybrid-assemblies*

Result: Carriage isolates are distributed across the phylogeny and intermixed with isolates that cause acute infection

Figure 2. *Maximum likelihood tree representing the core S. Agona genomes displaying the carriage status: sporadic, green; same episode blue; carriage, brown*

Result: Whilst >94 % isolates displayed the conserved arrangement of GS1.0, 7 imbalanced GSs were identified from 11 isolates

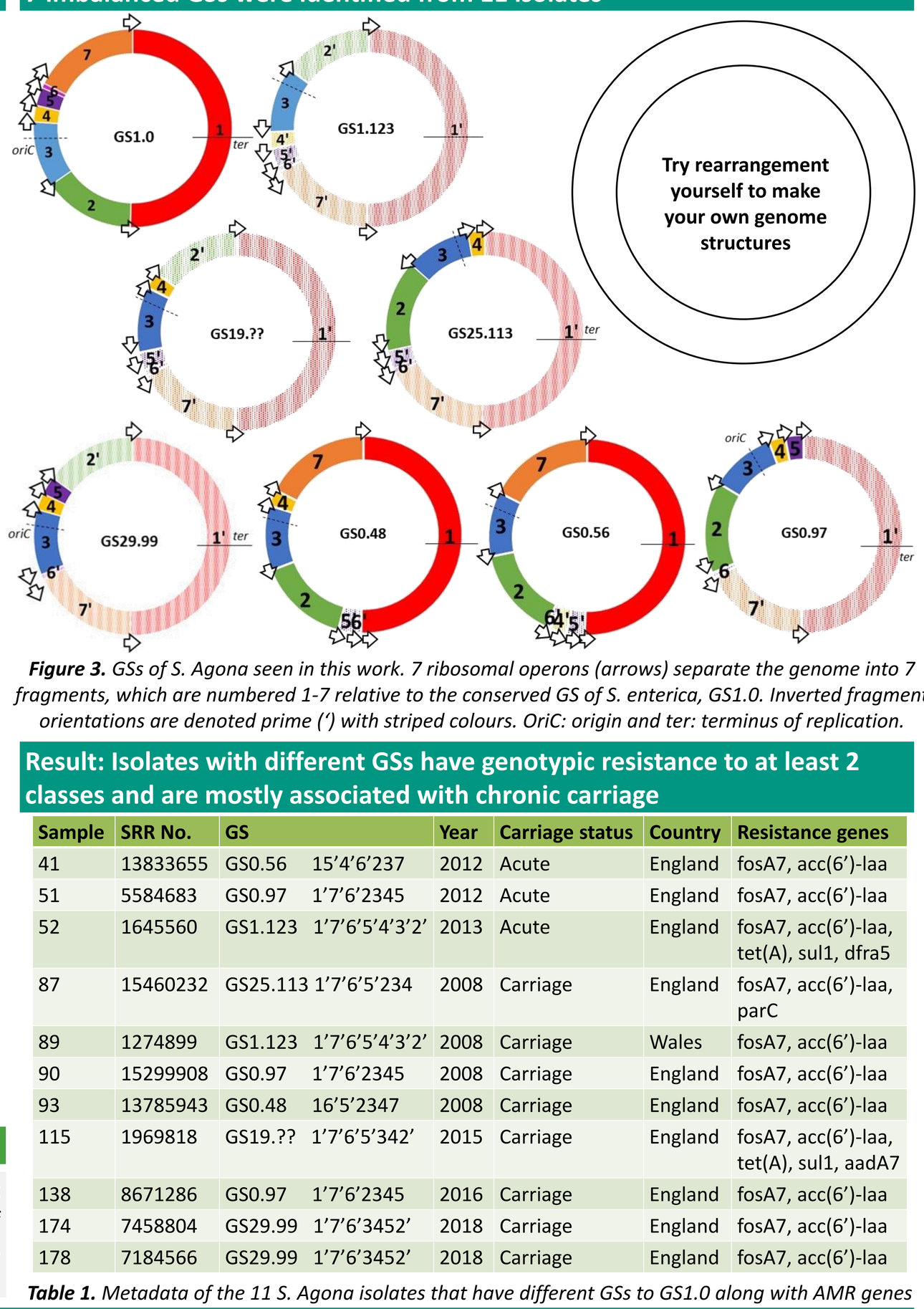

Figure 3. *GSs of S. Agona seen in this work. 7 ribosomal operons (arrows) separate the genome into 7 fragments, which are numbered 1-7 relative to the conserved GS of S. enterica, GS1.0. Inverted fragment orientations are denoted prime (') with striped colours. OriC: origin and ter: terminus of replication.*

Result: Isolates with different GSs have genotypic resistance to at least 2 classes and are mostly associated with chronic carriage

Sample	SRR No.	GS		Year	Carriage status	Country	Resistance genes
41	13833655	GS0.56	15'4'6'237	2012	Acute	England	fosA7, acc(6')-laa
51	5584683	GS0.97	1'7'6'2345	2012	Acute	England	fosA7, acc(6')-laa
52	1645560	GS1.123	1'7'6'5'4'3'2'	2013	Acute	England	fosA7, acc(6')-laa, tet(A), sul1, dfra5
87	15460232	GS25.113	1'7'6'5'234	2008	Carriage	England	fosA7, acc(6')-laa, parC
89	1274899	GS1.123	1'7'6'5'4'3'2'	2008	Carriage	Wales	fosA7, acc(6')-laa
90	15299908	GS0.97	1'7'6'2345	2008	Carriage	England	fosA7, acc(6')-laa
93	13785943	GS0.48	16'5'2347	2008	Carriage	England	fosA7, acc(6')-laa
115	1969818	GS19.??	1'7'6'5'342'	2015	Carriage	England	fosA7, acc(6')-laa, tet(A), sul1, aadA7
138	8671286	GS0.97	1'7'6'2345	2016	Carriage	England	fosA7, acc(6')-laa
174	7458804	GS29.99	1'7'6'3452'	2018	Carriage	England	fosA7, acc(6')-laa
178	7184566	GS29.99	1'7'6'3452'	2018	Carriage	England	fosA7, acc(6')-laa

Table 1. *Metadata of the 11 S. Agona isolates that have different GSs to GS1.0 along with AMR genes*

Conclusion

48-plex sequencing allows generation of complete assemblies for samples with at least ~20X theoretical coverage & improves affordability of routine long-read sequencing of bacteria (~£18 per *Salmonella*) to help identify GSs. We believe this variation provides a mechanism by which bacteria adapt to environmental pressures like chronic carriage.

References

1. Page, A. J., Ainsworth, E. V. and Langridge, G. C. (2020) *Socru*: Typing of genome-level order and orientation around ribosomal operons in bacteria. *Microb Genom,* **6**, 1-6.
2. Galié, S., García-Gutiérrez, C., Miguélez, E. M., Villar, C. J., and Lombó F. (2018) Biofilms in the Food Industry: Health Aspects and Control Methods. Front Microbiol, **9**, 1-18.
3. European Centre for Disease Prevention & Control/European Food Safety Authority (2018) Multi-country outbreak of Salmonella Agona possibly linked to ready-to-eat food – 26 July 2018.